

Marked-up version — egusphere-2026-2975

Revised title:

A physically grounded input distance fixes an ensemble persistent-homology test's geometry, and what that repair leaves untouched

Title of the submitted discussion paper:

Encoding-dependent verdicts and H_1 miscalibration in ensemble persistent homology of cyclone trajectories

Rongzhen Dai — Independent Researcher, Yangzhou, China

How to read this document. ~~Red-strike-through~~ marks text removed from the submitted preprint (egusphere-2026-2975, 23 June 2026); blue underline marks text added in the revision. Unmarked text is carried over verbatim.

This revision is close to a rewrite. Of the 13,340 words in the revised manuscript, 1,812 (13.6%) are carried over verbatim; 8,082 words were removed and 11,528 added, across 693 separate changes. The title, the abstract, the statement of the contribution and the whole of the great-circle analysis are new. A word-level comparison of a revision this extensive is necessarily dense, so it is offered for completeness rather than as the primary guide: **the substantive changes are set out point by point, each with a page and line number, in the author's response** accompanying this submission.

The figure was also replaced; both versions are reproduced below at the point where they occur. Table rows are collapsed to [table row] so that re-typeset tables do not swamp the comparison; the tables themselves appear in the revised manuscript.

~~Encoding-dependent verdicts and H_1 miscalibration~~ A physically grounded input distance resolves longitude-encoding ambiguity but not specification dependence in ensemble persistent homology of cyclone trajectories ~~**Rongzhen Dai¹**~~ ¹ Independent Researcher, Yangzhou, China. Correspondence: dairongzhen3@gmail.com —

Abstract ~~Ensemble topological data analysis (TDA) on multi-storm cyclone trajectories has been proposed as a tool to detect coherent perturbations — such as those associated with extreme~~ 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 -day windows of three benchmark geomagnetic events — that may not register in any single-storm intensity time series. We construct storms (25 event storms, 1,020 calendar-matched controls), three principled longitude encodings give three reproducible verdicts: one-sided H_1 resampling $p = 0.108$ for the cyclone point cloud (linear modular, unit sphere, and lat-linear-plus-longitude-circle cylinder) and apply the same ensemble-TDA pipeline to the same data: all storms whose lifetime overlaps the ± 15 -day peak window of Halloween 2003, St Patrick's Day 2015, or Gannon 2024 (25 event storms; 1,020 calendar-matched controls). Three findings emerge. First, the D1 pool H_1 permutation p -value depends a linear modular encoding on encoding choice in a way that flips the qualitative verdict: perm- $p_{H_1} = 0.130$ (linear), 0.009 (unit sphere), 0.214 (cylinder), on identical event and control sets. Second, single longitude convention, 0.021 for a unit sphere, 0.156 for a cylinder. Across six auditable control-draw orderings a 49-placebo calibration returns an H_0 false-positive rate close is nominal in H_1 throughout (9 to 12 of 147) while H_0 runs 4 to 16, crossing the nominal 5% 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 three encodings (8.2%, 4.1%, 6.1%), but an H_1 false-positive rate that is consistently above nominal in all three (8.2%, 10.2%, 14.3%; 42 cells. We report a design- and specification-limited directional consistency across encodings, not individually significant at $n = 49$ — the cylinder rate sits at the Wilson 95% upper bound) — H_1 calibration is inflated by a factor 1.6 $\times$ –2.9 $\times$ regardless residual: no detection, and no evidence of encoding choice. Third, this dimension asymmetry between H_0 (calibrated) and H_1 (inflated) recurs in stratified attribution across two basin cells and three single-event cells, and in subsample-size sensitivity tests at $N \in \{400, 600, 800, 1000\}$. We interpret the pattern as observational evidence that ensemble TDA on cyclone-trajectory data carries an intrinsic H_1 inflation that is not curable by lon-encoding choice alone. The solar-perturbation hypothesis cannot be tested through this pipeline until an encoding-invariant H_1 -calibrated protocol is in place. We frame the contribution as a methodological cautionary tale: pooled point-cloud TDA on geospatial trajectories with periodic coordinates is more fragile than its formal stability theorems suggest. We propose a minimum protocol of multi-encoding agreement testing plus dimension-resolved placebo calibration as a precondition for any positive ensemble-TDA claim in this class of problems: absence. —

1. Introduction

1.1 ~~Solar-cyclone hypothesis: a debated observational lineage~~ The hypothesis that solar-cyclone question, in brief Whether solar and geomagnetic activity perturbs tropical cyclone (TC) dynamics has a long and contested observational lineage. Elsner & ~~and~~ Kavvakov (2001) ~~brought the question to peer-reviewed satellite-era cyclone data.~~ They reported correlations between ~~Kp / Dst~~ geomagnetic indices and per-storm intensity ~~changes in the North Atlantic.~~ changes; Elsner & ~~and~~ Jagger (2008) found that linked US and Caribbean hurricane counts vary inversely with to sunspot number, and ~~proposed a heat-engine pathway in which an active Sun warms the upper troposphere via ozone-mediated ultraviolet absorption, modulating a storm's maximum potential intensity.~~ number; Pérez-Peraza et al. (2008) (2008), Mendoza and Mendoza & Pazos (2009) identified a 22-year cycle in Atlantic (2009), Hodges and Pacific hurricane occurrences that coincides with the magnetic solar cycle. Hodges & Elsner (2011) confirmed a and Hodges et al. (2014) reported solar-cycle dependence for US hurricanes over 1866–2008. Hodges, Jagger & Elsner (2014) refined the analysis with a Bayesian space-time model that resolves the modulation geographically. More recent work extends the analysis to other basins: dependences in various basins, and Li et al. (2023) report an asymmetric modulation of solar activity on TC frequency over the western North Pacific, and Nayak (2024) provides a contemporary review. 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, Ch. 7) reflects the consensus that galactic cosmic rays contribute negligible radiative forcing. Across ~~the lineage, reported~~ two decades effect sizes are modest, ~~proposed~~ mechanisms vary, and ~~null-hypothesis null~~ controls are unevenly applied.

1.2 ~~Mechanistic candidates: cosmic-ray-cloud and atmospheric ionization~~ Several physical mechanisms have been proposed to underwrite the observed correlations. Most prominent is the cosmic-ray-cloud (CRC) hypothesis. Heliospheric modulation of the galactic cosmic ray (GCR) flux is associated with variation in the rate of ion-induced new-particle formation in the lower atmosphere, and so with the abundance of cloud condensation nuclei (CCN). Marsh & Svensmark (2000) reported a correlation between GCR flux and low cloud cover during one solar cycle. Svensmark (2007) developed the finding into the broader cosmoecology programme. Laboratory work has tested the ion-induced nucleation step directly. The CLOUD experiment at CERN (Kirkby et al., 2011) showed that GCR-relevant ionization enhances nucleation rates of sulphuric-acid clusters under controlled conditions. The magnitude is sensitive to the presence of stabilising species such as ammonia. Numerical modelling by Yu & Luo (2014) explored how solar-cycle GCR variation translates to CCN concentration. They found a response larger than earlier estimates but still small relative to observed climate variability. Svensmark et al. (2016) used Forbush-decrease compositing to argue for a detectable cloud response, although the compositing scheme has itself been contested. Mechanistic candidates outside the CRC framework include direct geomagnetic-induced perturbation of mesospheric and stratospheric circulation. Solar-cycle signatures may propagate downward to the troposphere via wave-mean-flow interaction (Huth & Stryhal, 2024). Ionospheric coupling provides another pathway: solar-wind energy input is associated with auroral gravity waves that descend to the tropical troposphere and may co-vary with convective instability associated with rapid intensification (Mironova et al., 2018).

1.3 ~~Counter-evidence: the null literature~~ Against this body of suggestive correlations stands an equally substantial literature finding no robust effect. Kristjánsson et al. (2008) examined twenty-two Forbush-decrease events over MODIS-instrumented oceanic regions. They found no statistically significant response in cloud parameters. Calogovic et al. (2010) reached the same conclusion using a different compositing scheme. Laken et al. (2012) reviewed thirty-five years of satellite cloud observations alongside daily-timescale Forbush analyses. They concluded that evidence for a widespread GCR-cloud link remains elusive, with most apparent variability attributable to ENSO. Pierce (2017), drawing on findings from the CLOUD experiment, argued that ion-induced enhancement of CCN is too weak to translate into detectable cloud-radiation forcing at climate-relevant scales. The IPCC Sixth Assessment Report (IPCC AR6 WGI, 2021, Ch. 7.3.4.5) reflects the consensus position with high confidence: GCRs contribute a negligible effective radiative forcing over 1750–2019, and no robust association between GCRs and cloudiness has been established. The result is a literature in which positive correlation reports and null-result analyses coexist over two decades of investigation. We treat the question of whether solar as open, and geomagnetic activity perturbs cyclone dynamics use it here as the setting for a genuinely debated methodological one.

1.4 ~~Scope of the present work: encoding invariance as~~ 1.2 The problem: a precondition verdict that depends on unregistered choices Existing observational studies of the solar-cyclone question almost universally operate at the per-storm scale, asking whether a single storm's intensity time series correlates with concurrent geomagnetic indices. The scope choice matters. An alternative, scale. A more recently proposed scope applies persistent homology to multi-storm point clouds treats all the cyclones active in one time window as a single cloud of points in (latitude, longitude, intensity) coordinates over all TCs occurring within an event window, with 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 natural geometric object becoming radius at which each hole in the joint point cloud of all event-period fixes resulting union opens and the natural

~~summary statistic becoming its persistence diagram:~~ radius at which it closes, and summarises the cloud by the list of those (birth, death) pairs. A ~~coherent perturbation hypothesis, if true, might~~ worked example makes the idea concrete. Six storms whose tracks happen to encircle a quiet region of ocean leave a ~~signature on this~~ 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 ~~topology that is not visible~~ has and how long they survive, in ~~any~~ a way no single storm's ~~trajectory:~~ 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 ~~shift in scope~~ exponent of the matching cost is a parameter of the procedure, not ~~trivial.~~ ~~The present paper steps~~ 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 ~~level back from that scope.~~ ~~We do not,~~ has been done before: Beshkov and Tiesinga (2022) exchange the Euclidean metric for a geodesic one in the ~~end,~~ ~~advance~~ persistent homology of neural activity and recover different features under the two. And reporting a ~~positive or negative~~ 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 (Steenen 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 ~~claim about~~ grounds, the ~~solar–cyclone connection at~~ residual dependence on the ~~ensemble-TDA scope.~~ ~~Instead we ask~~ remaining unregistered choices is measured rather than argued, and the specification-averaged verdict is reported in a ~~methodological precondition question:~~ 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 ~~ensemble-TDA pipeline give a consistent answer about~~ points: the ~~same data under principled variations of how cyclone trajectories are represented as point clouds?~~ The longitude coordinate is ~~periodic.~~ ~~Three principled, published-style choices exist for handling~~ periodic, and three principled treatments of that ~~periodicity,~~ periodicity exist, each with ~~its own~~ 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 ~~conclusions depend~~ verdict depends on ~~this representational choice—~~ how the trajectories are represented, without ~~converging across them—~~ ~~converging,~~ is not yet ready to test the substantive ~~physical~~ hypothesis. ~~We carry out that test here, and report what we find.~~

1.5 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 to 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:~~ cloud — a linear modular normalization that ignores periodicity, a unit-sphere embedding ~~that maps (lat, lon) to (x, y, z) on S^2 ,~~ and a cylinder embedding ~~that places lat on a linear axis — and lon on a flat (cos, sin) circle.~~ We apply the same downstream pipeline — ~~Vietoris–Rips persistence to dimension 1, Wasserstein-2 distance to a bootstrap control distribution, permutation-based p-values —~~ to the same 25-storm event set (~~25 storms across Halloween 2003, St Patrick 2015, Gannon 2024~~) and 1,020-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 ϵ ; 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 ~~same calendar-matched~~ null is built by resampling control ~~pool (1,020 storms).~~ We report D1 pool,

49-placebo calibration, five-cell stratified attribution, and four-size subsample sensitivity for all three encodings on identical data. Three findings emerge. The D1 H₁ permutation p-value depends on encoding: 0.130 / 0.009 / 0.214 for linear / sphere / cylinder respectively. The H₀ placebo false-positive rate sits near nominal 5 % in every encoding (8.2 / 4.1 / 6.1 %). The H₁ placebo false-positive rate is consistently above nominal in every encoding (8.2 / 10.2 / 14.3 %; directional consistency, not individually significant at $n = 49$ placebos — see §4.2 for Wilson CI discussion). The H₁ inflation is dimension-asymmetric and encoding-robust — i.e. it persists storms. All storm sets are held fixed across encodings that otherwise disagree on the D1 verdict. We discuss what this pattern implies for any positive claim of ensemble-TDA evidence on solar-cyclone perturbation, and propose a minimum protocol of multi-encoding agreement plus dimension-resolved placebo calibration before such claims are advanced: representations, so across-representation differences cannot arise from event or control selection. This analysis is observational rather than a controlled experiment. We observational, not controlled: we manipulate the encoding of a periodic coordinate but not the underlying physical system; representation and the convergence of encoding-comparison findings reported in §4 resampling while leaving the physical system untouched, and what we report is observational convergence, within-pipeline robustness, not controlled-experiment causal isolation. The methodological claim of the paper rests on the encoding comparison as a within-pipeline robustness test; the substantive physical question (whether solar-cyclone perturbation exists at the ensemble-topology scope) remains open and would require an independent line of evidence beyond the scope of the present study. —

2. Data ****Catalog.**** IBTrACS v04r01 satellite-era subset (Knapp et al., 2010): 1965-08 to 2025-12, 2010; dataset: Knapp et al., 2019; 6,661 storms with valid latitude and longitude, 406,059 6-hourly fixes: fixes, downloaded from NOAA NCEI on 30 April 2026. The catalog is cached locally as ‘cache/ibtraes_sat.parquet’: 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 were chosen as events, one major geomagnetic storm benchmark geomagnetic storm per recent solar cycle: Halloween 2003 (peak Dst ≈ -422 nT) is the canonical -383 nT, Cycle 23 extreme: 23), St Patrick’s Day 2015 (peak Dst ≈ -223 nT) is (≈ -234 nT, the strongest Cycle 24 storm with adequate cross-basin tropical-cyclone TC activity in the ± 15 -day window: window), Gannon 2024 (peak Dst ≈ -412 nT) is the strongest (≈ -406 nT provisional, Cycle 25 storm to date. Other Dst ≤ -300 nT events in the satellite era are excluded for sample-window or independence reasons documented in Supplementary §S4. The three chosen events 25). They span three recent solar cycles, sample cycles and three different tropical-cyclone seasons, TC seasons and together provide are the only modern-era extreme events with adequate IBTrACS overlap to support an ensemble-TDA test: ensemble test; the full candidate log is Supplementary §S4. ****Event windows.**** Three peak Peak dates anchor the event windows: Halloween storms 2003-10-29 ± 15 d; St Patrick 2003-10-29, 2015-03-17 ± 15 d; Gannon 2024-05-10 and 2024-05-10, 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 §S4 is built — and recentring on the latter admits one fewer storm and changes no conclusion reported here (§S0, §S4). Every number in this paper uses the 25-storm set defined by the labelled dates. A storm enters the event-period event set if any of its IBTrACS fixes falls inside in any of the three windows. Membership is union-based. The resulting set has window (union rule): 25 unique storms across five basins {WP, NI, EP, SI, SP}, with 1,880 6-hourly fixes collectively, and these counts are fixes, held fixed across all encoding variants reported in §4 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 (§3.5) — so that any across-encoding differences in it includes fixes outside the ± 15 -day window and weights longer-lived storms (mean 75 fixes each) more heavily. The test outcome cannot be attributed 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 differences in which storms the control pool, more recent: mean genesis year 2011.6 against 2001.8, and more SI/WP (17 of 25). Both features are included. The union rule is deliberate. ****Basin composition is uneven across events.**** Table E1 reports the basin distribution. Halloween 2003 is WP-dominated (6 of 11 storms), reflecting the late northern-hemisphere typhoon season. St Patrick 2015 is SI/SP-dominated (7 of 9), reflecting the southern-hemisphere late summer. Gannon 2024 is SI-dominated (3 of 5), reflecting the southern-hemisphere transition month of May. The cross-basin imbalance is intrinsic to the three calendar choices that and are the three peak dates impose, and any per-event subset (D3 cells in §4.3) inherits the basin composition of its parent event window. ****Table E1.**** Basin $\times$ event distribution subject of the 25 event-period storms. Each row counts unique storms whose lifetime overlaps the ± 15 -day peak window matched-null controls of the corresponding event. “Other” captures storms whose IBTrACS BASIN entry uses the legacy SH code or §4.5. The basin-by-event breakdown is missing. [table-row] [table-row] [table-row] [table-row] [table-row] [table-row] Table S2. ****Control pool.**** For each event, the same calendar window in every other year between 1981 and 2024 (excluding 1981–2024, excluding the event year ± 2 to avoid solar-cycle phase contamination). Pooled across the three calendar windows, ± 2 . Pooled: 1,020 unique control storms with storms, 69,454 fixes. Event and control sets are held

fixed across all ~~three encodings reported below. Differences in test outcome are therefore not attributable to event-selection or control-selection variation.~~ representations. —

3. Methods

3.1 Persistent homology: a primer Persistent homology ~~is a tool from topological data analysis (TDA) that~~ summarises the multi-scale ~~geometrie~~ structure of a point cloud (Edelsbrunner & Harer, 2010; Carlsson, 2009; Ghrist, 2008; Wasserman, 2018). ~~Given a finite set of points X (2018), with an environmental-science primer in a metric space and a non-negative scale parameter ϵ , Ver Hoef et al. (2023). Over the Vietoris–Rips complex $VR(X, \epsilon)$ is the abstract simplicial complex whose k -simplices are subsets of $k+1$ points pairwise within distance ϵ . As ϵ grows from 0 to ∞ , $VR(X, \epsilon)$ grows monotonically. This nested family of complexes is the Vietoris–Rips filtration. Topological features — filtration, connected components in dimension zero (H_0), one-dimensional (H_0) and loops in dimension one (H_1), higher-dimensional voids in dimension k — (H_k) are born at some scale ϵ_{birth} and die at some larger scale ϵ_{death} . The collection of $(\epsilon_{\text{birth}}, \epsilon_{\text{death}})$ scales whose pairs for each feature is form the persistence diagram (PD) of the filtration. Features with large death minus birth persist across many scales and are interpreted as robust topological structure of the point cloud; features close to the diagonal are short-lived and treated as topological noise. diagram, which Cohen-Steiner et al. (2007) proved that the PD is stable under perturbations of the input: small changes to X produce only small changes to 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 diagram. Stability anchors PDs as well-defined summary statistics of point-cloud geometry. Ver Hoef et al. (2023) provide an accessible primer of these ideas in the context of environmental-science applications: finite classes. We restrict our analysis throughout to H_0 and H_1 because higher-dimensional H_k ; higher-degree homology is not computed. Neither channel has a literal physical counterpart — H_1 features ~~become combinatorially expensive without obvious physical interpretation at the present sample sizes.~~ 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 ~~Distance~~ Distances between persistence diagrams ~~To~~ We quantify ~~how different two persistence differences between~~ diagrams are, we use with the Wasserstein-2 degree-1 Wasserstein distance W_2 . Given two diagrams P and Q viewed as multisets of points in the extended plane (with the diagonal added to handle features without partner), $W_2(P, Q)$ is the minimum over partial matchings between P and Q of the Euclidean-distance cost summed in quadrature. Mileyko et al. (2011) showed that the space of persistence diagrams equipped with W_2 is W_1 , a complete and separable metric space. This result licenses on diagram space (Mileyko et al., 2011) that sums the statistical operations we perform on it: bootstrap sampling, computing mean distances to a reference set, and significance testing against a null. We use W_2 rather than the bottleneck distance W_∞ because W_2 accumulates mass cost over all matched feature pairs, and therefore reflects ensemble-level differences rather than only pairs. The exponent is not incidental to what follows, so we state the ~~worst-case feature.~~ The Wasserstein distance ~~between diagrams~~ implementation precisely. W_1 is computed with `persim 0.3.8` (Saul et al., 2019). `‘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; §S0 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 ~~of the point cloud and one physical distance~~ Each ~~storm contributes its 6-hourly fixes as vectors~~ fix is a vector in (latitude, longitude, wind-speed) wind space. ~~To bring the three dimensions onto comparable footing for a joint Vietoris–Rips filtration, each must be mapped~~ Latitude rescales linearly to $[0, 1]$: Latitude 1; wind is bounded, 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 ~~may be linearly resealed with no representational hazard.~~ Wind-speed a missing value is non-negative replaced by the median wind of the subset being encoded. That rule is itself a choice, and bounded-above the verdict is sensitive to it: replacing the per-subset median by physical limits; a single global constant moves the great-circle event p -value from 0.012 to 0.034. Neither rule is a natural standard, so we ~~divide by 200 kt~~ report both and clip. treat the difference as part of the specification dependence quantified in §4.7. Longitude is periodic. ~~Three periodic; three~~ encodings are tested.

3.3.1 Linear modular ~~encoding~~ The simplest choice maps $\text{lon} \in [-180^\circ, 180^\circ]$ to (3-D) $\text{lon}_n = (\text{lon} + 180) / 360 \in [0, 1]$ directly. Latitude maps to $180/360$, $\text{lat}_n = (\text{lat} + 90) / 180$; wind to $\text{wind}_n = \text{WMO_WIND} / 200$ (clipped). The point cloud lives in 3-D. The pipeline takes $(\text{lat}_n, \text{lon}_n, \text{wind}_n)$ as joint coordinates in a 3-D box. This encoding has the virtue of

dimensional minimality 90)/180, wind $\mathbf{n}$. Minimal and direct interpretability. Its known defect is that directly interpretable, but the metric does not respect the periodicity of longitude: two storms at $\text{lon} = \text{periodicity} - 179^\circ$ and $\text{lon} = +179^\circ$ are physically 2° apart on the globe but are placed $358^\circ/360^\circ \approx 0.994$ apart in lon though physically 2° apart. A second defect surfaced after the embedded representation. For cyclone trajectories 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 span multi-basin event windows, this discontinuity field places two representations of one meridian 1.0 apart at the antimeridian edge of the domain, which is encountered frequently in where the data: 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 to 0.108 and the event distance from 0.353 to 0.328, and rotating the branch cut by 0° , 90° and 180° gives 0.108, 0.081 and 0.022 (§S6; 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 encoding A more geometrically natural choice maps (4-D) (lat, lon) to the unit sphere $\rightarrow S^2 \in \mathbb{R}^3$ via spherical coordinates: $x = (\cos(\text{lat}) \cos(\text{lon}) \cos \text{lat} \cos \text{lon} + 1)/2, 1/2$, $y = (\cos(\text{lat}) \sin(\text{lon}) \cos \text{lat} \sin \text{lon} + 1)/2, 1/2$, $z = (\sin(\text{lat}) (\sin \text{lat} + 1)/2, \text{each component rescaled to } [0, 1]$. Wind $1/2$, with wind appended. Periodicity is appended as a fourth coordinate $w_n = \text{WMO_WIND} / 200$ (clipped). The point cloud lives in 4-D, with exact and the geographical position constrained to a 2-D submanifold of $[0, 1]^3$. This encoding handles longitude periodicity exactly (the antimeridian discontinuity vanishes) and avoids the flat-circle redundancy of a flat $(\cos \text{lon}, \sin \text{lon})$ circular embedding, which would carry an unwanted $\sin^2 + \cos^2 = 1$ constraint that may inflate H_1 noise. Polar distortion is partially absorbed by the sphere geometry. For cyclone tracks, which sit well away from the geographic poles, the sphere is effectively a rescaled equatorial band. avoided.

3.3.3 Cylinder encoding A third choice puts lat on a linear axis and lon (4-D) lat n linear, longitude on a flat $(\cos, \sin)$ circle: $\text{lat}_n = (\text{lat} + 90) / 180$, $\cos_n = (\cos(\text{lon}) + 1) / 2$, $\sin_n = (\sin(\text{lon}) + 1) / 2$, $w_n = \text{WMO_WIND} / 200$ (clipped). The point cloud lives in 4-D, with the geographical position constrained to the surface of a cylinder $S^1 \times \mathbb{R} \in \mathbb{R}^3$. Like the sphere, this handles circle, wind appended. Periodicity is exact and latitude and longitude periodicity exactly. Unlike the sphere, it decouples latitude from longitude — the joint metric is the L^2 combination of an unconstrained latitude axis with a circular longitude axis. The cylinder is more faithful to the physical decoupling between meridional and zonal storm motion at low latitudes, decouple, but it carries a known liability: the $\cos^2 + \sin^2 = 1$ constraint on the longitude pair means that storms spanning a wide longitudinal range wide-longitude ensembles can populate a near-circular ring in the $(\cos_n, \sin_n)$ plane, which the that Vietoris–Rips filtration is prone to may register as a persistent H_1 loop having with no physical content beyond the periodicity of longitude itself. periodicity.

3.3.4 Construction Construction, sub-sampling, and sub-sampling a reproducibility correction Each storm’s 6-hourly ensemble’s fixes are normalized under one of the three encodings above, then all storms in the relevant ensemble are concatenated. The point cloud is normalized, concatenated and sub-sampled uniformly at random to $N_{\text{SUB}} N_{\text{sub}} = 600$ points, both to keep ripser computation tractable and to standardize 600, with the point count of the filtration across encodings. The seed is set once fixed at script entry (`np.random.default_rng(2026)`) (`‘np.random.default_rng(2026)’`) and consumed sequentially. Three corrections separate the present pipeline from the single-run implementation behind the preprint, and §S0 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 RNG is then consumed sequentially by every subsample draw and every resampling iteration within interpreter hash seed of the script, original runs was not recorded; it retained only one coordinate normalizer at a time, variants being edited in place without version control, so that no two iterations 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 share while two logically equivalent implementations can still build a set in different orders under the same storm subset and the full end-to-end pipeline is reproducible bitwise on rerun. Sub-sample size sensitivity is reported in §4.4 across $N_{\text{SUB}} \in \{400, 600, 800, 1000\}$ for each encoding: seed.

3.3.5 Caveats: ambient dimension confound and embedding-induced homology Two caveats apply to the three-encoding comparison and must be carried into any interpretation of the §4 results. First, the three encodings differ not only in how longitude Ambient-dimension control The linear encoding is represented but also in the ambient dimension of the joint metric space. Linear modular places the point cloud in $[0, 1]^3$; unit-sphere 3-D while sphere and cylinder both add are 4-D, so a fourth coordinate to accommodate the periodic longitude. Differences in test outcome between linear and the two 4-D encodings therefore linear-versus-4-D difference could confound longitude representation with ambient-dimension choice. We do not separately rerun 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 form (for example by appending linear encoding with a constant zero coordinate or a Gaussian-noise nuisance dimension), which would be Gaussian fourth coordinate. That genuine added dimension leaves the natural follow-up placebo calibration nominal (H_1 false positives 2/49), while it moves the event verdict from $p = 0.037$ to isolate 0.072 as the encoding effect added variance dilutes the event H_1 signal (event H_1 features rise from 83 to 207). The two claims are kept separate: the dimensionality effect. A partial mitigation calibration behaviour is that dimension-robust and the H_0 false-positive rate event p -value is calibrated to nominal alpha in all three encodings (§4.2), which suggests the H_1 encoding-dependence is not a generic curse-of-dimensionality artefact, since a curse-of-dimensionality story would predict H_0 calibration to degrade with ambient dimension as well. Nonetheless, dimension-sensitive, so the confound is real at the level of the §4.1 comparison and the three-encoding comparison should be read with it acknowledged. Second, the sphere and cylinder embeddings carry their own intrinsic topology that the Vietoris–Rips filtration sees. The cylinder $S^1 \times \mathbb{R}$ has non-trivial first homology, $H_1(S^1 \times \mathbb{R}) = \mathbb{Z}$. A point cloud densely populating the cylindrical surface—as is the case for any ensemble spanning a wide longitudinal range—eventually registers the global longitude loop as a persistent H_1 feature at filtration scales comparable to the cylinder diameter, irrespective of whether the data-generating process carries any actual topological structure beyond periodicity. The unit sphere S^2 has trivial H_1 ($H_1(S^2) = 0$), but the wind coordinate appended as a fourth axis creates an $S^2 \times [0, 1]$ geometry whose H_1 behaviour at finite sample sizes is not zero in practice. We do not subtract the embedding-induced topological prior absent from the observed H_1 features. §4.2 diagnosis. 3.3.6 A more demanding analysis would compute persistence on a null embedding (e.g. random longitudes preserving latitude physically interpretable common distance The three encodings are three different geometric models, so their disagreement partly reflects that fact, and wind marginals) under each encoding and report event H_1 distance only after subtracting the embedding’s own H_1 contribution at matched point-cloud density. We defer this to future work. We note that the cylinder’s embedding-induced H_1 is a plausible candidate contributor to the cylinder placebo H_1 inflation reported in §4.2 (14.3 %), rather than to the cylinder D1 H_1 result itself (0.214, a null), since the embedding bias would inflate variance in placebo draws more readily than displace the mean need not indicate instability of a single real draw. These caveats notwithstanding, well-posed test. We therefore also build one physically interpretable pairwise distance and feed it directly to the finding reported in §4 stands. The H_1 perm- p value spans an order of magnitude across three principled encodings on identical data; at least one encoding is demonstrably uncalibrated against placebo (§4.2) by persistence computation as a factor $2.9\times$ nominal; distance matrix, so that only the distance step changes while storm selection, sub-sampling and the dimension-asymmetric pattern of H_0 calibrated versus H_1 inflated recurs across encodings; stratification cells (§4.3); statistic are identical: $d(i, j) = \sqrt{[(\gamma(i, j)/\pi)^2 + (\lambda \cdot |\Delta\text{wind}|/200)^2]}$, (1) where $\gamma(i, j)$ is the great-circle angle between fixes i^* and subsample sizes (§4.4). Any successor pipeline must reckon with i^* , 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 diagnostic regardless of whether the specific ambient-dimension and embedding-induced-homology contributions can be cleanly separated from the encoding choice encoders whose normalized coordinates all live on $[0, 1]$. Results are in future work. §4.5.

3.4 H_0 and H_1 features of the storm ensemble H_0 captures the connectivity of the storm-ensemble point cloud at multiple scales. H_1 captures one-dimensional loops in the same cloud. Neither dimension has a literal physical counterpart— H_1 features are not closed storm trajectories, and H_0 components are not literal storm clusters. They are summary statistics of the joint (lat, lon, wind) geometry that aggregate information across all storms in the ensemble, in a way that mean intensity or mean track length statistics do not. H_0 with the infinite-death component removed reflects the multi-scale connectivity pattern. H_1 reflects whether the ensemble in normalised phase space contains non-convex structure such as loops, holes, or recurrence-like features. A perturbation that systematically shifts the joint distribution of storm fixes would change H_0 and H_1 in characteristic ways; whether the present pipeline can read out such shifts robustly is the methodological question the paper investigates. 3.5 Prior TDA in atmospheric and cyclone studies: prior applications Within atmospheric science, persistent science Persistent homology has been used to recognise recognised atmospheric river patterns rivers in satellite-derived gridded water-vapour fields (Muszynski et al., 2019); to quantify 2019), a diurnal cycle in tropical-cyclone single-storm cloud fields from geostationary infrared imagery (Tymochko et al., 2020); 2020) and to characterise weather regimes through the via attractor topology of the atmospheric attractor (Strommen et al., 2022). The present application differs from each. Muszynski 2022); Ghil and Sciamarella (2023) review its wider use in the climate sciences, and Inatsu et al. operate on gridded fields rather than point clouds. Tymochko et al. operate on single-storm imagery (2017) come closest in image space. Strommen et al. operate on attractor reconstructions in phase space. subject, identifying individual cyclones from a superlevel-set filtration with a merge-tree reconstruction. Our scope is — the joint point cloud of multiple many storms’ fixes in geographical-intensity space. We are not aware of prior work that has tested — differs from each, and we put the encoding-invariance question explicitly at the cyclone-ensemble TDA scope, which is the gap the present paper addresses. that scope.

3.6 3.5 Test statistic statistic, resampling, and calibration For each encoding, representation the same downstream statistic

is used. The event ensemble's persistence diagram is compared against a bootstrap resampling distribution of control PDs diagrams by mean Wasserstein-2 W_1 distance from the event PD to a reference set of 30–50 fifty control PDs drawn at random from diagrams; the full bootstrap pool. The null distribution is the same statistic computed with each control PD diagram substituted for the event PD (jackknifing diagram, jackknifing the reference set). The permutation reference; and the resampling p-value is the fraction of null draws whose distance equals or exceeds 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, §S9 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 §§4.3–4.4. 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 to 0.051 and on 0.044 from 0.033 to 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 $B = 10,000$, holding the reference set fixed so that the estimand does not move with B (§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 $B = 10,000$ 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 (§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 (§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 (§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 reference set is drawn from the pool of null draws, so enlarging that pool changes it: recomputing the linear encoding at $B = 10,000$ without pinning it moved the event distance from 0.352771 to 0.349897 and the p-value from 0.0370 to 0.0338 — a change in the estimand, not a gain in precision. Every $B = 10,000$ 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 to 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 For each encoding we report four pipeline outputs. The and implementation D1 pool fuses all three events: pool: 25 event storms vs against 1,020 control storms, with 1,000 bootstrap iterations. $B = 1000$. The 50-placebo 49-placebo calibration uses the same calendar windows (Oct 29 / Mar 17 / May 10) at random years between 1985 and 2020 1985–2020 outside the real event years ± 2 , with ± 2 , $B = 100$ bootstrap iterations per placebo (one placebo is skipped for too few storms). each. Stratified attribution disaggregates the D1 pool into covers two basin cells (D2-WP, D2-SI) and three single-event cells (D3-Halloween, D3-StPatrick, D3-Gannon), with 500 bootstrap iterations each at $B = 500$, corrected by the procedure of Benjamini and Benjamini-Hochberg multiple-comparison correction. Subsample Hochberg (1995). Subsample-size sensitivity reruns the D1 pool at N_{SUB} covers $N_{\text{sub}} \in \{400, 600, 800, 1000\}$ with 500 bootstrap iterations each. **Resampling unit and method.** In every pipeline run, each resampling iteration draws a sample of n storms uniformly at random *without* replacement from the control pool (n matches the event sample size), then concatenates all 6-hourly fixes belonging to those storms before sub-sampling and Vietoris-Rips filtration. The unit matters. The resampling unit is therefore the storm, not the individual fix; within-storm temporal autocorrelation in the 6-hourly track is preserved by construction, and the

iteration-over-iterations distribution reflects between-storm variability under the calendar-matched control distribution. We use the loose term “bootstrap” elsewhere in the manuscript for compatibility with TDA-on-cyclones literature usage, but strictly speaking the procedure is a storm-level subsample-based permutation null rather than a Bernoulli (with-replacement) bootstrap in the sense of Efron (1979). We do not resample at the fix level, which would break the natural dependence structure within a track and inflate the apparent independence of the resampling. We also do not resample at the storm level *with* replacement, which would produce duplicated storms in some iterations and equally inflate apparent independence relative to the calendar-matched control distribution we wish to test against. 3.8 Implementation 1000}. Python 3.11; numpy 1.26, scipy 1.13, pandas 2.2, matplotlib 3.9, [numpy, scipy, pandas, matplotlib](#); ripser 0.6.14 (Bauer, 2021, 2021); persim 0.3.8 (Saul 0.3.8; gudhi 3.12 (see Otter et al., 2019). Vietoris–Rips persistent homology is computed to dimension 1 via ripser. Otter et al. (2017) 2017, for a survey alternative computational pipelines: of persistent-homology computation). The analysis scripts (with the three encodings as switchable variants) and the run logs containing 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 numerical result reported below 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 available 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 §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 corresponding author: 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: §S9 sets out what falls outside it, including the five values recomputed at $B = 10,000$ and eighteen of the thirty-six cells of the matched-null comparison of §4.5, and says where each is anchored instead. —

4. Results

4.1 The D1 pool: signal flips the verdict moves with the encoding Table 1 reports gives the D1 pool result pooled event-versus-control comparison under all the three encodings on and, for reference, under the same event and control sets: physical distance of §4.5. **Table 1.** D1 pool ensemble TDA under three longitude encodings. Each row: event-PD Wasserstein-2 encodings, on identical event and control sets. Columns: mean distance from the event diagram to the mean control PD; bootstrap fifty reference diagrams; null mean over 1,000 control draws of 25 storms each; permutation p-value as draws; H_1 resampling p at $B = 1000$ and at $B = 10,000$; and the fraction of null draws whose distance equals or exceeds the event distance. Sources: ‘derived/11_4d_run.log’, ‘derived/11_sph_run.log’, ‘derived/11_cyl_run.log’. [table row] [table row] [table row] [table row] The H_1 perm-p value spans roughly an order of magnitude across the three encodings. Three numbers, three verdicts. Under the linear modular encoding the H_1 test is non-significant (perm-p = 0.130); a naive readout would conclude no ensemble-TDA signal. Under the unit-sphere encoding the H_1 test crosses the conventional $\alpha = 0.05$ threshold by a wide margin (perm-p = 0.009); a naive readout would conclude a positive signal. The cylinder tips back to null. Under it the test is again non-significant (perm-p = 0.214), with the event H_1 distance sitting essentially at the null mean, a pattern consistent with the unit-sphere positive being tied to properties range of the sphere metric rather than H_1 resampling p over six event-subsample seeds at $B = 1000$, seed 2026 having been frozen before the analyses of §§4.1–4.5 and the five others added afterwards. The linear row is reported under both longitude conventions: the archived one, under which §§4.2–4.4 were also computed, and the single convention of §3.3.1, whose value is one point in the range 0.022 to a coherent topological signature shared across all three reasonable representations of the same trajectory data. The H_0 perm-p values are broadly consistent across encodings (0.082 to 0.190, none crossing 0.05) but vary by a factor of more than two. Three principled encodings of the same data, all of them defensible at face value, give three qualitatively different verdicts on the D1 pool H_1 result. Before any positive or negative claim can be sustained, the question must be asked which encoding’s verdict is to be trusted. The placebo calibration in §4.2 begins to address this. 4.2 Placebo calibration: H_0 calibrated, H_1 inflated in every encoding The 49-placebo calibration is the same across encodings: 49 fake event-date sets (one of 50 was skipped for too few storms), each a random year in {1985,...,2020} excluding real event years ± 2 , at the same calendar windows. Each placebo runs the full pipeline with 100 bootstrap iterations. Table 2 reports the false-positive rate at $\alpha = 0.05$ in both dimensions. **Table 2.** Placebo calibration of the D1 pool pipeline under three longitude encodings. Same 49 dates per encoding. False-positive rate is the fraction of 49 placebos whose perm-p falls below 0.05, with the Wilson 95 % CI for nominal 5 % at $n=49$ being approximately [1.4 %, 14.7 %], so a placebo-rate observation outside that interval is formally inconsistent with the pipeline holding nominal alpha at the conventional 5 % significance level. Sources: ‘derived/12_4d_run.log’, ‘derived/12_sph_run.log’, ‘derived/12_cyl_run.log’. [table row] [table row] [table row] [table row]

[table row] Three observations. First, the H_0 false-positive rate sits near the nominal 5 % rate in all three encodings (4.1 % to 8.2 %), with all values inside the Wilson 95 % confidence interval for nominal 5 %. The H_0 channel of the pipeline is calibrated, encoding-robust. Second, the H_1 false-positive rate is systematically above nominal 5 % in every encoding, ranging from 8.2 % to 14.3 %. The factor of inflation grows as the encoding moves from linear modular (1.6 $\times$) to unit-sphere (2.0 $\times$) to cylinder (2.9 $\times$). Third, the cylinder H_1 rate sits at the upper bound of the Wilson 95 % CI for nominal 5 % — under a strict reading, it is borderline incompatible with nominal calibration; under a permissive reading, it is the worst case of a pattern visible in all three encodings. The pattern is dimension-asymmetric. H_0 holds nominal alpha encoding-robustly; H_1 fails nominal alpha in a direction consistent across encodings. The encoding choice modulates the magnitude of the H_1 inflation but does not eliminate it. This has direct implications for the §4.1 D1 verdict. The unit-sphere D1 perm-p $H_1 = 0.009$ must be read against an H_1 calibration that produces false positives at $\approx 2\times$ nominal. Adjusting upward [0.108 spanned](#) by the calibration factor ($0.009 \times 10.2/5.0 \approx 0.018$) still leaves a value below 0.05, so [branch cut \(§S6\)](#). A positive control reproduces the qualitative signal does not vanish, but the marginal-significance reading is degraded. More importantly, the calibration finding undermines the precondition [archived cylinder seed row bit](#) for ever calling a positive D1 H_1 result a clean detection: a positive result in H_1 from this pipeline rests on a noise channel known to be 1.6 $\times$ –2.9 $\times$ over nominal. ****Three further diagnostics.**** First, across the 49 cylinder placebos, the H_1 event-control distance has $CV \approx 0.50$ (mean 0.248, std 0.125), compared to the H_0 CV of ≈ 0.24 (mean 1.47, std 0.36). The H_1 channel is twice as variable relative to its scale, consistent with H_1 features in this pipeline being dominated by a small number of high-persistence loops whose presence-absence varies sharply across event-window draws. Second, the H_1 null mean itself varies with $CV \approx 0.48$ across placebos, versus $CV \approx 0.12$ for H_0 , confirming that the H_1 null geometry is unstable across event windows in a way the H_0 null is not. Sources for both diagnostics: ‘derived/placebo_test.csv’ (cylinder run, $n = 49$), cross-placebo statistics computed by inline pandas summary at audit time. Third, with $n = 49$ placebos and $\alpha = 0.05$ the Wilson 95 % CI for nominal calibration spans [1.4 %, 14.7 %]: the cylinder H_1 rate (14.3 %) sits at the upper bound, and we cannot strictly reject the null of nominal calibration on $n = 49$ alone, although the directional consistency across encodings makes the inflation reading the more parsimonious one. We recommend $n \geq 200$ placebos for any future replication of this calibration assessment, which would tighten the Wilson CI to approximately [2.5 %, 9.0 %] at the 5 % nominal rate and resolve the H_1 -specific inflation more sharply.

4.3 Stratified attribution: same dimension asymmetry recurs Disaggregating the D1 pool across two basin cells (D2_WP, D2_SI) and three single-event cells (D3_Halloween, D3_StPatrick, D3_Gannon) gives the picture in Table 3. ****Table 3.**** Stratified attribution under three longitude encodings. Five cells: D2_WP, D2_SI, D3_Halloween, D3_StPatrick, D3_Gannon. We report uncorrected perm-p H_1 for each cell, with Benjamini–Hochberg multiple-comparison correction reported in parentheses next to cells whose uncorrected p falls below 0.10, where the correction divides the rank-ordered p-values by an FDR-derived constant computed across the five cells separately within each encoding column rather than pooled across encodings. Sources: ‘derived/13_4d_run.log’, ‘derived/13_sph_run.log’, ‘derived/13_cyl_run.log’. [bit](#). [table row] [table row] [table row] [table row] [table row] [table row] [table row] [table row] Two cells (D2_WP, D3_StPatrick) sit near or below [Three representations of the uncorrected same 25 storms](#) give three verdicts. Under the archived longitude convention the linear and sphere encodings cross $\alpha = 0.05$ threshold 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 and cylinder encodings, and well above α under sphere. After BH correction across encoding, on where the five cells, no cell crosses α in any encoding. [circle was cut](#). The cell that returns [relative structure](#) is the strongest deviation differs by encoding: linear’s most-extreme cell is D2_WP at 0.046; sphere’s is D3_Halloween at 0.238 (still null); cylinder’s is D3_StPatrick at 0.044. None of these survives multiple-comparison correction. The same dimension asymmetry recurs. H_0 cell-level perm-p values (omitted from Table 3 for compactness; see ‘derived/13_*_run.log’) are above 0.10 in nearly every cell [more stable part](#): under every encoding — H_0 one of the six event-subsample seeds the sphere p-value is uniformly null at the cell scale. H_1 cell-level perm-p values straddle α at smallest of the uncorrected level 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 [some encoding-cell combinations](#), but reach BH significance the first three rows occurs in none. The pooled D1 H_1 verdict under any one seed of six. Supplementary §S5 gives the six seeds in full, together with the ENSO and solar state of every event and placebo window. An earlier single encoding (0.130 run reported 0.130 / 0.009 / 0.214) 0.214 for these rows; §S0 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. ****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; §S1 gives the per-encoding breakdown of the deterministic ordering. [table row] [table row] [table row] [table row] 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 **localizable** 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 (§S0). 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 independently 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 to 2.2 times larger for H_1 than for H_0 and that of the null mean itself 3.9 to 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. Ambient dimension is likewise excluded by the 4-D linear control of §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. ****Table 3.**** Stratified attribution. Uncorrected H_1 resampling p per cell **under** 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 **own disaggregation**. Combined with the encoding sensitivity five cells. §S0 records why these values replace those of the pooled D1 perm-p itself, this means the apparent unit-sphere D1 H_1 signal is not localized to a coherent single-cell source.

4.4 Subsample-size sensitivity: non-monotonic in every encoding Sub-sampling the event and control point cloud to $N_{\text{SUB}} \in \{400, 600, 800, 1000\}$ prior to Vietoris–Rips filtration probes whether the D1 H_1 verdict is sensitive to a fixed sub-sample size. Table 4 reports the result. ****Table 4.**** D1 pool H_1 perm-p across subsample sizes under three longitude encodings. 500 bootstrap iterations per cell. Sources: ‘derived/14_d_run.log’, ‘derived/14_sph_run.log’, ‘derived/14_cyl_run.log’. earlier versions. [table row] [table row] [table row] [table row] [table row] [table row] [table row] No cell survives correction in any encoding. The most extreme cell is D2 WP in all three, at 0.048 to 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 to 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 ****Table 4.**** D1 pool H_1 resampling p across subsample sizes, $B = 500$ per cell, under the archived longitude convention. [table row] [table row] [table row] [table row] [table row] [table row] In Table 4 the linear encoding **yields** is monotonic in N_{SUB} and strengthens at $N = 1000$; the sphere and cylinder encodings are not monotonic and weaken there. We read the table as ordinary sampling variability of the H_1 channel; it is not a **monotonic dependence on N_{SUB}** . Under linear, size dependence, since the **smallest perm-p** same event realization tested at different subsample sizes draws different point subsets, so each entry is at $N_{\text{SUB}} = 600$ (0.124, still null); under sphere, 800 (0.004); under cylinder, 600 (0.182). 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 $B = 1000$ **row** 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 §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 to 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; §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; §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 to 0.034 (§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. ****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. [table row] [table row] [table row] [table row] [table row] [table row] Under W_1 the difference survives all four matchings, non-monotonically and informatively so: matching basin alone sharpens it, matching era alone weakens substantially it the most, and the joint match sits between, pointing to era over basin as the principal confounder. Those twelve verdicts are one-sided (§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; §S9 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 encoding, including 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 value is 41.3 % for the event set against a control-draw mean of 38.0 %, the 81st percentile of 1,000 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 (§S7). 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. Sections 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. ****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 1,000 control draws, with the event's percentile in that control distribution. The decomposition of the matching cost that carries the excess is Table S5. A positive control reconstructs the archived event diagram exactly (90 features, total persistence 0.9076). [table row] [table row] [table row] [table row] [table row] [table row] [table row] [table row] [table row] 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° to 7.2° , equivalently to wind-speed differences of 4 to 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 — $\sum c_i$ against $(\sum c_i^2)^{1/2}$ — and, because the assignment is re-optimised at each exponent, they need not aggregate the same vector. Table S5 gives the measured decomposition. Squaring moves the weight toward the largest costs: the single most persistent feature takes 14.2 % of $\sum c_i$ but 48.7 % of $\sum c_i^2$, the four bins below persistence 0.04 fall together from 64.5 % to 31.1 %, and the bin carrying the event excess drops from 26.5 % to 17.9 %. The feature that gains

is the one §4.6.1 found unremarkable. Rematching is small: the assignment changes for 21 % to 35 % of matched pairs, yet holding the $p = 1$ assignment fixed and merely squaring its costs reproduces 99.2 % to 99.5 % of the shift, so 0.5 % to 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 % to 18.6 %: the separation in those units falls from 2.93 to 1.07 and the resampling p-value moves from 0.003 to 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 to 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 to 1.65 standard deviations across the range while the sphere ~~where~~ decays from 1.76 to 0.52, the cylinder from 0.59 to -0.13 and the great-circle $\lambda = 1$ from 1.58 to 0.64. This follows §4.6.1: under the linear encoding the difference 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 ~~boundary~~ 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 ~~α~~ contributing terms as $N_{\text{eff}} = 0.05 \cdot (\sum_i w_i)^2 / \sum_i w_i^2$, 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 ~~pattern~~ functional itself is consistent with a noise channel whose realisation depends on which subsample standard — the seed happens to draw, rather than a stable signal as a function effective sample size of point-cloud size. The sphere column in Table 4 deserves explicit comment. Read in isolation, the sphere $D1\ H_1 = 0.009$ (Table 1) Kish (1965) and the sphere $N_{\text{SUB}} \in \{600, 800\}$ sub-sample p-values $\{0.008, 0.004\}$ (Table 4) jointly invite participation ratio of condensed-matter physics — and we present it as such; what we add is the mapping to the matched costs of a ~~positive narrative:~~ low p-values robust to sub-sampling at mid-range N . Read in persistence-diagram comparison. Both counts follow from the context pair of Tables 1, 2, 3, 4 jointly, the same numbers read very differently. The linear diagrams, with no null distribution and cylinder encodings disagree with sphere on $D1$; all three encodings agree that H_1 placebo calibration fails nominally; no encoding stratifies the pooled signal resampling required, so they are available before one commits to a single cell; exponent for inference. **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 ~~$N_{\text{SUB}} = 1000$ row degrades sharply~~ 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. [table row] [table row] [table row] [table row] [table row] [table row] [table row] [table row] 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 % to 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 ~~encoding including sphere.~~ one of the sixty subsample-by-representation combinations; on the median of ten subsamples the overshoot runs from 9 % to 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 S7 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 ~~sphere sub-table~~ 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 to 3.7, wider than the 1.91 to 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 §S7: 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 ~~coherent island inside~~ 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 ~~otherwise incoherent multi-encoding picture.~~ A single-encoding ensemble TDA pipeline 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. Supplementary §S8 reproduces the two pre-registration records that ~~selected sphere~~ were deposited before their runs, and the specification matrix that separates what was registered from what was not. Reporting a ~~priori would have reported a positive signal~~ 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 (Girardi et al., 2024) supplies post-selection inference over that distribution, noting that the **broader-picture** 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. ****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 $B = 10,000$ 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. [table row] [table row] [table row] [table row] 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 to 0.0443 and from 0.0419 to 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. ****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. [table row] [table row] [table row] [table row] [table row] [table row] 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). 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 to 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.

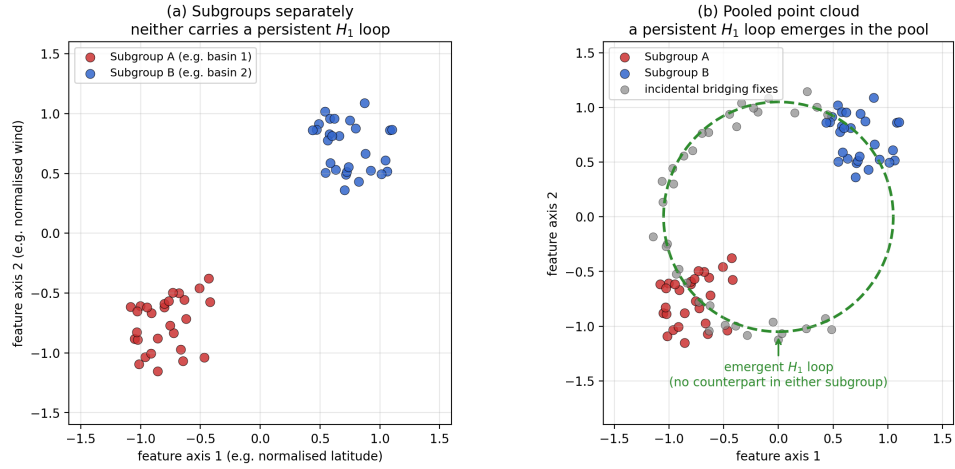
4.7.3 What this design does and **does not support:** resolve The direction of the effect is consistent across every specification examined, and its magnitude does not reach the conventional threshold once neither the exponent nor the subsample is selected. That conclusion is one-sided (§3.5), and a two-sided reading only strengthens it, since the $M = 10$ family values of Table 9 already sit above α (§S9). 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 (§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 to 0.674 (§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 ~~The three findings of §4 admit a single interpretation: the ensemble-TDA pipeline as currently configured is fragile in two coupled ways. Two failures. The pipeline is encoding-fragile in H₁: a defensible variation in how longitude is~~

represented can flip the H_1 verdict. And it is dimension-asymmetric in calibration: H_0 holds nominal α across all three encodings tested in §4.2, while H_1 over-rejects by $1.6\times$ to $2.9\times$ across the same encodings, indicating that the H_1 channel of the pipeline returns false positives at a rate substantially above the nominal 5 % under any of the encodings we tested. The first fragility says the pipeline’s H_1 verdict on the same data is not single-valued. The second says the H_1 channel that returns those verdicts sits above nominal α to begin with.

5.1 The H_1 verdict cannot be read directly off one encoding. The standard practice in ensemble-TDA studies on geographical point clouds has been to choose encoding, nor off one encoding — typically a linear modular normalization — and report its result as the result. The choice is not innocent. The three encodings in this paper are not exotic variations: each is published, each is physically motivated, and each, applied alone to the same trajectory data, alone, would have produced a paper claiming the answer to reporting a different verdict on the same question with three different verdicts and three different sets data. Reporting one of consequences for them as “the answer” without the literature on solar perturbations of cyclone activity. The linear modular encoding is the dimensional minimum and is what a naive implementation would write first. The unit-sphere embedding is the geometrically faithful choice and is what a methodologically careful implementation would adopt to handle longitude periodicity. The cylinder is the decoupled-axes choice. It is what a physicist would write if asked to keep latitude and longitude as independent observables. None of the three is wrong in itself. But all three disagree on the H_1 perm- p (0.130, 0.009, 0.214) for the same event set and the same control pool. The disagreement is not noise around a common value; it is an order-of-magnitude swing that crosses both above and below the conventional $\alpha = 0.05$ line. We do not see how to defend reporting the sphere value (0.009) as the answer without also reporting the linear (0.130) and cylinder (0.214) values, others, and without doing so under transparent the caveat that the answer is encoding-conditional. This may seem like a methodological house-keeping observation. We argue it encoding-conditional, is more than that. In TDA the not defensible. The stability theorem of Cohen-Steiner et al. (2007) is sometimes read as guaranteeing that small input perturbations produce small does not rescue this: it bounds output perturbations. That is true change under the a “fixed” input metric the theorem refers to. The theorem and says nothing about the choice of metric itself. 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 (linear vs sphere vs cylinder) 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 “small perturbation” “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 input — it 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 diagrams 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 choice resampling verdict. The spread measured in §4.1 is a spread of which 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 §4.2 is a property of the sparse H_1 channel: the metric to use. The stability theorem is silent on 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 choice. The encoding-flip we observe here “topological aggregation instability” as a working descriptive term, not a formal result; the cell-level null pattern of §4.3 is consistent with the stability theorem, not a violation of it; it is a reminder that the choice of metric on a periodic geographical coordinate is itself a modelling choice, not a free parameter. 5.2 H_1 inflation is intrinsic, not encoding-specific. The H_1 placebo over-rejection (Table 2) is the more striking observation. Across all three encodings, false-positive rate at $\alpha = 0.05$ is at $1.6\times$ nominal at the friendliest end (linear, 8.2 %) and at $2.9\times$ nominal at the unfriendliest (cylinder, 14.3 %). In none of the three is H_1 false-positive rate close to the nominal 5 %, and in none of the three does H_0 over-reject in a comparable way. The pattern is consistent with an intrinsic feature of pooled cyclone-trajectory point clouds: H_1 loops in the joint (lat, lon, wind) cloud are sparse and small-effect under the null, it, and the bootstrap null distribution for them is therefore narrow. A modestly outlying event ensemble can pierce that narrow null even when no physically coherent ensemble-level effect is present. The H_0 channel does not have this property — H_0 is dominated by the global connected component plus variance measurements of §4.2 give it a

[Figure removed in revision — superseded by the figure below]



[Figure added in revision]

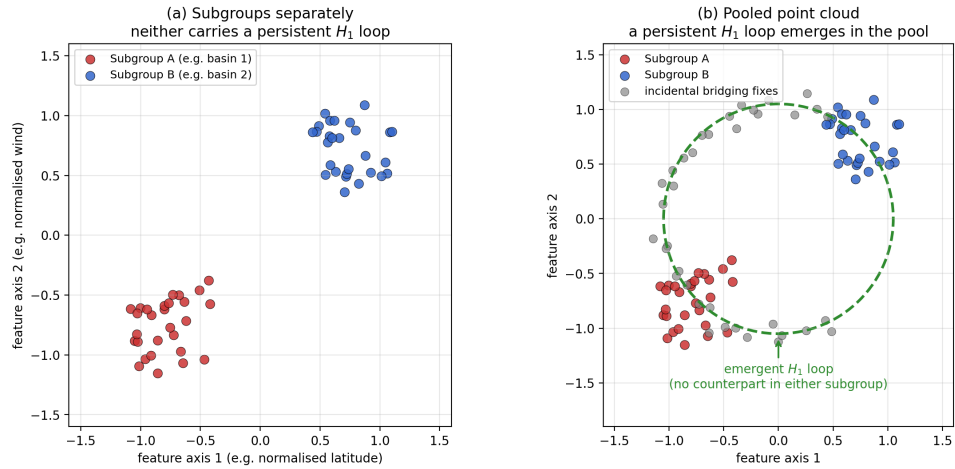


Figure 1. Schematic of topological aggregation instability. Left: two subgroups (red and blue point clouds), each occupying a distinct region of (lat, lon, wind) phase space; subgroups, neither carrying a persistent H_1 loop on its own. Right: the pooled point cloud, in which a shared Vietoris–Rips filtration generates a persistent H_1 loop spanning the gap between the two subgroups — a feature that has with no counterpart in either subgroup considered separately, and subgroup, that disappears under disaggregation. Pooling is not a linear operation on persistence diagrams; subgroup cancellation diagrams. 5.3 Metric, realisation, and what the diagnostic is for Sections 4.6 and 4.7 change what the repair of §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 to 0.087, the second from 0.012 to 0.069, and the third, tested on the linear encoding alone, over 0.036 to 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 topological setting 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 inspectable from a general rule — under the pooled diagram alone. This figure shows why H_1 specifically linear encoding the difference is the dimension-at-risk for aggregation artifacts in this class of data. Two dimensions, two behaviours. H_0 is less susceptible because adding cross-subgroup-connecting points reduces, not creates, connected-component features — joining two clusters under a Vietoris–Rips filtration eliminates a feature present in the H_0 channel rather 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 opening one, so 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 H_0 null distribution is broader and less easily pierced by a modest outlier event ensemble. H_1 carries degree- p Wasserstein distance sums the asymmetry: inserting a bridging point between two subgroup clouds opens a loop that did not exist in either alone. The H_1 inflation observed in our placebo calibration (§4.2) is consistent with this geometric picture as one contributing mechanism. We use p -th powers of the label *topological aggregation instability* as a working descriptive term for matched costs (Cohen-Steiner et al., 2010), raising p concentrates the observed failure mode rather than a formal theoretical concept. A rigorous treatment — for example, a theorem characterizing the conditions under which pooling two contractible point clouds connected by sparse bridge points produces a persistent H_1 feature in the union that has no counterpart in either subgroup — is left to future work. The figure schematic and the §4.3 stratified-attribution null pattern are evidence consistent with the label, not proof of a general principle. We retain the working label in the paper because the cell-level null pattern of §4.3 is informative for future studies even without a formal characterization, and the label gives a working vocabulary for the failure mode that any subsequent theoretical treatment can refine, reject, or replace.

5.6 Pathways toward H_1 -calibrated ensemble TDA The minimum protocol of §5.4 constrains what should be reported but does not fix the underlying H_1 inflation diagnosed in §4.2 and §5.2. Three pathways toward a fix appear worth attempting. We do not attempt any of them in the present paper. We sketch each so that a follow-up study — by us or others — can pick one and report whether it meets the calibration requirement that the current pipeline does not. The first pathway replaces the test statistic. Wasserstein-2 between persistence diagrams produces a null whose narrowness in H_1 reflects, in part, the W_2 metric concentrating mass on a small number of high-persistence features and assigning negligible weight to the dense low-persistence tail. Alternative summary statistics distribute mass differently. Persistence landscapes (Bubenik, 2015) convert each PD into a function statistic on the real line, 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 L^2 -distances on the resulting function space have a different null geometry that may be wider in the relevant H_1 regime. Persistence images (Adams et al., 2017) provide a vectorized PD representation amenable to kernel methods on point clouds (Reininghaus spread over many (Fan et al., 2015), with and maximum-based procedures attain the kernel bandwidth offering a tunable handle on 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 §4.6.3 is one way of reading it: it envelopes the widening of the null variance. A simpler first-pass attempt would normalize the H_1 statistic itself: dividing the event H_1 distance by in all sixty combinations examined, while as an estimate of the control bootstrap standard deviation transforms size of that widening it does not improve on a raw distance into constant, and its sub-sampling scatter is wide enough that no ranking of representations follows. Its practical use is as a z -score 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 may widen 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 (§S9) — 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 without-changing microphysical literature; the underlying TDA framework. The second pathway enlarges the event sample from the 25 storms used here to all satellite-era extreme geomagnetic events meeting three scopes interrogate different signatures, and a $K_p \geq 8$ threshold (approximately 50 events). Sample size alone 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 cure. Narrowness priority follow-up and beyond the scope of the H_1 null present study.

5.5 A structural limit on what this design can test A second limitation is a shape property; a larger 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 §S4 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 sample may not by itself widen the null, but it raises the bar a modest outlier ensemble must clear to register against the null, and a $2\times$ sample increase would tighten the Wilson 95 % CI eligibility depends in part on the placebo false-positive rate very quantity the test examines. For a methodological benchmark this is harmless and indeed convenient, since one wants windows with enough data to exercise the pipeline. For a test of solar or geomagnetic forcing it is not: a design in which exposure is defined partly by approximately $\sqrt{2}$, allowing finer resolution 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 ~~residual encoding-specific calibration bias~~. The third ~~pathway replaces~~ 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 ~~Vietoris–Rips persistence framework entirely~~. Mapper-graph homology, cubical persistence on gridded fields, and witness complexes each carry different sensitivity to point-cloud aggregation. The cost is re-implementing the full pipeline at the framework level. The payoff is independence from the specific Vietoris–Rips H_1 inflation diagnosed here, which may be tied ~~source archive~~ to a single ~~computational choice~~ 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 ~~to ensemble TDA on cyclone trajectories in general~~. We frame a reporting format, since reordering a uniformly resampled control pool selects a different realisation while the estimand does not move (§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 Supplementary §S2. The present paper 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 §4.7.1 and Table 1 show. The two statistic-level remedies a ~~methodological diagnostic that stands on its own~~. A natural follow-up would attempt pathway 1 (H_1 statistic normalization, then persistence landscapes, then persistence-image kernels) on try first do not help (§4.2); the same data, productive move was the physically interpretable distance of §4.5, which removes the representational ambiguity that those remedies leave in place. Persistence images with the goal of meeting the placebo H_1 calibration target of nominal 5 %; conditional on that follow-up reaching nominal calibration, a further study could return to the substantive solar-cyclone question with a calibrated tool. The present paper does not commit to a specific timetable for either follow-up step. We report the present diagnosis now because it has standalone methodological value: it warns any current user of the existing pipeline against a positive H_1 claim kernel methods (Adams et al., 2017; Reininghaus et al., 2015), larger event samples, and motivates the search for a fix by us or by other groups. non-Vietoris–Rips frameworks remain open directions. —

6. Conclusions We applied an ensemble-TDA pipeline to test ~~Testing~~ whether three benchmark extreme geomagnetic events (Halloween 2003, St Patrick 2015, Gannon 2024) perturb the ~~cyclone-activity~~ topology of the ~~cyclone activity~~ in their immediate ± 15 -day windows. The result is mixed. Three 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 (lat, lon, wind) point cloud — linear modular, unit-sphere, 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 to 0.087), with which 600 of the ensemble’s fixes enter the diagram (0.012 to 0.069), and lat-linear-plus-longitude-cylinder — give three qualitatively different $D1 H_1$ perm-p-values (0.130, 0.009, 0.214) with which fifty control diagrams form the reference set (0.036 to 0.047, tested on the same event set linear encoding alone and crossing no threshold), and it is absent under the true degree-2 and the same control pool, demonstrating that exact bottleneck distance. The reason is legible in the ~~pooled-ensemble test verdict in this pipeline depends on diagram~~: the choice of representation excess sits at a level that crosses both above and below the conventional $\alpha = 0.05$ line of demarcation between null and positive results in this literature. A 49-placebo calibration shows that the H_0 false-positive rate sits near the nominal 5 % in every encoding (8.2 %, 4.1 %, 6.1 %); intermediate persistence while the H_1 false-positive rate single most persistent feature is consistently above nominal in every encoding (8.2 %, 10.2 %, 14.3 %; directional consistency across encodings, with cylinder at the Wilson 95 % upper bound 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 $n = 49$). Stratified attribution 1.87 standard deviations. The substantive outcome is a design- and subsample-size sensitivity recur 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 same pattern: H_1 results that

disagree across encodings, no cell-level signal surviving BH-correction effect at all, because event eligibility depends in any encoding, no monotonic dependence part on N_{SUB} . The pipeline is encoding-fragile in H_1 and dimension-asymmetric in calibration. Under these constraints we cannot defend either a positive or negative substantive claim about ensemble-topology evidence for or against the solar-cyclone perturbation hypothesis at this scope. We frame the contribution as a methodological cautionary tale: pooled cyclone activity itself (§5.5). Pooled point-cloud TDA topology on geospatial trajectories with periodic coordinates is more fragile than should therefore prefer a physically interpretable input distance, register the exponent of its formal stability theorems suggest. We propose matching cost and its reference set or report a minimum protocol — multi-encoding agreement testing plus dimension-resolved placebo calibration plus pre-registered stratified attribution — as a precondition for any positive ensemble-TDA claim 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 this class of problems. Pathways toward an H_1 -calibrated successor pipeline are sketched in §5.6 as starting points for any successor study: both homology channels. —

Code and data availability IBTrACS v04r01 is publicly available at <https://www.ncei.noaa.gov/products/international-best-track-archive>. Analysis scripts (`'scripts/11_event_ensemble_tda.py'`, `'12_placebo_test.py'`, `'13_stratified_attribution.py'`, `'14_subsample_sensitivity.py'`) The analysis code — a single pipeline module with the three longitude encodings and the great-circle distance as switchable variants (`'normalize_pe_sphere'`, `'normalize_pe_cylinder'`, variants, the placebo, matched-null, bottleneck, landscape and subsample drivers, and the in-line linear modular form), specification record and verification script of §3.7 — together with the run logs and derived run logs (`'derived/{11,12,13,14}_{4d,sph,cyl}_run.log'`) reproducing every numerical tables, is archived at Zenodo (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 reported printed here, and the frozen outputs of the superseded runs discussed in Tables 1–4 §3.3.4 and Table E1, §S0, so that the corrections described there can be inspected rather than taken on trust. Run-log paths quoted in this paper are available from relative to the corresponding author on reasonable request. `'v7/'` directory of that archive.

Author contributions RD designed the study, implemented the three-encoding pipeline, executed the bootstrap and placebo runs, performed the stratified-attribution analyses, and subsample-sensitivity analyses, drafted wrote the manuscript, and revised it for submission: manuscript. RD is the sole author and accepts responsibility for the content.

Competing interests The author declares no competing interests.

Acknowledgements Analysis An AI system (Anthropic Claude) was performed using Anthropic Claude Code used as a coding assistant, with and writing assistant; all design decisions, statistical interpretations, encoding-comparison framing, and writing by conclusions are the author. The longitude-periodicity bug in the v1–v3 single-encoding pipeline was identified through an internal pre-submission audit of the v3 manuscript; the diagnosis motivated the three-encoding analysis reported here: author's. IBTrACS data are provided by NOAA NCEI. Financial support This research has received no external funding.

References Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., & Ziegelmeier, L. (2017). L.: Persistence images: a stable vector representation of persistent homology. *J. homology, J. Mach. Learn. Res.*, 18, 1–35. 1–35, 2017. Adams, H. and Moy, M.: Topology applied to machine learning: from global to local, *Front. Artif. Intell.*, 4, 668302, <https://doi.org/10.3389/frai.2021.668302>, 2021. Bauer, U. (2021). U.: Ripser: efficient computation of Vietoris-Rips Vietoris-Rips persistence barcodes. *J. 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. (2015). P.: Statistical topological data analysis using persistence landscapes. *J. landscapes, J. Mach. Learn. Res.*, 16, 77–102. 77–102, 2015. Calogovic, J., Albert, C., Arnold, F., Beer, J., Desorgher, L., & Flueckiger, E. O. (2010). 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. (2009). G.: Topology and data. *Bull. Amer. Math. Soc.*, 46(2), 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., & Harer, J. (2007). J.: Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1),

103–120. <https://doi.org/10.1007/s00454-006-1276-5> *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. (2010). — **Computational L.: Computational Topology: An Introduction**. Introduction, American Mathematical Society, Providence, RI. <https://doi.org/10.1090/mbk/069> RI, <https://doi.org/10.1090/mbk/069>, 2010.

Elsner, J. B., & B. and Jagger, T. H. (2008). H.: United States and Caribbean tropical cyclone activity related to the solar ~~eye~~. **Geophys. cycle, Geophys. Res. Lett.**, *Lett.*, 35, L18705. <https://doi.org/10.1029/2008gl034431> L18705, <https://doi.org/10.1029/2008GL034431>, 2008.

Elsner, J. B., & B. and Kavnikov, S. P. (2001). P.: Hurricane intensity changes associated with geomagnetic ~~variation~~. **Atmos. variation, Atmos. Sci. Lett.**, *Lett.*, 2, 86–93. <https://doi.org/10.1006/asle.2001.0043> 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. (2008). R.: Barcodes: the persistent topology of ~~data~~. **Bull. Amer. data, B. Am. Math. Soc.**, 45(1), 61–75. <https://doi.org/10.1090/s0273-0979-07-01191-3> Ver Hoef, L., Adams, H., King, E. J., & Ebert-Uphoff, I. (2023). *A primer 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 ~~topological data analysis to support image analysis tasks in environmental science~~. **Artif. Intell. Earth Syst.**, 2(1), AIES-D-22-0039.1. <https://doi.org/10.1175/aies-d-22-0039.1> the sign flipping score test, *Psychometrika*, 89, 542–568, <https://doi.org/10.1007/s11336-024-09973-6>, 2024.

Hodges, R. E., & E. and Elsner, J. B. (2011). B.: Evidence linking solar variability with US ~~hurricanes~~. **Int. hurricanes, Int. J. Climatol.**, *Climatol.*, 31, 1897–1907. <https://doi.org/10.1002/joc.2196> 1897–1907, <https://doi.org/10.1002/joc.2196>, 2011.

Hodges, R. E., Jagger, T. H., & and Elsner, J. B. (2014). B.: The sun–hurricane connection: diagnosing the solar impacts on hurricane frequency over the North Atlantic basin using a space–time ~~model~~. **Nat. Hazards*, model, *Nat. Hazards*, 73, 1063–1084. <https://doi.org/10.1007/s11069-014-1120-9> 1063–1084, <https://doi.org/10.1007/s11069-014-1120-9>, 2014.*

Huth, R., & R. and Stryhal, J. (2024). J.: Association of the 11-year solar cycle with correlation and teleconnection structures in tropospheric ~~circulation~~. **J. circulation, J. Space Weather Space Clim.**, *Clim.*, 14, 31. <https://doi.org/10.1051/swse/2024031> IPCC AR6 WGI (2021). **Climate 31*, <https://doi.org/10.1051/swse/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*~~ [Masson-Delmotte, V., et al. (eds.)], Change, Chapter 7, Cambridge University Press, ~~Chapter 7~~. Cambridge, UK, 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., et al. (2011). 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*, nucleation, *Nature*, 476, 429–433. <https://doi.org/10.1038/nature10343> 429–433, <https://doi.org/10.1038/nature10343>, 2011.*

Kish, L.: Survey Sampling, John Wiley and Sons, New York, 1965.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & and Neumann, C. J. (2010). J.: The International Best Track Archive for Climate Stewardship (IBTrACS). **Bull. Amer. Meteor. Soc.**, 91(3), 363–376. <https://doi.org/10.1175/2009bams2755.1> (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. (2008). K.: Cosmic rays, cloud condensation nuclei and clouds — a reassessment using MODIS ~~data~~. **Atmos. data, Atmos. Chem. Phys.**, *Phys.*, 8, 7373–7387. <https://doi.org/10.5194/acp-8-7373-2008> 7373–7387, <https://doi.org/10.5194/acp-8-7373-2008>, 2008.

Laken, B. A., Pallé, E., Calogovic, J., & and Dunne, E. M. (2012). M.: A cosmic ray–climate link and cloud ~~observations~~. **J. observations, J. Space*

Weather Space *Clim.**, *Clim.*, 2, A18. https://doi.org/10.1051/swse/2012018_A18, https://doi.org/10.1051/swsc/2012018_2012. Li, D., Xiao, Z., Xu, J., & Zhao, L. (2023). 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 the possible mechanism. *Front. Earth Sci.**, 11, 1169685. <https://doi.org/10.3389/feart.2023.1169685>. 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., & D. and Svensmark, H. (2000). H.: Low cloud properties influenced by cosmic rays. *Phys. rays, Phys. Rev. Lett.**, *Lett.*, 85, 5004–5007. <https://doi.org/10.1103/physrevlett.85.5004>, <https://doi.org/10.1103/PhysRevLett.85.5004>, 2000. Mendoza, B., & B. and Pazos, M. (2009). M.: A 22-yr hurricane cycle and its relation with geomagnetic activity. *J. Atmos. Sol.-Terr. Phys.**, *Phys.*, 71, 2047–2054. <https://doi.org/10.1016/j.jastp.2009.09.012>, 2009. Mileyko, Y., Mukherjee, S., & Harer, J. (2011). J.: Probability measures on the space of persistence diagrams. *Inverse Problems**, *diagrams, Inverse Probl.*, 27, 124007. <https://doi.org/10.1088/0266-5611/27/12/124007>. Mironova, I. A., et al. (2018). Rapid intensification of tropical cyclones in the context of the solar wind–magnetosphere–ionosphere–atmosphere coupling. *J. Atmos. Sol.-Terr. Phys.**, 182, 1–4. <https://doi.org/10.1088/0266-5611/27/12/124007>, 2011. Muszynski, G., Kashinath, K., Kurlin, V., Wehner, M., & Prabhat (2019). and Prabhat: Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets. *Geosci. Model Dev.**, *Dev.*, 12, 613–628. <https://doi.org/10.5194/gmd-12-613-2019>, 2019. Nayak, S. (2024). S.: Effects of solar variability on tropical cyclone activity. *Earth activity, Earth Space Sci.**, *Sci.*, 11, e2023EA003500. <https://doi.org/10.1029/2023ea003500>, 2024. Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A.: A roadmap for the computation of persistent homology. *EPJ Data Sci.**, *Sci.*, 6, 17. <https://doi.org/10.1140/epjds/s13688-017-0109-5>, 2017. Pérez-Peraza, J., Velasco, V., & Kavnikov, S. (2008). S.: Wavelet coherence analysis of Atlantic hurricanes and cosmic rays. *Geofísica Internacional**, 47(3), 231–244. <https://doi.org/10.22201/igeof.00167169p.2008.47.3.85>, 2008. Pierce, J. R. (2017). R.: Cosmic rays, aerosols, clouds, and climate: recent findings from the CLOUD experiment. *J. Geophys. Res.-Atmos.**, *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., & Kwitt, R. (2015). R.: A stable multi-scale kernel for topological machine learning. *Proc. learning, in: Proceedings of the IEEE Conf. Conference on Computer Vision and Pattern Recognition (CVPR)**, 4741–4748. [doi:10.1109/CVPR.2015.7299106](https://doi.org/10.1109/CVPR.2015.7299106), <https://doi.org/10.1109/evpr.2015.7299106>, 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., N. and Tralie, C., et al. (2019). C.: Scikit-TDA: topological data analysis for Python. <https://scikit-tda.org/>, Python, available at: <https://scikit-tda.org/>, 2019. 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., Hederes, 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. (1951). H.: The interpretation of interaction in contingency tables. *J. R. tables, J. Roy. Stat. Soc. B**, 13(2), 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. Steegen, 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., & Otter, N. (2022). N.: A topological perspective on weather regimes. *Climate Dynamics**, *regimes, Clim. Dynam.*, 60, 1415–1434. <https://doi.org/10.1007/s00382-022-06395-x>, 2022. Svensmark, H. (2007). H.: Cosmoclimatology: a new theory

emerges. ~~*Astron. Geophys.*~~, 48(1), 1.18–1.24. ~~<https://doi.org/10.1111/j.1468-4004.2007.48118.x>~~ 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. (2016). H.: The response of clouds and aerosols to cosmic ray decreases. ~~*J. Geophys. Res. Space Phys.*~~, *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/>, 2015. Tymochko, S., Munch, E., Dunion, J., Corbosiero, K., & and Torn, R. (2020). R.: Using persistent homology to quantify a diurnal cycle in Hurricane Felix. ~~*Pattern Recognit. Lett.*~~, *Felix, Pattern Recogn. Lett.*, 133, 137–143. ~~<https://doi.org/10.1016/j.patrec.2020.02.022>~~ 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. (2018). L.: Topological data analysis. ~~*Annu. Rev. Stat. Appl.*~~, *Appl.*, 5, 501–532. ~~<https://doi.org/10.1146/annurev-statistics-031017-100045>~~ 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., & F. and Luo, G. (2014). G.: Effect of solar variations on particle formation and cloud condensation nuclei. ~~*Environ. Res. Lett.*~~, *Environ. Res. Lett.*, 9, 045004. ~~<https://doi.org/10.1088/1748-9326/9/4/045004>~~ 045004, <https://doi.org/10.1088/1748-9326/9/4/045004>, 2014.