



A physically grounded input distance resolves longitude-encoding ambiguity but not specification dependence in ensemble persistent homology of cyclone trajectories

Rongzhen Dai

independent researcher: Yangzhou, China

Correspondence: Rongzhen Dai (dairongzhen3@gmail.com)

Received: 24 May 2026 – Discussion started: 23 June 2026

Revised: 6 September 2026 – Accepted: 8 October 2026 – Published: 20 October 2026

Abstract. We show that a physically grounded input distance fixes an ensemble persistent-homology test's geometry, and measure what that repair leaves untouched. On the ± 15 d windows of three benchmark geomagnetic storms (25 event storms, 1020 calendar-matched controls), three principled longitude encodings give three reproducible verdicts: one-sided H_1 resampling $p = 0.108$ for a linear modular encoding on a single longitude convention, 0.021 for a unit sphere, 0.156 for a cylinder. Across six auditable control-draw orderings a 49-placebo calibration is nominal in H_1 throughout (9–12 of 147) while H_0 runs 4–16, crossing the nominal rate at its upper ordering. A great-circle distance removes the coordinate ambiguity and returns one value, and the event–control difference then survives per-window, basin, era and joint matching, though the era-matched design fails two-sided under the conservative doubling convention. It does not survive a change of the matching exponent or of the event subsample, and is absent under the true degree-2 and exact bottleneck distances. Selecting none of these, the specification-averaged p -value is 0.082 over exponents (basin-matched) and 0.122 and 0.134 over wider unmatched families, with the event above the control median in all 42 cells. We report a design- and specification-limited directional residual: no detection, and no evidence of absence.

1 Introduction

1.1 The solar–cyclone question, in brief

Whether solar and geomagnetic activity perturbs tropical cyclone (TC) dynamics has a long and contested observational lineage. Elsner and Kavnikov (2001) reported correlations between geomagnetic indices and per-storm intensity changes; Elsner and Jagger (2008) linked US and Caribbean hurricane counts inversely to sunspot number; Pérez-Peraza et al. (2008), Mendoza and Pazos (2009), Hodges and Elsner (2011) and Hodges et al. (2014) reported solar-cycle dependences in various basins, and Li et al. (2023) and Nayak (2024) extend and review the picture. Proposed mechanisms range from the cosmic-ray–cloud pathway (Marsh and Svensmark, 2000; Svensmark, 2007; Kirkby et al., 2011; Yu

and Luo, 2014; Svensmark et al., 2016) to downward wave–mean-flow coupling (Huth and Stryhal, 2024) and ionospheric gravity-wave forcing (Prikryl et al., 2019). Against this stands an equally substantial null literature: Forbush-decrease cloud analyses find no robust response (Kristjánsson et al., 2008; Calogovic et al., 2010; Laken et al., 2012), the ion-induced enhancement of cloud condensation nuclei appears too weak to matter climatically (Pierce, 2017), and IPCC AR6 WGI (2021, chap. 7) reflects the consensus that galactic cosmic rays contribute negligible radiative forcing. Across two decades effect sizes are modest, mechanisms vary, and null controls are unevenly applied. We treat the question as open, and use it here as the setting for a methodological one.

1.2 The problem: a verdict that depends on unregistered choices

Existing observational studies almost universally operate at the per-storm scale. A more recently proposed scope treats all the cyclones active in one time window as a single cloud of points in (latitude, longitude, intensity) coordinates and asks about its shape. The tool for that is persistent homology (Carlsson, 2009; Ghrist, 2008; Edelsbrunner and Harer, 2010): one grows a ball of radius r around every point, records the radius at which each hole in the resulting union opens and the radius at which it closes, and summarises the cloud by the list of those (birth, death) pairs. A worked example makes the idea concrete. Six storms whose tracks happen to encircle a quiet region of ocean leave a ring of points with an empty middle; the balls around them meet each other before they fill the middle, so a hole opens at the radius where neighbours touch and closes at the radius where the middle is covered, and the width of that interval is the one number the method keeps about it. A perturbation shared across many storms, if present, might change how many such holes the ensemble has and how long they survive, in a way no single storm's trajectory reveals.

Related randomisation two-sample tests for collections of persistence diagrams were introduced by Robinson and Turner (2017), using losses built from within-group pairwise matching costs, and later extended by Islambekov and Pathirana (2024). Our resampling construction below differs in form, but shares the same dependence on the chosen diagram distance. Robinson and Turner are explicit about which guarantee their test carries and which it does not: under the null the p -value is uniform, but they know of no theory that it will be small when the distributions differ. The exponent of the matching cost is a parameter of the procedure, not a quantity the data fix.

Three further bodies of work bound what follows. Stability theory does reach across a change of metric: Chazal et al. (2014) bound the bottleneck distance between the Vietoris–Rips and Čech diagrams of two totally bounded spaces by the Gromov–Hausdorff distance between them. What that bounds is the distance between diagrams, not the verdict a test built on those diagrams returns, and it is the verdict that moves here. Substituting an intrinsic distance for an ambient one has been done before: Beshkov and Tiesinga (2022) exchange the Euclidean metric for a geodesic one in the persistent homology of neural activity and recover different features under the two. And reporting a result over the set of defensible analyses rather than one selected from it is established elsewhere as specification curve analysis (Simonsohn et al., 2020) and multiverse analysis (Steen et al., 2016), with post-selection inference supplied by Girardi et al. (2024). What this paper adds to those three is a single dataset on which the representation is fixed on physical grounds, the residual dependence on the remaining unregistered choices is measured rather than argued,

and the specification-averaged verdict is reported in a setting where the specifications are topological rather than regression choices.

Several steps of that comparison are chosen by the analyst, and the method does not fix them. Where to place the points: the longitude coordinate is periodic, and three principled treatments of that periodicity exist, each with defenders. How to charge for a mismatch between two lists of holes: one may add the mismatches as they are or add their squares, and nothing in the method says which. Which points enter the fixed-size subsample, and which control diagrams form the reference set against which the event is measured. A reader handed a p -value cannot tell how much of it came from the data and how much from those steps, and the author often cannot either, because each is made once, early, and never varied. If they can move a verdict across the conventional threshold, then every ensemble-topology result in the applied literature carries an unquantified term. Invariance under them is thus a precondition: a pipeline whose verdict depends on how the trajectories are represented, without converging, is not yet ready to test the substantive hypothesis.

1.3 This paper

We establish three things on one dataset. We show that the input geometry can be fixed on physical grounds: replacing the three coordinate encodings by a great-circle distance on the sphere the storms live on removes the ambiguity of which geometry to trust and returns a single verdict where the encodings returned three. We measure what that repair leaves untouched, and demonstrate that it is partial – the verdict the physical distance returns still moves with the exponent of the matching cost (resampling p from 0.003–0.087), with which 600 of the ensemble's fixes enter the diagram, and with which fifty control diagrams form the reference set, three quantities that had been fixed without being registered. We report, third, a test that selects no exponent and averages over the subsample realisation, and we give its specification-averaged p -value for three widths of family.

The central statement is therefore a qualified one. A physically interpretable input geometry removes one arbitrary choice; the event-level verdict still depends on the diagram metric and on the finite-sample realisation, and a test that selects neither leaves the effect below threshold.

The vehicle is a cyclone-trajectory dataset with a clean structure for the purpose. We construct three encodings of the (lat, lon, wind) point cloud – a linear modular normalization that ignores periodicity, a unit-sphere embedding and a cylinder embedding – and apply the same downstream pipeline to the same 25-storm event set and 1020-storm control pool. Persistence is computed by the Vietoris–Rips construction, whose object is the clique complex of the proximity graph, in which a simplex enters as soon as all its vertices are pairwise within r ; that complex is closely related to but distinct from the Čech complex the nerve theorem ties to the union of balls.

Two persistence lists are compared by matching their holes and summing the cost of the matching, and the null is built by resampling control storms. All storm sets are held fixed across representations, so across-representation differences cannot arise from event or control selection. This analysis is observational, not controlled: we manipulate the representation and the resampling while leaving the physical system untouched, and what we report is within-pipeline robustness, not causal isolation.

2 Data

Catalog. IBTrACS v04r01 satellite-era subset (Knapp et al., 2010; dataset: Knapp et al., 2019): 6661 storms with valid latitude and longitude, 406 059 6 hourly fixes, downloaded from NOAA NCEI on 30 April 2026. The retrieved archive has SHA-256 e569d503e70165b142f445aa93162682b6d9364d10d68682fa76c184c571b782 and the satellite-era subset actually analysed 02f85e0ceb1deb38a358bf0ee2ee747a726850de920517ae0e29aac3a3b729fd, which the verification script checks before it recounts anything.

Event selection. Three events, one benchmark geomagnetic storm per recent solar cycle: Halloween 2003 (peak Dst ≈ -383 nT, Cycle 23), St Patrick's Day 2015 (≈ -234 nT, the strongest Cycle 24 storm with adequate cross-basin TC activity in the window), Gannon 2024 (≈ -406 nT provisional, Cycle 25). They span three solar cycles and three TC seasons and are the only modern-era extreme events with adequate IBTrACS overlap to support an ensemble test; the full candidate log is shown in Sect. S5 in the Supplement.

Event windows. Peak dates 29 October 2003, 17 March 2015 and 10 May 2024, each ± 15 d. Two date conventions are in play – the storm-label dates the analysis code uses, and the Kyoto WDC index-minimum dates, one day later for two of the three, on which the candidate enumeration of Sect. S5 is built – and recentring on the latter admits one fewer storm and changes no conclusion reported here (Sects. S1 and S5). Every number in this paper uses the 25-storm set defined by the labelled dates. A storm enters the event set if any of its fixes falls in any window (union rule): 25 unique storms across basins {WP, NI, EP, SI, SP}, 1880 fixes, held fixed across all representations. The ensemble then uses all lifetime fixes of each selected storm – the storm, not the individual fix, is the unit of exposure and of resampling (Sect. 3.5) – so it includes fixes outside the ± 15 d window and weights longer-lived storms (mean 75 fixes each) more heavily. The test therefore probes the topology of window-overlapping storms, a wider object than the window interval alone, and a window-cropped variant is a natural sensitivity for a larger study.

Composition. The event set is basin-uneven (WP-dominated for Halloween, SI/SP for St Patrick, SI for Gannon) and, relative to the control pool, more recent: mean gen-

esis year 2011.6 against 2001.8, and more SI/WP (17 of 25). Both features are intrinsic to the three calendar choices and are the subject of the matched-null controls of Sect. 4.5. The basin-by-event breakdown is shown in Table S3 in the Supplement.

Control pool. For each event, the same calendar window in every other year 1981–2024, excluding the event year ± 2 . Pooled: 1020 unique control storms, 69 454 fixes. Event and control sets are held fixed across all representations.

3 Methods

3.1 Persistent homology: a primer

Persistent homology summarises the multi-scale structure of a point cloud (Edelsbrunner and Harer, 2010; Carlsson, 2009; Ghrist, 2008; Wasserman, 2018), with an environmental-science primer in Ver Hoef et al. (2023). Over the Vietoris–Rips filtration, connected components (H_0) and loops (H_1) are born and die at scales whose pairs form the persistence diagram, which Cohen-Steiner et al. (2007) proved stable under perturbations of the input in a fixed input metric. One H_0 class never dies, and that essential class is dropped from every diagram, event and control alike, before any distance is computed (`finite_pd`), so all H_0 figures below are over the finite classes. We restrict throughout to H_0 and H_1 ; higher-degree homology is not computed. Neither channel has a literal physical counterpart – H_1 features are not closed storm tracks, H_0 components are not storm clusters – and both aggregate across storms in a way mean-intensity statistics do not.

3.2 Distances between persistence diagrams

We quantify differences between diagrams with the degree-1 Wasserstein distance W_1 , a complete separable metric on diagram space (Mileyko et al., 2011) that sums the cost over all matched pairs. The exponent is not incidental to what follows, so we state the implementation precisely. W_1 is computed with `persim.wasserstein` (persim 0.3.8; Saul and Tralie, 2019), whose signature carries no exponent argument and whose returned value is the plain sum of the matched costs under a Euclidean ground metric. Earlier versions of this manuscript described the same quantity as a Wasserstein-2 distance – the variant that adds the squares of the matched costs and takes a square root – a description that misnamed the implementation while leaving it correct; Sect. S1 records the correction. Because the exponent turns out to change the verdict, we recompute the load-bearing results under the true degree-2 distance and under the exact bottleneck distance W_∞ , the latter with `gudhi` 3.12 (The GUDHI Project, 2015). The three are different metrics, not three points on one continuum: W_∞ replaces the Euclidean ground norm by the sup norm and the sum by a maximum, so it should not be read as the $p \rightarrow \infty$ limit of the same family.

3.3 Three longitude encodings and one physical distance

Each fix is a vector in (latitude, longitude, wind) space. Latitude rescales linearly to $[0, 1]$; wind is $\text{WMO_WIND}/200$, clipped. The intensity channel is largely imputed: WMO_WIND is missing for 268 335 of the 406 059 satellite-era fixes, or 66.1 %, and a missing value is replaced by the median wind of the subset being encoded. That rule is itself a choice, and the verdict is sensitive to it: replacing the per-subset median by a single global constant moves the great-circle event p -value from 0.012–0.034. Neither rule is a natural standard, so we report both and treat the difference as part of the specification dependence quantified in Sect. 4.7. Longitude is periodic; three encodings are tested.

3.3.1 Linear modular (3-D)

$\text{lon_n} = (\text{lon} + 180)/360$, $\text{lat_n} = (\text{lat} + 90)/180$, wind_n . Minimal and directly interpretable, but the metric does not respect periodicity: -179 and $+179^\circ$ sit ≈ 0.994 apart though physically 2° apart. A second defect surfaced after the analyses below were run: the archived IBTrACS snapshot does not use a single longitude convention, so 2.19 % of fixes are recorded east of $+180^\circ$ while others carry the negative representation of the same meridians, and lon_n applied to that field places two representations of one meridian 1.0 apart at the edge of the domain, which is where the large-scale structure of a Vietoris–Rips filtration is set. Reducing the field to a single convention and holding everything else fixed moves the linear H_1 resampling p -value from 0.037–0.108 and the event distance from 0.353–0.328, and rotating the branch cut by 0, 90 and 180° gives 0.108, 0.081 and 0.022 (Sect. S7; run logs `results/independent_checks/lon_*.log` and `rot_*.log`, the first of which is a positive control reproducing the archived event distance and p -value bit for bit). The sphere, cylinder and great-circle representations pass longitude through a periodic function before any distance is taken and are unaffected.

3.3.2 Unit-sphere (4-D)

$(\text{lat}, \text{lon}) \rightarrow S^2$ via $x = (\cos \text{lat} \cos \text{lon} + 1)/2$, $y = (\cos \text{lat} \sin \text{lon} + 1)/2$, $z = (\sin \text{lat} + 1)/2$, with wind appended. Periodicity is exact and the flat-circle redundancy is avoided.

3.3.3 Cylinder (4-D)

lat_n linear, longitude on a flat $(\cos, \sin)$ circle, wind appended. Periodicity is exact and latitude and longitude decouple, but the $\cos^2 + \sin^2 = 1$ constraint means wide-longitude ensembles can populate a ring that Vietoris–Rips

may register as a persistent H_1 loop with no physical content beyond periodicity.

3.3.4 Construction, sub-sampling, and a reproducibility correction

Each ensemble's fixes are normalized, concatenated and sub-sampled uniformly at random to $N_{\text{sub}} = 600$, with the seed fixed at script entry (`np.random.default_rng(2026)`) and consumed sequentially. Three corrections separate the present pipeline from the single-run implementation behind the preprint, and Sect. S1 tabulates them with the values they replace: that implementation built its control array from an unordered set, so it drew a different control sample in every process even with the random seed fixed, and the interpreter hash seed of the original runs was not recorded; it retained only one coordinate normalizer at a time, variants being edited in place without version control, so that its linear row cannot be audited while its sphere and cylinder rows reproduce to within a few percent; and because the earlier values are not reproducible from the archived method, we do not carry them forward. Determinism is now specified as an explicit sort of the stratum and identifier lists, in place of a pinned interpreter hash seed, since pinning the hash seed reproduces one historical run while two logically equivalent implementations can still build a set in different orders under the same seed.

3.3.5 Ambient-dimension control

The linear encoding is 3-D while sphere and cylinder are 4-D, so a linear-versus-4-D difference could confound longitude representation with ambient dimension. Two controls separate them. A 4-D linear encoding with a constant fourth coordinate is a code-path identity check, since a constant coordinate contributes zero to every pairwise distance, and it reproduces the 3-D result exactly: event distance 0.3528, null mean 0.2910, null standard deviation 0.0290, resampling p 0.0370 in H_1 and 0.0520 in H_0 , with the pairwise distances agreeing to 6.6×10^{-8} . The evidential weight is therefore carried by the second control, a 4-D linear encoding with a Gaussian fourth coordinate. That genuine added dimension leaves the placebo calibration nominal (H_1 false positives 2/49), while it moves the event verdict from $p = 0.037$ –0.072 as the added variance dilutes the event H_1 signal (event H_1 features rise from 83–207). The two claims are kept separate: the calibration behaviour is dimension-robust and the event p -value is dimension-sensitive, so the confound is real at the level of the Sect. 4.1 comparison and absent from the Sect. 4.2 diagnosis.

3.3.6 A physically interpretable common distance

The three encodings are three different geometric models, so their disagreement partly reflects that fact, and need not indicate instability of a single well-posed test. We therefore also build one physically interpretable pairwise distance and feed it directly to the persistence computation as a distance matrix, so that only the distance step changes while storm selection, sub-sampling and the statistic are identical:

$$d(i, j) = \sqrt{[(\gamma(i, j)/\pi)^2 + (\lambda \cdot |\Delta \text{wind}|/200)^2]}, \quad (1)$$

where $\gamma(i, j)$ is the great-circle angle between fixes i and j , so longitude periodicity is exact and no encoding artefact or spurious cylinder loop arises, and λ is an intensity weight sensitivity-tested at $\lambda \in \{0, 0.5, 1.0\}$. $\lambda = 0$ is pure track geometry; $\lambda = 1$ places geometry and intensity on equal footing, matching the encoders whose normalized coordinates all live on $[0, 1]$. Results are in Sect. 4.5.

3.4 Prior TDA in atmospheric science

Persistent homology has recognised atmospheric rivers in gridded water-vapour fields (Muszynski et al., 2019), a diurnal cycle in single-storm cloud imagery (Tymochko et al., 2020) and weather regimes via attractor topology (Strommen et al., 2022); Ghil and Sciamarella (2023) review its wider use in the climate sciences, and Inatsu et al. (2017) come closest in subject, identifying individual cyclones from a superlevel-set filtration with a merge-tree reconstruction. Our scope – the joint point cloud of many storms’ fixes – differs from each, and we put the encoding-invariance question at that scope.

3.5 Test statistic, resampling, and calibration

For each representation the event diagram is compared against a resampling distribution of control diagrams by mean W_1 distance to a reference set of fifty control diagrams; the null is the same statistic with each control diagram substituted for the event diagram, jackknifing the reference; and the resampling p -value is the fraction of null draws whose distance is at least the event distance. This is a Monte-Carlo empirical null obtained by resampling the control pool, not a label-permutation test: no labels are exchanged, and the draws are without replacement.

That statistic is one-sided: the pre-specified alternative is that the event ensemble lies farther from the reference set than the controls do, and this convention was fixed in the discussion-paper version and is not changed here. Because sidedness is a convention rather than a result, Sect. S10 gives the two-sided companion of every value on which a threshold statement in this paper rests, under both of the conventions available, and where a statement below turns on which side of α a value lies we name the convention it rests on. Each iteration draws n storms matching the event count and

concatenates their fixes, so the resampling unit is the storm and within-track autocorrelation is preserved. Placebo false-positive rates are compared to nominal with exact-binomial tests and per-proportion Wilson (1927) intervals. The finite-sample correction $(1 + \#\{\text{null} \geq \text{observed}\})/(1 + B)$ sometimes recommended for resampling tests changes no qualitative verdict and no placebo classification here; its size stays below 0.001 for the $B = 1000$ event runs, reaching 0.0099 at the $B = 100$ used per placebo and 0.002 at the $B = 500$ of Sect. 4.3 and 4.4.

A resampling p -value estimated from B draws carries its own sampling error, and near a decision threshold that error is not negligible. At $B = 1000$ the Wilson 95 % interval on an estimate of 0.037 runs from 0.027–0.051 and on 0.044 from 0.033–0.059: both contain 0.05, so at that resolution the statement that either lies below α is not separable from Monte-Carlo noise. We therefore recomputed the five load-bearing results at $B = 10000$, holding the reference set fixed so that the estimand does not move with B (Sect. 3.6). The intervals tighten by roughly a factor of three and none now contains 0.05: linear 0.0343 [0.0309, 0.0380], sphere 0.0120 [0.0100, 0.0143], cylinder 0.1516 [0.1447, 0.1588], the basin-matched subsample-averaged great-circle result 0.0427 [0.0389, 0.0468], and the same under index-minimum event dates 0.0329 [0.0296, 0.0366]. Where a statement below turns on which side of α an estimate lies, it rests on the $B = 10000$ value; everything else stays at $B = 1000$, being internal comparisons in which every entry shares one resolution.

Three further additions relative to a single-run analysis are motivated by the fragility we find. (i) *A fixed ordering, with Monte-Carlo uncertainty reported*. The control-draw order is a free implementation choice; we report the placebo false-positive count as a distribution over six auditable orderings, a deterministic sort and five seeded permutations of it. Its width is largely Monte-Carlo (Sect. 4.2), so reordering a uniformly resampled control pool selects a different realisation while the estimand does not move; the durable requirement is therefore the narrower one, to fix the ordering by an explicit sort key so the run is reproducible and to report the Monte-Carlo uncertainty of the resulting p -value alongside its point value. (ii) *Matched null*. In addition to uniform control draws we draw controls preserving the event ensemble’s per-window (seasonal) storm counts (11/9/5), its genesis basin, its genesis era in 5 year bins, and the joint basin-by-era composition (Sect. 4.5). (iii) *Subsample averaging*. Where a statistic carries a conclusion, it is averaged over $M = 10$ independent subsamples of the event ensemble and of every control draw alike (Sect. 4.7).

3.6 A third unregistered choice: the reference set

The event statistic is the mean distance from the event diagram to a reference set of fifty control diagrams, and which fifty is itself a choice that was never registered. The refer-

ence set is drawn from the pool of null draws, so enlarging that pool changes it: recomputing the linear encoding at $B = 10000$ without pinning it moved the event distance from 0.352771–0.349897 and the p -value from 0.0370–0.0338 – a change in the estimand, not a gain in precision. Every $B = 10000$ result reported here therefore pins the reference set to the one the archived $B = 1000$ run used. Drawing five further reference sets of fifty from the same pool, and changing nothing else, moves the linear resampling p -value over 0.036–0.047 against 0.037 for the archived set: a spread of the same order as the Monte-Carlo interval at $B = 1000$ and, unlike that interval, one that more null draws do not reduce. It flips no verdict here, and a study whose estimate sat closer to the threshold could be moved across it by this choice alone. We record the reference indices with the results and recommend that future work register the reference set alongside the exponent and the subsample.

3.7 Pipeline runs and implementation

D1 pool: 25 event against 1020 control storms, $B = 1000$. The 49-placebo calibration uses the same calendar windows at random years 1985–2020 outside event years ± 2 , $B = 100$ each. Stratified attribution covers two basin cells and three single-event cells at $B = 500$, corrected by the procedure of Benjamini and Hochberg (1995). Subsample-size sensitivity covers $N_{\text{sub}} \in \{400, 600, 800, 1000\}$. Python 3.11; numpy, scipy, pandas, matplotlib; ripser 0.6.14 (Bauer, 2021); persim 0.3.8; gudhi 3.12 (see Otter et al., 2017, for a survey of persistent-homology computation). The encodings and the great-circle distance are switchable variants of one pipeline module, so all representations share the same selection and resampling code, and the library version does not affect the verdict: numpy 1.26 and 2.4 give bit-identical resampling p -values once the control order is fixed.

Every specification reported here is defined in one place and every artefact carries its own configuration. A specification record fixes the encoding, the exponent and ground metric of the matching cost, the null design and stratum, the control-draw ordering, the subsample size and count, the number of null draws, the reference-set size, the random seed, the imputation rule and the stratum-sorting rule. Output file names are derived from that record and never typed, and each output header records the resolved encoder function actually called, the SHA-256 of the input archive, the library versions and the code revision – the resolved encoder in direct answer to the failure of Sect. 3.3.4, in which a label and the implementation behind it had come apart. A verification script re-derives all 87 registered specifications from the archived data and compares each with the value printed here; all 87 agree to the precision at which they are reported. That check covers the resampling p -value of every registered specification and, at $p = 1$ under the Euclidean ground metric, the archived event distance as well. It does not cover every number in this paper: Sect. S10 sets out what falls outside

it, including the five values recomputed at $B = 10000$ and eighteen of the thirty-six cells of the matched-null comparison of Sect. 4.5, and says where each is anchored instead.

4 Results

4.1 The D1 pool: the verdict moves with the encoding

Table 1 gives the pooled event-versus-control comparison under the three encodings and, for reference, under the physical distance of Sect. 4.5.

Three representations of the same 25 storms give three verdicts. Under the archived longitude convention the linear and sphere encodings cross $\alpha = 0.05$ and the cylinder does not; under the single convention only the sphere does. A single-encoding readout would report a positive signal or a null on the same data depending on which encoding was chosen and, for the linear encoding, on where the circle was cut. The relative structure is the more stable part: under every one of the six event-subsample seeds the sphere p -value is the smallest of the three and the cylinder p -value the largest, while the position of any of them relative to 0.05 is not reproducible, and the particular combination printed in the first three rows occurs in one seed of six. Section S6 gives the six seeds in full, together with the ENSO and solar state of every event and placebo window. An earlier single run reported 0.130/0.009/0.214 for these rows; Sect. S1 records why those values are superseded.

4.2 Placebo calibration: nominal in H_1 , with a residual ordering spread

The 49-placebo calibration runs the full pipeline on 49 fake event-date sets at the same calendar windows in random years. Table 2 reports the pooled false-positive count in each homology channel as a distribution over equally valid control-draw orderings.

Every H_1 ordering is nominal-compatible. The H_0 spread is the wider of the two in absolute terms and its upper ordering crosses the nominal rate, so what distinguishes the channels is not that H_0 is stable and H_1 is not: the H_1 verdict sits near α at every ordering while the H_0 verdict is far from it at most. The reproducible pipeline is H_1 -nominal with a mild ordering spread, and an earlier single run's H_1 count of 16/147, read at the time as an inflation, cannot be placed among the auditable orderings (Sect. S1). Much of even the 9–12 spread is finite-resampling noise: a per-placebo decomposition attributes roughly half to two-thirds of the cross-ordering movement to the $B = 100$ term alone, so raising B would compress it. Neither pooled count is a set of independent trials – the three encodings share the same 49 dates – so the exact-binomial p -values understate the sampling uncertainty and are used only to place the counts relative to the boundary, which the per-encoding Wilson intervals inde-

Table 1. D1 pool ensemble TDA under three longitude encodings, on identical event and control sets. Columns: mean distance from the event diagram to fifty reference diagrams; null mean over 1000 control draws; H_1 resampling p at $B = 1000$ and at $B = 10\,000$; and the range of the H_1 resampling p over six event-subsample seeds at $B = 1000$, seed 2026 having been frozen before the analyses of Sect. 4.1–4.5 and the five others added afterwards. The linear row is reported under both longitude conventions: the archived one, under which Sect. 4.2–4.4 were also computed, and the single convention of Sect. 3.3.1, whose value is one point in the range 0.022–0.108 spanned by the branch cut (Sect. S7). A positive control reproduces the archived cylinder seed row bit for bit. Boldface identifies the resampling p -values in the two pH1 columns ($B = 1000$ and $B = 10\,000$), irrespective of whether they are below the 0.05 threshold.

Representation	event H_0 dist	event H_1 dist	null mean H_1	pH₁, $B = 1000$	pH₁, $B = 10\,000$	pH ₁ over six seeds
Linear modular, archived convention	1.505	0.353	0.291	0.037	0.0343	0.037–0.491
Linear modular, single convention	–	0.328	–	0.108	–	–
Unit-sphere	2.251	0.656	0.509	0.021	0.0120	0.014–0.139
Cylinder	2.120	0.546	0.492	0.156	0.1516	0.153–0.617
Great-circle, $\lambda = 1$ (Sect. 4.5)	–	0.546	0.408	0.012	–	0.012–0.124

Table 2. Pooled placebo false-positive count, of $147 = 49$ placebos $\times 3$ encodings, across six auditable control-draw orderings: a deterministic sort and five seeded permutations of it, with the main resampling seed held fixed. The six values in each channel are listed in ascending order, not in the order of the permutations that produced them. Exact-binomial p -values against nominal 5 % are quoted at the extremes of each channel; Sect. S2 gives the per-encoding breakdown of the deterministic ordering.

Channel	ordering values (of 147)	pooled range	exact binomial at the extremes
H_1	9, 9, 10, 10, 11, 12	9–12	$p = 0.45$ at 9; $p = 0.086$ at 12
H_0	4, 5, 8, 11, 11, 16	4–16	$p = 0.0035$ at 16

pently support: all contain 5 %, at H_1 1/49, 4/49, 4/49 against H_0 2/49, 1/49, 2/49 on the deterministic ordering.

The ordering sensitivity co-occurs with a measurable difference between the channels. On the reproducible cache the coefficient of variation of the event-to-null distance is 2.0–2.2 times larger for H_1 than for H_0 and that of the null mean itself 3.9–4.4 times larger. Both are measured across placebos, so neither establishes that the H_1 null is narrower within any one placebo; what they establish is that the H_1 summaries move more from placebo to placebo. Alongside them sits a population of threshold-straddling placebos: 17 of the 147 combinations, 11.6 %, have a resampling p between 0.02 and 0.10 under either the published implementation or the deterministic one. Across three orderings on a matched null, the H_0 p in fact swings marginally more than H_1 (mean range 0.158 against 0.135); the H_1 swing straddles α because the event distance sits near the upper tail of the H_1 null.

Two candidate explanations do not account for it. Against the exact bottleneck distance over three orderings on a matched null, W_1 gives comparable H_1 totals (11/15/11 against 12/10/12), and per placebo the bottleneck has the larger cross-ordering range (mean 0.214 against 0.135), a sign test on the paired ranges rejecting equality at $p = 1.5 \times 10^{-6}$: the exact bottleneck is if anything the more order-sensitive of the two. It is a different-metric control, changing the aggregation from a sum to a maximum and the ground norm from Euclidean to sup, and it retains matching to the diagonal, which our implementation constructs explicitly. Am-

bient dimension is likewise excluded by the 4-D linear control of Sect. 3.3.5, and two statistic-level remedies also fail to change the picture, an L^2 -persistence-landscape statistic (Bubenik, 2015) giving 12/147 and a z -score normalization of the event distance 14/147, against 9/147 for the sorted W_1 baseline. These placebo counts are the pipeline’s empirical rejection frequency under realistic historical variability, folding methodological behaviour together with the climate, era and basin variability of the placebo years.

4.3 Stratified attribution

Table 3 breaks the pooled verdict into two basin cells and three single-event cells.

No cell survives correction in any encoding. The most extreme cell is D2_WP in all three, at 0.048–0.142 uncorrected, which is a weaker localization than an encoding-dependent extreme would be: the three representations agree on which cell is least consistent with its null, and none of them puts that cell past correction. Halloween 2003, the canonical extreme, is null in all three at 0.248–0.338. Cell-level H_0 p -values are above 0.18 everywhere. The pooled D1 verdict is not localizable to a coherent single-cell source under any encoding.

4.4 Subsample-size sensitivity

In Table 4 the linear encoding is monotonic in N_{sub} and strengthens at $N = 1000$; the sphere and cylinder encodings are not monotonic and weaken there. We read the ta-

Table 3. Stratified attribution. Uncorrected H_1 resampling p per cell at $B = 500$, with the Benjamini–Hochberg value in parentheses where the uncorrected p is below 0.10. The correction is applied within each encoding over that encoding’s five cells. Section S1 records why these values replace those of earlier versions.

Cell	n event	n control	Linear H_1	Sphere H_1	Cylinder H_1
D2_WP	9	240	0.082 (BH 0.41)	0.142	0.048 (BH 0.24)
D2_SI	9	194	0.168	0.670	0.590
D3_Halloween	11	519	0.318	0.248	0.338
D3_StPatrick	9	314	0.338	0.182	0.306
D3_Gannon	5	187	0.674	0.850	0.788

Table 4. D1 pool H_1 resampling p across subsample sizes, $B = 500$ per cell, under the archived longitude convention.

N_{sub}	Linear	Unit-sphere	Cylinder
400	0.062	0.214	0.314
600	0.048	0.018	0.112
800	0.048	0.004	0.206
1000	0.026	0.044	0.450

ble as ordinary sampling variability of the H_1 channel; it is not a size dependence, since the same event realization tested at different subsample sizes draws different point subsets, so each entry is one seeded sample. The $N = 600$ row and the corresponding column of Table 1 differ because the two are computed at different resolutions, $B = 500$ here against $B = 1000$ there, on independently drawn nulls; the differences (0.048 against 0.037, 0.018 against 0.021, 0.112 against 0.156) are of the size the Monte-Carlo intervals of Sect. 3.5 predict. Section 4.7 replaces the single realization by an average over ten.

4.5 A physically interpretable common distance

We replace the three encodings by the single great-circle-plus-intensity distance of Eq. (1) and repeat the D1 test and the calibration. Under $\lambda = 1$ the H_1 resampling p -value stays within [0.006, 0.017] across three control-draw orderings without crossing α , and the placebo calibration is nominal across seven such orderings, with per-ordering H_1 false positives of 2–4 of 49. On the axis on which the three arbitrary encodings disagreed, the physically grounded distance gives a single value. That is stability across control-draw orderings, one of the pipeline’s two random choices; Sect. 4.7 shows it does not extend to the other.

The verdict depends on the intensity weight. With pure track geometry ($\lambda = 0$) the event H_1 is not significant (resampling $p = 0.157$), and significance appears as the intensity channel is weighted up ($0.157 \rightarrow 0.062 \rightarrow 0.012$ for $\lambda = 0 \rightarrow 0.5 \rightarrow 1$), so under this single subsample the difference is located in the intensity channel; Sect. 4.7 shows that after averaging over subsample realisations the gradient

across λ largely disappears, so this localization holds for one subsample only. A related sensitivity has the same character: the imputation rule for the 66.1 % of missing wind values moves the same p -value from 0.012–0.034 (Sect. 3.3).

Substituting a distance intrinsic to the space the data occupy for an ambient embedding is not a new action. Beshkov and Tiesinga (2022) replace the Euclidean metric by a geodesic one in the persistent homology of neural population activity and recover different features under the two. What differs here is that the intrinsic distance is fixed by the geometry the storms move on rather than estimated from the data, so the substitution removes a coordinate choice without introducing an estimated one.

A residual difference remains, of undetermined cause, and its survival is conditional on the metric. Table 5 gives it under four matched-null designs and three metrics.

Under W_1 the difference survives all four matchings, non-monotonically and informatively so: matching basin alone sharpens it, matching era alone weakens it the most, and the joint match sits between, pointing to era over basin as the principal confounder. Those twelve verdicts are one-sided (Sect. 3.5), and the convention decides two of them, both era-matched, so that under the more conservative reading the difference survives three of the four designs and fails in the one already weakest; Sect. S10 gives the companion values under each convention. Under the true W_2 and the exact W_∞ none of the twelve cells reaches α , though the event lies between the 68th and the 87th percentile of the control distribution in every W_∞ cell, and the era-matched design is the weakest of the four under both. The residual is therefore specification-conditional; it is not a metric-robust finding.

Three considerations preclude attributing that residual to solar activity even where it survives: it is confined to the intensity channel, being null for pure geometry; it is partly an era effect, the event storms’ greater recency and the era dependence of recorded intensity accounting for a substantial share; and at three events, three specific periods can differ from same-calendar controls for reasons – era and instrumental intensity trends, ENSO phase, the idiosyncratic meteorology of those windows – that no resampling can separate from a solar cause. One covariate is uncontrolled by all four designs: the fraction of fixes carrying an observed wind

Table 5. The great-circle $\lambda = 1$ event–control H_1 comparison under four matched-null designs and three metrics, each across three control-draw orderings, at $B = 1000$. W_1 is the summed matching cost, W_2 the true degree-2 distance, W_∞ the exact bottleneck. The twelve W_1 entries reproduce the archived values bit for bit, nine of them from a full re-run starting at the raw archive.

Matched-null design	W_1	true W_2	exact W_∞
Per-window (seasonal) count, 11/9/5	0.006/0.010/0.004	0.147/0.144/0.139	0.198/0.163/0.186
Genesis basin	0.003/0.004/0.003	0.087/0.107/0.102	0.133/0.149/0.138
Genesis era, 5 year bins	0.043/0.034/0.009	0.324/0.267/0.143	0.284/0.321/0.221
Joint basin $\times$ era	0.012/0.013/0.011	0.133/0.125/0.098	0.184/0.164/0.144

value is 41.3 % for the event set against a control-draw mean of 38.0 %, the 81st percentile of 1000 draws, though aggregated to the storm the two distributions are indistinguishable (Mann–Whitney $p = 0.993$). Because the draw randomises storms and not fixes, the axis was tested at that unit under the unmatched baseline draw and one of the four matched designs only (Sect. S8). We report the residual as a possibility motivating a larger-sample study.

Two quantities were held fixed throughout everything reported so far, and neither is addressed by the basin, era and ordering controls: the exponent of the matching cost, held at one, and the random subsample, drawn once for the event ensemble under a single seed while every control draw received its own. Section 4.6 and 4.7 vary both.

4.6 Where the difference sits, and why the exponent moves the verdict

Having established a calibrated verdict under the physically interpretable distance, we next ask what that verdict is made of, and where in the persistence diagram the difference sits. That question turns out to have a definite answer, and the answer settles which exponent has power against the difference.

4.6.1 The excess is at intermediate persistence

Each H_1 feature has a persistence, the difference between its death and its birth radius. Binning the event ensemble’s H_1 features by persistence and comparing the total in each bin against the 1000 control draws locates the excess; Table 6 gives the profile.

The excess is concentrated in the band $0.020 \leq \text{persistence} < 0.040$, where the event total is 2.5 times the control median and exceeds all but 4 of the 1000 control draws, while the single largest feature is unremarkable: the event’s maximum persistence is exceeded by 184 of the 1000 draws, the 82nd percentile. Under Eq. (1) that band corresponds to great-circle separations of $3.6\text{--}7.2^\circ$, equivalently to wind-speed differences of 4–8 kt, and two things limit the translation. The bin edges were chosen for readability and were not registered in advance: shifting them by half a bin moves the strongest bin to $0.060 \leq \text{persistence} < 0.120$ and lowers its percentile to 98.3, while an equal-log grid places it at $0.016 \leq \text{persistence} < 0.032$ at 98.1. And the excess is

much weaker in pure track geometry, the same band reaching the 94th percentile at $\lambda = 0$. What survives all three binnings is qualitative – the excess sits at intermediate persistence and not in the largest features – and the channel carrying it is intensity, with geometry alone insufficient.

4.6.2 Why the exponent changes the verdict

The two exponents aggregate the vector of matching costs differently – $\Sigma_i c_i$ against $(\Sigma_i c_i^2)^{1/2}$ – and, because the assignment is re-optimised at each exponent, they need not aggregate the same vector. Table S6 gives the measured decomposition. Squaring moves the weight toward the largest costs: the single most persistent feature takes 14.2 % of Σc_i but 48.7 % of Σc_i^2 , the four bins below persistence 0.04 fall together from 64.5 %–31.1 %, and the bin carrying the event excess drops from 26.5 %–17.9 %. The feature that gains is the one Sect. 4.6.1 found unremarkable. Rematching is small: the assignment changes for 21 %–35 % of matched pairs, yet holding the $p = 1$ assignment fixed and merely squaring its costs reproduces 99.2 %–99.5 % of the shift, so 0.5 %–0.8 % of the exponent effect is rematching and the rest is reweighting.

The exponent therefore acts on the noise. The ratio of the event distance to the mean control distance is close to 1.3 under both exponents, moving by 2.5 % in the basin-matched design, while the null standard deviation as a fraction of its mean grows from 9.8 %–18.6 %: the separation in those units falls from 2.93–1.07 and the resampling p -value moves from 0.003–0.087. The difference between event and control does not shrink, and the noise floor against which it is measured doubles.

This raises the question of how far the mechanism generalises: is a larger exponent always the blunter instrument, or does that depend on the representation? The direction is not universal, and this is where the representations divide cleanly. Under the linear encoding the event-to-control ratio rises with the exponent (1.212–1.262) and the p -value stays between 0.036 and 0.039, while under the other five it falls; with ten subsamples the linear encoding keeps a separation of 1.51–1.65 standard deviations across the range while the sphere decays from 1.76–0.52, the cylinder from 0.59 to –0.13 and the great-circle $\lambda = 1$ from 1.58–0.64. This follows Sect. 4.6.1: under the linear encoding the difference

Table 6. Persistence profile of the event H_1 excess under the great-circle distance at $\lambda = 1$, from the single archived event sub-sample: per-bin feature counts and total persistence for the event ensemble against the median of the 1000 control draws, with the event’s percentile in that control distribution. The decomposition of the matching cost that carries the excess is shown in Table S6. A positive control reconstructs the archived event diagram exactly (90 features, total persistence 0.9076). Boldface highlights the [0.020, 0.040) intermediate-persistence bin, in which the event total persistence is most extreme relative to the controls.

Persistence bin	Event features	Control median, count	Event total persistence	Control median, total persistence	Event percentile
[0, 0.005)	50	51	0.0902	0.0821	69.0
[0.005, 0.010)	14	12	0.0966	0.0858	65.7
[0.010, 0.020)	11	9	0.1523	0.1198	74.8
[0.020, 0.040)	11	4	0.3028	0.1188	99.6
[0.040, 0.080)	3	1	0.1487	0.0500	92.1
[0.080, 1.000)	1	0	0.1169	0.0000	80.9
Total	90	78	0.9076	0.5139	99.7

is present in the largest features as well, so squaring keeps it, and under the others it sits at intermediate scales, where squaring discards it.

4.6.3 A diagnostic available before the exponent is chosen

The change in width follows from how many matched costs actually contribute. For a statistic that sums weights $w_1, \dots, w_m$, define the effective number of contributing terms as $N_{\text{eff}} = (\sum_i w_i)^2 / \sum_i w_i^2$, which equals m when the weights are equal and approaches 1 when a single term dominates; setting $w_i = c_i$ gives the count for the summed statistic and $w_i = c_i^2$ that for the squared one. The functional itself is standard – the effective sample size of Kish (1965) and the participation ratio of condensed-matter physics – and we present it as such; what we add is the mapping to the matched costs of a persistence-diagram comparison. Both counts follow from the pair of diagrams, with no null distribution and no resampling required, so they are available before one commits to a single exponent for inference.

Were the contributions independent and of comparable relative variance, and were the matched costs the same under both exponents, replacing the summed statistic by the squared one would widen the null by $\sqrt{(N_1/N_2)}$. None of those three conditions holds exactly here, and the third fails measurably: re-optimising the assignment moves N_2 itself by 3 %–7 %. We therefore report $\sqrt{(N_1/N_2)}$ as an empirical envelope under those approximations, never as a derived limit. In Table 7 it exceeds the measurement in every one of the sixty subsample-by-representation combinations; on the median of ten subsamples the overshoot runs from 9 %–50 %, with a median of 24 %. Envelope and measurement co-vary, at a rank correlation of 0.83 on that median reading, with an exact permutation p of 0.029 over all 720 orderings. Section S8 records the archived-subsample figures, the bootstrap check behind the sixty-for-sixty count, and two checks that

the association is not a proxy for feature count or for the concentration of the persistence distribution alone.

Three limits bound what this establishes. The six specifications come from one dataset and one set of 1000 control draws, so the correlation describes consistency among six related quantities and not agreement across six independent cases. Used to estimate the size of the widening, the quantity is worse than the trivial baseline: its root-mean-square error across the six is 0.44, against a standard deviation of 0.30 in the measured widenings themselves. And the counts carry a scatter wide enough that the envelope for a single representation spans about 1.3–3.7, wider than the 1.91–2.60 spread between the six, so that four of them share the interval [2.36, 2.45]. No ranking of representations by this quantity follows, and what the six support is the direction of the widening together with an empirical envelope on its size. Two further caveats are in Sect. S8: it envelopes the widening of the null and not the shape of the p -value curve, and its counts come from event-to-reference costs while what they are compared against is a property of the control null.

4.7 A test that selects neither the exponent nor the subsample

Because the exponent changes the verdict, quoting one exponent understates the uncertainty of the conclusion. What follows is an exploratory robustness analysis; it is not a pre-registered confirmatory test: the family was assembled after the sensitivity had been seen, and how wide to draw it is itself a judgement. Section S9 reproduces the two pre-registration records that were deposited before their runs, and the specification matrix that separates what was registered from what was not.

Reporting a result across the set of defensible analyses rather than one chosen from it is established practice in other fields. Specification curve analysis (Simonsohn et al., 2020) and multiverse analysis (Steege et al., 2016) display the distribution of outcomes over the analytic choices, and PIMA

Table 7. The effective number of contributing matching costs under the two exponents, the envelope $\sqrt{(N_1/N_2)}$ their ratio implies under the approximations stated below, and the widening measured from the null distributions as the ratio of their coefficients of variation. Columns 4 and 5 recompute the envelope over ten event sub-samples in place of the single archived one.

Representation	N_{eff} summed/squared	Envelope, archived subsample	Median over ten subsamples	Range over ten	Measured widening
linear	29.9/8.1	1.92	1.873	1.70–2.17	1.25
unit-sphere	25.5/5.0	2.25	2.319	2.14–2.50	1.88
cylinder	17.1/2.5	2.60	2.496	2.36–2.60	1.97
great-circle $\lambda = 1$	24.1/4.3	2.36	2.505	2.33–2.55	2.00
great-circle $\lambda = 0.5$	16.5/3.0	2.34	2.387	2.23–2.45	2.09
great-circle $\lambda = 0$	8.7/2.4	1.91	1.979	1.90–2.05	1.81

Table 8. Resampling p -values under a single event subsample, the archived realisation, and after averaging over $M = 10$ independent subsamples, for the uniform and basin-matched designs. Separations in units of the null standard deviation are in parentheses. The $B = 10000$ recomputation of the basin-matched $M = 10$ cell is 0.0427 [0.0389, 0.0468]. At $M = 1$ the recomputation reproduces the archived event distances exactly. Boldface highlights the resampling p -values after averaging over ten subsamples ($M = 10$), as distinct from the archived single-subsample results ($M = 1$).

Design	$M = 1$ (archived)	$M = 10$
Unmatched D1	0.012 ($\sigma = 2.91$)	0.069 ($\sigma = 1.58$)
Basin-matched	0.003 ($\sigma = 3.18$)	0.044 ($\sigma = 1.87$)

(Girardi et al., 2024) supplies post-selection inference over that distribution, noting that the specification curve alone controls only the weak family-wise error rate. What follows carries that construction into ensemble topology, where the specifications are the matching exponent, the subsample realisation and the reference set.

4.7.1 Averaging the subsample realisation

The two unregistered quantities require different treatment. The exponent is a choice an analyst makes, and a choice can be tested over. The subsample is a Monte Carlo realisation, and averaging is the appropriate response; treating a realisation as an option would amount to selecting over randomness. We therefore recompute both sides symmetrically, averaging the statistic over $M = 10$ independent subsamples of the event ensemble and of each control draw, with the reference set held fixed. Table 8 gives the effect.

The change comes from the event side: the event statistic falls by 0.068 and 0.063 in the two designs while the null standard deviation falls only from 0.0474–0.0443 and from 0.0419–0.0377, subsampling accounting for 13 % and 19 % of the null variance. Averaging is unbiased, so a fall of this size means the single event subsample drawn under seed 2026 was a high one, and converting the six-seed scan of Table 1 back to distances confirms it: seed 2026 sits at 0.545

with the other five between 0.462 and 0.483, and the $M = 10$ average of 0.4773 lies among those five.

4.7.2 A test that adopts no exponent, and how wide to draw the family

The statistic is evaluated over a grid of seven exponents between 1 and 2, each standardised against the control draws, and the grid is averaged. The event and every control draw pass through the same transformation and the same average, so the combined statistic is calibrated by the resampling scheme that calibrates the individual tests, and multiplicity and dependence among family members are handled by that construction without an assumption of asymptotic independence; analytic combination rules for dependent p -values exist (Liu and Xie, 2020), and the resampling null makes one unnecessary here. The averaging rule is stated in advance: averaging asks whether the difference is present throughout the family, while the smallest p -value asks only whether some exponent reaches significance, so we report the average as the primary figure and the minimum alongside. The construction extends to the encoding and the intensity weight, because the 1000 control draws are the same storm sets in the same order for every representation and share one reference set.

Under a single subsample 12 of the 28 cells and 12 of the 42 reach $p < 0.05$; after averaging none does, the largest cell p -value being 0.363. All four widths of family are reported in Table 9, so that the dependence of the answer on that judgement is visible; adding the exact bottleneck gives twelve further values between 0.133 and 0.321 (Table 5), none below 0.05. Averaging leaves three descriptive findings standing – the linear representation flat across the exponent range against the other five decaying, the era-matched design weakest under the squared statistic and the bottleneck alike, and the directional consistency below – and removes one: under a single subsample the p -value falls monotonically with the intensity weight over a thirteenfold range, while after averaging the range is a factor of 1.75 in which the two weighted settings are no longer separable (1.54 against 1.58 standard deviations).

Table 9. Specification-averaged family p -values, under a single subsample and after $M = 10$ averaging, for four widths of family and two combination rules. Restricted to the single cell the minimum-based combination returns the archived resampling p -value exactly. Boldface highlights the three specification-averaged $M = 10$ p -values discussed in the Abstract and Conclusions (0.082, 0.122 and 0.134).

Family	Cells	Average, $M = 1$	Average, $M = 10$	Minimum, $M = 1$	Minimum, $M = 10$
$p = 1$, great-circle $\lambda = 1$ alone	1	0.012	0.069	–	–
Exponent, basin-matched	7	0.032	0.082	0.007	0.073
Exponent, unmatched	7	0.052	0.140	0.023	0.107
Exponent $\times$ encoding, unmatched	28	0.054	0.122	–	–
Exponent $\times$ encoding $\times \lambda$, unmatched	42	0.081	0.134	–	–

Reporting only whether each specification crosses 0.05 discards information. Across the 42 exponent-by-representation specifications the event lies above the control median in every one, at a median 86th percentile and a minimum of 63rd, and the event-to-control distance ratio exceeds one in 40 of the 42, ranging from 0.964–1.338. Those 42 values are strongly dependent – one dataset, one set of 1000 control draws and, at $M = 1$, one event subsample – so a count such as 42 of 42 describes consistency, is not a test, and cannot be converted into a p -value. The construction that does convert directional consistency into a single figure is the subsample averaging above, where the separation remains positive at $\sigma = 1.87$ and $\sigma = 1.58$ without reaching 1.96.

4.7.3 What this design does and does not resolve

The direction of the effect is consistent across every specification examined, and its magnitude does not reach the conventional threshold once neither the exponent nor the subsample is selected. That conclusion is one-sided (Sect. 3.5), and a two-sided reading only strengthens it, since the $M = 10$ family values of Table 9 already sit above α (Sect. S10). Under the basin-matched design the averaged separation is $\sigma = 1.87$ against the 1.96 required at the 5 % level.

Two readings of that outcome are available and this design cannot separate them. In the first a directionally consistent residual is present and the sample resolves it too coarsely to reach the threshold. In the second the design does not identify the target effect at all – a live possibility, because event eligibility depends in part on cyclone activity itself, so exposure is defined partly by outcome (Sect. 5.5). Nothing in the statistics distinguishes these, and no extrapolation from three event windows can: the three are not interchangeable units, holding 11, 9 and 5 storms, and each single-event cell is individually null at 0.248–0.674 (Sect. 4.3), so we have no basis for assuming that a further window would contribute a comparable increment of signal. We therefore report a design- and specification-limited directional residual, not a detection and not evidence of absence.

5 Discussion

5.1 The verdict cannot be read off one encoding, nor off one run

The three encodings are each published, each physically motivated, and each, applied alone, would have produced a paper reporting a different verdict on the same data. Reporting one of them as “the answer” without the others, and without the caveat that the answer is encoding-conditional, is not defensible. The stability theorem of Cohen-Steiner et al. (2007) does not rescue this: it bounds output change under a *fixed* input metric and says nothing about the choice of metric on a periodic coordinate, which is itself a modelling choice. Škraba and Turner (2020) sharpen the stability picture in the Wasserstein setting, which is where our statistic lives, and their bounds too are stated for a fixed exponent. The spread we report is of the kind Silberzahn et al. (2018) documented across analysts working independently on one dataset; here all three verdicts come from a single pipeline in which one parameter was fixed without being recorded, so the spread is reproducible from the archive, with no differences of judgement between teams involved.

This bears on what an encoding-specific result could mean. Re-encoding a periodic coordinate changes the pairwise distance and hence the input metric to the filtration, and nothing forces two embeddings of the same data to give the same diagram. An effect that appears under one embedding and not others is therefore consistent with two readings – a genuine physical coupling legible only in one geometry, or a metric-induced amplification with no physical content – and the topology alone cannot separate them. This is why encoding invariance is a *necessary* but not a *sufficient* condition for attribution: a physically interpretable distance removes the ambiguity of which geometry to trust, and even a difference that survives it requires independent, non-topological evidence before it can be called a coupling.

The stability results usually invoked to licence persistence as a robust summary do cover a change of the underlying space. Chazal et al. (2014) establish q -tameness for the Vietoris–Rips and Čech complexes of a totally bounded space, and bound the bottleneck distance between the dia-

grams of two such spaces by the Gromov–Hausdorff distance between them; Cohen–Steiner et al. (2007) bound the effect of perturbing a function on a fixed space. What neither bounds is a resampling verdict. The spread measured in Sect. 4.1 is a spread of p -values, and a bound on the distance between two diagrams does not by itself bound the difference between tail probabilities computed from them against a resampled null.

5.2 The H_1 fragility is a property of the sparse channel

The behaviour of Sect. 4.2 is a property of the sparse H_1 channel: the metric is excluded, since the exact bottleneck is if anything more order-sensitive, and so is the ambient dimension, since a 4-D linear control stays calibrated. A related aggregation mechanism sharpens the point. Because a persistence diagram is a global summary of the whole cloud, pooling subgroups can open loops at scales neither subgroup supports alone (Fig. 1). We label this *topological aggregation instability* as a working descriptive term, not a formal result; the cell-level null pattern of Sect. 4.3 is consistent with it, and the variance measurements of Sect. 4.2 give it a measured footing.

5.3 Metric, realisation, and what the diagnostic is for

Section 4.6 and 4.7 change what the repair of Sect. 4.5 can be said to have achieved. Fixing the input geometry on physical grounds removes one arbitrary choice, while leaving untouched three quantities downstream of it: how the mismatch between two lists of features is charged, which finite sample of points enters the diagram, and which reference diagrams the comparison is built against. On this dataset the first moves the event-level resampling p -value from 0.003–0.087, the second from 0.012–0.069, and the third, tested on the linear encoding alone, over 0.036–0.047 without crossing the threshold – sizes comparable to the effect being tested, which is the sense in which the repair is partial.

The reason the exponent matters here is legible in the diagram: the event–control difference sits at intermediate persistence, and squaring the matching cost spends the statistic’s sensitivity on the part of the diagram where the event is least distinctive and the resampling noise is largest. That is a statement about this dataset and not a general rule – under the linear encoding the difference is present in the largest features as well, and squaring helps it. The same trade-off is documented elsewhere. Bendich et al. (2016) found that in populations of brain artery trees the highest-persistence features had far less distinguishing power with respect to age than features of intermediate rank, and Adams and Moy (2021) draw on that case to argue that persistence length alone should not decide what counts as signal; since the degree- p Wasserstein distance sums the p th powers of the matched costs (Cohen–Steiner et al., 2010), raising p concentrates the statistic on the largest of them. The parallel

in high-dimensional testing is explicit: extreme-value statistics are the more powerful choice when a few components carry the effect, sums across components are used against effects that are individually weak and spread over many (Fan et al., 2015), and maximum-based procedures attain the optimal detection boundary only over part of the sparse range (Donoho and Jin, 2004). Which of the two situations obtains can be read off the diagrams before the exponent is committed to, and the effective count of Sect. 4.6.3 is one way of reading it: it envelopes the widening of the null in all sixty combinations examined, while as an estimate of the size of that widening it does not improve on a constant, and its subsampling scatter is wide enough that no ranking of representations follows. Its practical use is as a warning before the fact, never as a calibrated forecast.

5.4 What follows for the solar–cyclone question

Under the arbitrary encodings the pipeline does not give a single-valued verdict, so it cannot by itself support a positive or negative substantive claim at the ensemble-topology scope. Under the physically interpretable distance, the representation-level ambiguity is removed, and a residual event–control H_1 difference survives all four matched nulls under W_1 , three of the four under the more conservative two-sided convention (Sect. S10) – but that residual is confined to the intensity channel, is attenuated by era matching, is absent under the true W_2 and the exact bottleneck, and cannot be attributed to solar activity at three events. The cleanest defensible statement is therefore a cautious one: the ensemble-topology scope is not yet a tool that settles the solar–cyclone question, and the one representation that yields a calibrated difference surfaces a possibility that a larger-sample study should test. This non-detection at the ensemble-topology scope is consistent with both the older positive per-storm literature and the null microphysical literature; the three scopes interrogate different signatures, and a non-detection at one is not evidence either way at another. Causal isolation would require a design with many more events and an intensity- and era-controlled null – for example a synthetic-event injection with known perturbation strength – identified here as the priority follow-up and beyond the scope of the present study.

5.5 A structural limit on what this design can test

A second limitation is structural rather than statistical, and it bounds what this design can ever be used for. The three events were not selected on geomagnetic magnitude alone: the inclusion criteria of Sect. S5 also require sufficient cross-basin tropical-cyclone activity in the window, one benchmark per recent solar cycle, and ± 2 year independence. Cyclone activity is part of the outcome under study, so event eligibility depends in part on the very quantity the test examines. For a methodological benchmark this is harmless and indeed convenient, since one wants windows with enough data to

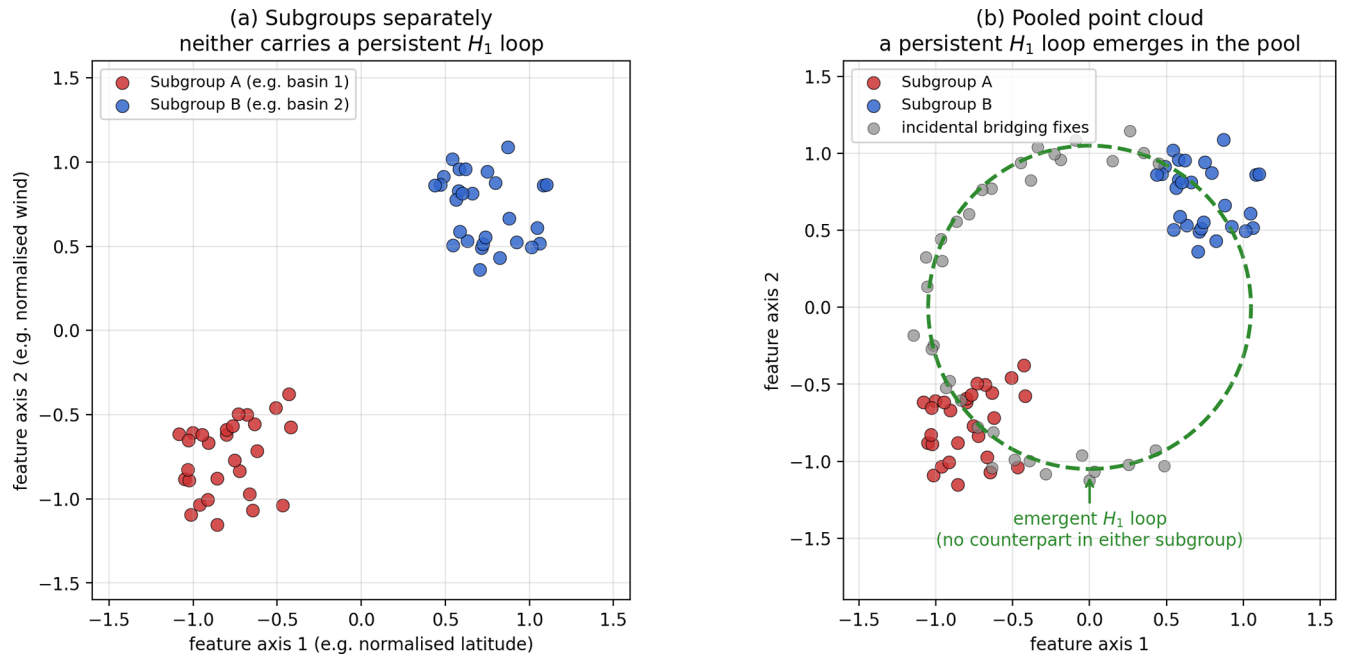


Figure 1. Schematic of topological aggregation instability. Left: two subgroups, neither carrying a persistent H_1 loop alone. Right: the pooled cloud, in which a shared Vietoris–Rips filtration generates an H_1 loop spanning the gap – a feature with no counterpart in either subgroup, that disappears under disaggregation. Pooling is not a linear operation on persistence diagrams.

exercise the pipeline. For a test of solar or geomagnetic forcing it is not: a design in which exposure is defined partly by outcome cannot establish that direction of influence, however the statistics come out. Any physical test of the solar–cyclone question must therefore define its exposure from geomagnetic and solar-wind indices alone, before any cyclone data is consulted.

5.6 A minimum protocol, and remedies attempted

We propose five pre-registration requirements.

1. *Representation.* Report any positive H_1 claim under at least two principled representations, preferring a physically interpretable distance wherever one exists. Where a periodic coordinate is unavoidable, reduce the source archive to a single convention and register the position of the branch cut.
2. *Matching exponent.* Register the exponent of the matching cost, or report a statistic that selects none.
3. *Subsample realisation.* Average it, never select it, over $M \geq 10$ for the event and for every control draw alike.
4. *Reference set.* Register it alongside the exponent and the subsample.
5. *Control-draw ordering and Monte-Carlo error.* Fix the ordering by an explicit sort key, never by a pinned interpreter hash seed. Use enough placebos ($n \geq 200$) and

null draws that the Monte-Carlo interval on a p -value is narrow relative to α , and report that interval alongside the estimate, separately for H_0 and H_1 .

Varying the ordering is a diagnostic check on requirement 5 rather than a reporting format, since reordering a uniformly resampled control pool selects a different realisation while the estimand does not move (Sect. 3.5). A template implementing these, with an operational agreement rule and a stratified-attribution requirement guarding against pooled-significant-but-stratified-null artefacts of the Simpson (1951) family, is shown in Sect. S3.

The present analysis does not fully meet its own protocol: the placebo set is $n = 49$, below requirement 5's $n \geq 200$, and each placebo uses $B = 100$ null draws, which inflates the ordering spread. Neither changes the qualitative conclusions, though the verdict is unstable across event-subsample realisations, as Sect. 4.7.1 and Table 1 show. The two statistic-level remedies a follow-up would try first do not help (Sect. 4.2); the productive move was the physically interpretable distance of Sect. 4.5, which removes the representational ambiguity that those remedies leave in place. Persistence images with kernel methods (Adams et al., 2017; Reininghaus et al., 2015), larger event samples, and non-Vietoris–Rips frameworks remain open directions.

6 Conclusions

Testing whether three benchmark extreme geomagnetic events perturb cyclone-activity topology in their ± 15 d windows, we find three principled longitude encodings giving three reproducible verdicts on identical data (H_1 resampling $p = 0.108, 0.021$ and 0.156), while the 49-placebo calibration is nominal in H_1 at every auditable control-draw ordering.

A great-circle input distance removes the ambiguity of which geometry to trust and returns a single value, and under the summed matching cost the event–control difference survives all four matched nulls one-sided, and three of the four two-sided under the conservative doubling convention. That repair is partial in a way the encodings had concealed: the same difference moves with the exponent of the matching cost (0.003 – 0.087), with which 600 of the ensemble’s fixes enter the diagram (0.012 – 0.069), and with which fifty control diagrams form the reference set (0.036 – 0.047 , tested on the linear encoding alone and crossing no threshold), and it is absent under the true degree-2 and the exact bottleneck distance. The reason is legible in the diagram: the excess sits at intermediate persistence while the single most persistent feature is at the 82nd percentile of its null, so squaring the matching cost spends the statistic’s sensitivity where the event is least distinctive, doubling the noise floor while moving the event-to-control ratio by 2.5 %. The effective number of contributing matching costs, computable from the diagrams before an exponent is committed to, envelopes that widening in all sixty combinations examined, by a median of 24 %.

Selecting neither the exponent nor the subsample gives a specification-averaged resampling p -value of 0.082 over exponents under the basin-matched design, and 0.122 and 0.134 over the wider unmatched families, with the event above the control median in all 42 cells and the averaged separation at 1.87 standard deviations. The substantive outcome is a design- and specification-limited directional residual: no detection, and no evidence of absence. The design cannot separate a residual too coarsely resolved from a design that does not identify the effect at all, because event eligibility depends in part on cyclone activity itself (Sect. 5.5). Pooled point-cloud topology on geospatial trajectories should therefore prefer a physically interpretable input distance, register the exponent of its matching cost and its reference set or report a statistic that selects neither, average over subsample realisations instead of quoting one, and fix the control-draw ordering by an explicit sort key while reporting the Monte-Carlo uncertainty of the resulting p -value in both homology channels.

Code and data availability. IBTrACS v04r01 is publicly available at <https://www.ncei.noaa.gov/products/international-best-track-archive> (last access: 30 April 2026).

The analysis code – a single pipeline module with the three longitude encodings and the great-circle distance as switchable variants, the placebo, matched-null, bottleneck, landscape and subsample drivers, and the specification record and verification script of Sect. 3.7 – together with the run logs and derived tables, is archived at Zenodo (<https://doi.org/10.5281/zenodo.22333475>, Dai, 2026) under the MIT licence. The archive includes the verification report that re-derives all 87 registered results and compares each with the value printed here, and the frozen outputs of the superseded runs discussed in Sects. 3.3.4 and S1, so that the corrections described there can be inspected rather than taken on trust. Run-log paths quoted in this paper are relative to the `v7/directory` of that archive.

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/npg-33-521-2026-supplement>.

Competing interests. The author has declared that there are no competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. An AI system (Anthropic Claude) was used as a coding and writing assistant; all design decisions, statistical interpretations, and conclusions are the author’s. IBTrACS data are provided by NOAA NCEI.

Review statement. This paper was edited by Wansuo Duan and reviewed by three anonymous referees.

References

- Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., and Ziegelmeier, L.: Persistence images: a stable vector representation of persistent homology, *J. Mach. Learn. Res.*, 18, 1–35, 2017.
- Adams, H. and Moy, M.: Topology applied to machine learning: from global to local, *Front. Artif. Intel. Ap.*, 4, 668302, <https://doi.org/10.3389/frai.2021.668302>, 2021.
- Bauer, U.: Ripser: efficient computation of Vietoris–Rips persistence barcodes, *J. Appl. Comput. Topol.*, 5, 391–423, <https://doi.org/10.1007/s41468-021-00071-5>, 2021.
- Bendich, P., Marron, J. S., Miller, E., Pieloch, A., and Skwerer, S.: Persistent homology analysis of brain artery trees, *Ann. Appl. Stat.*, 10, 198–218, <https://doi.org/10.1214/15-AOAS886>, 2016.

- Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. Roy. Stat. Soc. B*, 57, 289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>, 1995.
- Beshkov, K. and Tiesinga, P.: Geodesic-based distance reveals nonlinear topological features in neural activity from mouse visual cortex, *Biol. Cybern.*, 116, 53–68, <https://doi.org/10.1007/s00422-021-00906-5>, 2022.
- Bubenik, P.: Statistical topological data analysis using persistence landscapes, *J. Mach. Learn. Res.*, 16, 77–102, 2015.
- Calogovic, J., Albert, C., Arnold, F., Beer, J., Desorgher, L., and Flueckiger, E. O.: Sudden cosmic ray decreases: no change of global cloud cover, *Geophys. Res. Lett.*, 37, L03802, <https://doi.org/10.1029/2009GL041327>, 2010.
- Carlsson, G.: Topology and data, *B. Am. Math. Soc.*, 46, 255–308, <https://doi.org/10.1090/S0273-0979-09-01249-X>, 2009.
- Chazal, F., de Silva, V., and Oudot, S.: Persistence stability for geometric complexes, *Geometriae Dedicata*, 173, 193–214, <https://doi.org/10.1007/s10711-013-9937-z>, 2014.
- Cohen-Steiner, D., Edelsbrunner, H., and Harer, J.: Stability of persistence diagrams, *Discrete Comput. Geom.*, 37, 103–120, <https://doi.org/10.1007/s00454-006-1276-5>, 2007.
- Cohen-Steiner, D., Edelsbrunner, H., Harer, J., and Mileyko, Y.: Lipschitz functions have L^p -stable persistence, *Found. Comput. Math.*, 10, 127–139, <https://doi.org/10.1007/s10208-010-9060-6>, 2010.
- Dai, R.: Analysis code and outputs for “A physically grounded input distance resolves longitude-encoding ambiguity but not specification dependence in ensemble persistent homology of cyclone trajectories”, Zenodo [code], <https://doi.org/10.5281/zenodo.22333475>, 2026.
- Donoho, D. and Jin, J.: Higher criticism for detecting sparse heterogeneous mixtures, *Ann. Stat.*, 32, 962–994, <https://doi.org/10.1214/009053604000000265>, 2004.
- Edelsbrunner, H. and Harer, J. L.: Computational Topology: An Introduction, American Mathematical Society, Providence, RI, <https://doi.org/10.1090/mbk/069>, 2010.
- Elsner, J. B. and Jagger, T. H.: United States and Caribbean tropical cyclone activity related to the solar cycle, *Geophys. Res. Lett.*, 35, L18705, <https://doi.org/10.1029/2008GL034431>, 2008.
- Elsner, J. B. and Kavnikov, S. P.: Hurricane intensity changes associated with geomagnetic variation, *Atmos. Sci. Lett.*, 2, 86–93, <https://doi.org/10.1006/asle.2001.0043>, 2001.
- Fan, J., Liao, Y., and Yao, J.: Power enhancement in high-dimensional cross-sectional tests, *Econometrica*, 83, 1497–1541, <https://doi.org/10.3982/ECTA12749>, 2015.
- Ghil, M. and Sciamarella, D.: Review article: Dynamical systems, algebraic topology and the climate sciences, *Nonlin. Processes Geophys.*, 30, 399–434, <https://doi.org/10.5194/npg-30-399-2023>, 2023.
- Ghrist, R.: Barcodes: the persistent topology of data, *B. Am. Math. Soc.*, 45, 61–75, <https://doi.org/10.1090/S0273-0979-07-01191-3>, 2008.
- Girardi, P., Vesely, A., Lakens, D., Altoè, G., Pastore, M., Calcagni, A., and Finos, L.: Post-selection inference in multiverse analysis (PIMA): an inferential framework based on the sign flipping score test, *Psychometrika*, 89, 542–568, <https://doi.org/10.1007/s11336-024-09973-6>, 2024.
- Hodges, R. E. and Elsner, J. B.: Evidence linking solar variability with US hurricanes, *Int. J. Climatol.*, 31, 1897–1907, <https://doi.org/10.1002/joc.2196>, 2011.
- Hodges, R. E., Jagger, T. H., and Elsner, J. B.: The sun–hurricane connection: diagnosing the solar impacts on hurricane frequency over the North Atlantic basin using a space–time model, *Nat. Hazards*, 73, 1063–1084, <https://doi.org/10.1007/s11069-014-1120-9>, 2014.
- Huth, R. and Stryhal, J.: Association of the 11 year solar cycle with correlation and teleconnection structures in tropospheric circulation, *J. Space Weather Spac.*, 14, 31, <https://doi.org/10.1051/swsc/2024031>, 2024.
- Inatsu, M., Kato, H., Katsuyama, Y., Hiraoka, Y., and Ohbayashi, I.: A cyclone identification algorithm with persistent homology and merge-tree, *SOLA*, 13, 214–218, <https://doi.org/10.2151/sola.2017-039>, 2017.
- IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Chapt. 7, Cambridge University Press, Cambridge, UK, <https://doi.org/10.1017/9781009157896>, 2021.
- Islambekov, U. and Pathirana, H.: Vector summaries of persistence diagrams for permutation-based hypothesis testing, *Found. Data Sci.*, 6, 41–61, <https://doi.org/10.3934/fods.2024002>, 2024.
- Kirkby, J., Curtius, J., Almeida, J., Dunne, E., Duplissy, J., Ehrhart, S., Franchin, A., Gagné, S., Ickes, L., Kürten, A., Kupc, A., Metzger, A., Riccobono, F., Rondo, L., Schobesberger, S., Tsagkogeorgas, G., Wimmer, D., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Downard, A., Ehn, M., Flagan, R. C., Haider, S., Hansel, A., Hauser, D., Jud, W., Junninen, H., Kreissl, F., Kvashin, A., Laaksonen, A., Lehtipalo, K., Lima, J., Lovejoy, E. R., Makhmutov, V., Mathot, S., Mikkilä, J., Minginette, P., Mogo, S., Nieminen, T., Onnela, A., Pereira, P., Petäjä, T., Schnitzhofer, R., Seinfeld, J. H., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Vanhanen, J., Viisanen, Y., Vrtala, A., Wagner, P. E., Walther, H., Weingartner, E., Wex, H., Winkler, P. M., Carslaw, K. S., Worsnop, D. R., Baltensperger, U., and Kulmala, M.: Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation, *Nature*, 476, 429–433, <https://doi.org/10.1038/nature10343>, 2011.
- Kish, L.: Survey Sampling, John Wiley and Sons, New York, ISBN 978-0-471-48900-9, 1965.
- Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The International Best Track Archive for Climate Stewardship (IBTrACS), *B. Am. Meteorol. Soc.*, 91, 363–376, <https://doi.org/10.1175/2009BAMS2755.1>, 2010.
- Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., and Schreck, C. J.: International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4, NOAA National Centers for Environmental Information [data set], <https://doi.org/10.25921/82ty-9e16>, 2019.
- Kristjánsson, J. E., Stjern, C. W., Stordal, F., Fjæraa, A. M., Myhre, G., and Jónasson, K.: Cosmic rays, cloud condensation nuclei and clouds – a reassessment using MODIS data, *Atmos. Chem. Phys.*, 8, 7373–7387, <https://doi.org/10.5194/acp-8-7373-2008>, 2008.

- Laken, B. A., Pallé, E., Calogovic, J., and Dunne, E. M.: A cosmic ray – climate link and cloud observations, *J. Space Weather Spac.*, 2, A18, <https://doi.org/10.1051/swsc/2012018>, 2012.
- Li, D., Xiao, Z., Xu, J., and Zhao, L.: Asymmetric modulation of solar activity on tropical cyclone frequency over the western North Pacific, *Front. Earth Sci.*, 11, 1169685, <https://doi.org/10.3389/feart.2023.1169685>, 2023.
- Liu, Y. and Xie, J.: Cauchy combination test: a powerful test with analytic p -value calculation under arbitrary dependency structures, *J. Am. Stat. Assoc.*, 115, 393–402, <https://doi.org/10.1080/01621459.2018.1554485>, 2020.
- Marsh, N. D. and Svensmark, H.: Low cloud properties influenced by cosmic rays, *Phys. Rev. Lett.*, 85, 5004–5007, <https://doi.org/10.1103/PhysRevLett.85.5004>, 2000.
- Mendoza, B. and Pazos, M.: A 22-yr hurricane cycle and its relation with geomagnetic activity, *J. Atmos. Sol.-Terr. Phys.*, 71, 2047–2054, <https://doi.org/10.1016/j.jastp.2009.09.012>, 2009.
- Mileyko, Y., Mukherjee, S., and Harer, J.: Probability measures on the space of persistence diagrams, *Inverse Probl.*, 27, 124007, <https://doi.org/10.1088/0266-5611/27/12/124007>, 2011.
- Muszynski, G., Kashinath, K., Kurlin, V., Wehner, M., and Prabhat: Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets, *Geosci. Model Dev.*, 12, 613–628, <https://doi.org/10.5194/gmd-12-613-2019>, 2019.
- Nayak, S.: Effects of solar variability on tropical cyclone activity, *Earth Space Sci.*, 11, e2023EA003500, <https://doi.org/10.1029/2023EA003500>, 2024.
- Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., and Harrington, H. A.: A roadmap for the computation of persistent homology, *EPJ Data Sci.*, 6, 17, <https://doi.org/10.1140/epjds/s13688-017-0109-5>, 2017.
- Pérez-Peraza, J., Velasco, V., and Kavnikov, S.: Wavelet coherence analysis of Atlantic hurricanes and cosmic rays, *Geofís. Int.*, 47, 231–244, <https://doi.org/10.22201/igeof.00167169p.2008.47.3.85>, 2008.
- Pierce, J. R.: Cosmic rays, aerosols, clouds, and climate: recent findings from the CLOUD experiment, *J. Geophys. Res.-Atmos.*, 122, 8051–8055, <https://doi.org/10.1002/2017JD027475>, 2017.
- Prikryl, P., Nikitina, L., and Rušin, V.: Rapid intensification of tropical cyclones in the context of the solar wind–magnetosphere–ionosphere–atmosphere coupling, *J. Atmos. Sol.-Terr. Phys.*, 183, 36–60, <https://doi.org/10.1016/j.jastp.2018.12.009>, 2019.
- Reininghaus, J., Huber, S., Bauer, U., and Kwitt, R.: A stable multi-scale kernel for topological machine learning, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 7–12 June 2015, 4741–4748, <https://doi.org/10.1109/CVPR.2015.7299106>, 2015.
- Robinson, A. and Turner, K.: Hypothesis testing for topological data analysis, *J. Appl. Comput. Topol.*, 1, 241–261, <https://doi.org/10.1007/s41468-017-0008-7>, 2017.
- Saul, N. and Tralie, C.: Scikit-TDA: topological data analysis for Python, *Zenodo*, <https://doi.org/10.5281/zenodo.2533369>, 2019 (also available at: <https://scikit-tda.org/>, last access: 16 October 2026).
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., Fong, N., Gamez-Djokic, M., Glenz, A., Gordon-McKeon, S., Heaton, T. J., Hederos, K., Heene, M., Hofelich Mohr, A. J., Högden, F., Hui, K., Johannesson, M., Kalodimos, J., Kaszubowski, E., Kennedy, D. M., Lei, R., Lindsay, T. A., Liverani, S., Madan, C. R., Molden, D., Molleman, E., Morey, R. D., Mulder, L. B., Nijstad, B. R., Pope, N. G., Pope, B., Prenoveau, J. M., Rink, F., Robusto, E., Roderique, H., Sandberg, A., Schlüter, E., Schönbrodt, F. D., Sherman, M. F., Sommer, S. A., Sotak, K., Spain, S., Spörlein, C., Stafford, T., Stefanutti, L., Tauber, S., Ullrich, J., Vianello, M., Wagenmakers, E.-J., Witkowiak, M., Yoon, S., and Nosek, B. A.: Many analysts, one data set: making transparent how variations in analytic choices affect results, *Adv. Meth. Pract. Psych.*, 1, 337–356, <https://doi.org/10.1177/2515245917747646>, 2018.
- Simonsohn, U., Simmons, J. P., and Nelson, L. D.: Specification curve analysis, *Nat. Hum. Behav.*, 4, 1208–1214, <https://doi.org/10.1038/s41562-020-0912-z>, 2020.
- Simpson, E. H.: The interpretation of interaction in contingency tables, *J. Roy. Stat. Soc. B*, 13, 238–241, <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>, 1951.
- Škraba, P. and Turner, K.: Wasserstein stability for persistence diagrams, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2006.16824>, 2020.
- Steenen, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W.: Increasing transparency through a multiverse analysis, *Perspect. Psychol. Sci.*, 11, 702–712, <https://doi.org/10.1177/1745691616658637>, 2016.
- Strommen, K., Chantry, M., Dorrington, J., and Otter, N.: A topological perspective on weather regimes, *Clim. Dynam.*, 60, 1415–1445, <https://doi.org/10.1007/s00382-022-06395-x>, 2022.
- Svensmark, H.: Cosmoclimatology: a new theory emerges, *Astron. Geophys.*, 48, 1.18–1.24, <https://doi.org/10.1111/j.1468-4004.2007.48118.x>, 2007.
- Svensmark, J., Enghoff, M. B., Shaviv, N. J., and Svensmark, H.: The response of clouds and aerosols to cosmic ray decreases, *J. Geophys. Res.-Space*, 121, 8152–8181, <https://doi.org/10.1002/2016JA022689>, 2016.
- The GUDHI Project: GUDHI User and Reference Manual, GUDHI Editorial Board, <https://gudhi.inria.fr/> (last access: 16 October 2026), 2015.
- Tymochko, S., Munch, E., Dunion, J., Corbosiero, K., and Torn, R.: Using persistent homology to quantify a diurnal cycle in Hurricane Felix, *Pattern Recogn. Lett.*, 133, 137–143, <https://doi.org/10.1016/j.patrec.2020.02.022>, 2020.
- Ver Hoef, L., Adams, H., King, E. J., and Ebert-Uphoff, I.: A primer on topological data analysis to support image analysis tasks in environmental science, *Artif. Intell. Earth Syst.*, 2, e220039, <https://doi.org/10.1175/AIES-D-22-0039.1>, 2023.
- Wasserman, L.: Topological data analysis, *Annu. Rev. Stat. Appl.*, 5, 501–532, <https://doi.org/10.1146/annurev-statistics-031017-100045>, 2018.
- Wilson, E. B.: Probable inference, the law of succession, and statistical inference, *J. Am. Stat. Assoc.*, 22, 209–212, <https://doi.org/10.1080/01621459.1927.10502953>, 1927.
- Yu, F. and Luo, G.: Effect of solar variations on particle formation and cloud condensation nuclei, *Environ. Res. Lett.*, 9, 045004, <https://doi.org/10.1088/1748-9326/9/4/045004>, 2014.