https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

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

Rongzhen Dai[1]

[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 geomagnetic events — that may not register in any single-storm intensity time series. We construct three principled longitude encodings 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 on encoding choice in a way that flips the qualitative verdict: perm-p_H1 = 0.130 (linear), 0.009 (unit-sphere), 0.214 (cylinder), on identical event and control sets. Second, a 49-placebo calibration returns an H_0 false-positive rate close to the nominal 5 % 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 %; 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×–2.9× regardless 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.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

1. Introduction

1.1 Solar–cyclone hypothesis: a debated observational lineage

The hypothesis that solar and geomagnetic activity perturbs tropical cyclone (TC) dynamics has a long and contested observational lineage. Elsner & Kavlakov (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. Elsner & Jagger (2008) found that US and Caribbean hurricane counts vary inversely with 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. Pérez-Peraza et al. (2008) and Mendoza & Pazos (2009) identified a ~22-year cycle in Atlantic and Pacific hurricane occurrences that coincides with the magnetic solar cycle. Hodges & Elsner (2011) confirmed a 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: 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. Across the lineage, reported effect sizes are modest, proposed mechanisms vary, and null-hypothesis 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 cosmoclimatology 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).

2

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

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 and geomagnetic activity perturbs cyclone dynamics as a genuinely debated one.

1.4 Scope of the present work: encoding-invariance as a precondition

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, more recently proposed scope applies persistent homology to multi-storm point clouds in (latitude, longitude, intensity) coordinates over all TCs occurring within an event window, with the natural geometric object becoming the joint point cloud of all event-period fixes and the natural summary statistic becoming its persistence diagram. A coherent perturbation hypothesis, if true, might leave a signature on this ensemble topology that is not visible in any single storm's trajectory. The shift in scope is not trivial.

The present paper steps one level back from that scope. We do not, in the end, advance a positive or negative physical claim about the solar–cyclone connection at the ensemble-TDA scope. Instead we ask a methodological precondition question: does the ensemble-TDA pipeline give a consistent answer about 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 that periodicity, each with its own defenders. A pipeline whose conclusions depend on this representational choice — without converging across them — is not yet ready to test the substantive physical hypothesis. We carry out that test here, and report what we find.

1.5 This paper

We construct three encodings of the (lat, lon, wind) point 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-

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

based p-values — to the same event set (25 storms across Halloween 2003, St Patrick 2015, Gannon 2024) and the same calendar-matched 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 H_1 permutation p-value depends on encoding: 0.130 / 0.009 / 0.214 for linear / sphere / cylinder respectively. The H_0 placebo false-positive rate sits near nominal 5 % in every encoding (8.2 / 4.1 / 6.1 %). The H_1 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 H_1 inflation is dimension-asymmetric and encoding-robust — i.e. it persists 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.

This analysis is observational rather than a controlled experiment. We manipulate the encoding of a periodic coordinate but not the underlying physical system, and the convergence of encoding-comparison findings reported in §4 is observational convergence, 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, 6,661 storms with valid latitude and longitude, 406,059 6-hourly fixes. The catalog is cached locally as cache/ibtracs_sat.parquet.

Event selection. Three events were chosen as one major geomagnetic-storm benchmark per recent solar cycle. Halloween 2003 (peak Dst ≈ −422 nT) is the canonical Cycle 23 extreme. St Patrick's Day 2015 (peak Dst ≈ −223 nT) is the strongest Cycle 24 storm with adequate cross-basin tropical-cyclone activity in the ±15-day window. Gannon 2024 (peak Dst ≈ −412 nT) is the strongest 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 span three recent solar cycles, sample three different tropical-cyclone seasons, and together provide the only modern-era extreme events with adequate IBTrACS overlap to support an ensemble-TDA test.

Event windows. Three peak dates anchor the event windows: Halloween storms 2003-10-29 ± 15 d; St Patrick 2015-03-17 ± 15 d; Gannon 2024-05-10 ± 15 d. A storm enters the event-period set if any of its IBTrACS fixes falls inside any of the three windows. Membership is union-based. The resulting set has 25 unique storms across five basins {WP, NI, EP, SI, SP}, with 1,880 6-hourly fixes collectively, and these counts are held

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

fixed across all encoding variants reported in §4 so that any across-encoding differences in test outcome cannot be attributed to differences in which storms 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 calendar choices that 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 × event distribution of the 25 event-period storms. Each row counts unique storms whose lifetime overlaps the ±15-day peak window of the corresponding event. "Other" captures storms whose IBTrACS BASIN entry uses the legacy SH code or is missing.

Event	Total	WP	NI	EP	SI	SP	Other
Halloween 2003	11	6	1	1	1	0	2
St Patrick 2015	9	2	0	0	4	3	0
Gannon 2024	5	1	1	0	3	0	0
Union (unique)	25	9	2	1	8	3	2

Control pool. For each event, the same calendar window in every other year between 1981 and 2024 (excluding the event year ±2 to avoid solar-cycle phase contamination). Pooled across the three calendar windows, 1,020 unique control storms with 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.

3. Methods

3.1 Persistent homology: a primer

Persistent homology is a tool from topological data analysis (TDA) that summarises the multi-scale geometric structure of a point cloud (Edelsbrunner & Harer, 2010; Carlsson, 2009; Ghrist, 2008; Wasserman, 2018). Given a finite set of points X in a metric space and a non-negative scale parameter ε, the Vietoris–Rips complex VR(X, ε) 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, ε) grows monotonically. This nested family of complexes is the Vietoris–Rips filtration. Topological features — connected components in dimension zero (H_0), one-dimensional loops in dimension one (H_1), higher-dimensional voids in dimension k — are born at some scale ε_birth and die at some larger scale ε_death. The collection of (ε_birth, ε_death) pairs for each feature is the persistence diagram (PD) of the filtration. Features with large death-minus-birth persist

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

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. 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 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. We restrict our analysis to H_0 and H_1 because higher-dimensional features become combinatorially expensive without obvious physical interpretation at the present sample sizes.

3.2 Distance between persistence diagrams

To quantify how different two persistence diagrams are, we use the Wasserstein-2 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 a complete and separable metric space. This result licenses 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 over all matched feature pairs, and therefore reflects ensemble-level differences rather than only the worst-case feature. The Wasserstein distance between diagrams is computed with persim 0.3.8 (Saul et al., 2019).

3.3 Three longitude encodings of the point cloud

Each storm contributes its 6-hourly fixes as vectors in (latitude, longitude, wind-speed) space. To bring the three dimensions onto comparable footing for a joint Vietoris–Rips filtration, each must be mapped to [0, 1]. Latitude is bounded, and may be linearly rescaled with no representational hazard. Wind speed is non-negative and bounded above by physical limits; we divide by 200 kt and clip. Longitude is periodic. Three encodings are tested.

3.3.1 Linear modular encoding

The simplest choice maps lon ∈ [−180°, 180°] to lon_n = (lon + 180) / 360 ∈ [0, 1] directly. Latitude maps to lat_n = (lat + 90) / 180; wind to wind_n = WMO_WIND / 200 (clipped). The point cloud lives in 3-D. The pipeline takes (lat_n, lon_n, wind_n) as joint coordinates in a 3-D box. This encoding has the virtue of dimensional minimality and direct interpretability. Its known defect is that the metric does not respect the periodicity of longitude: two storms at lon = −179° and lon = +179° are physically 2° apart on the globe but are placed 358°/360° ≈ 0.994 apart in the embedded representation. For cyclone trajectories that span multi-basin event windows, this discontinuity at the antimeridian is encountered frequently in the data.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

3.3.2 Unit-sphere encoding

A more geometrically natural choice maps (lat, lon) to the unit sphere $S^2 \subset R^3$ via spherical coordinates: x = (cos(lat) cos(lon) + 1) / 2, y = (cos(lat) sin(lon) + 1) / 2, z = (sin(lat) + 1) / 2, each component rescaled to [0, 1]. Wind is appended as a fourth coordinate w_n = WMO_WIND / 200 (clipped). The point cloud lives in 4-D, with 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 redundancy of a flat (cos lon, sin 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.

3.3.3 Cylinder encoding

A third choice puts lat on a linear axis and lon on a flat (cos, sin) circle: lat_n = (lat + 90) / 180, cos_n = (cos(lon) + 1) / 2, sin_n = (sin(lon) + 1) / 2, w_n = WMO_WIND / 200 (clipped). The point cloud lives in 4-D, with the geographical position constrained to the surface of a cylinder $S^1 \times R \subset R^3$. Like the sphere, this handles 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, 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 can populate a near-circular ring in the (cos_n, sin_n) plane, which the Vietoris–Rips filtration is prone to register as a persistent H_1 loop having no physical content beyond the periodicity of longitude itself.

3.3.4 Construction and sub-sampling

Each storm's 6-hourly fixes are normalized under one of the three encodings above, then all storms in the relevant ensemble are concatenated. The point cloud is sub-sampled uniformly at random to N_SUB = 600 points, both to keep ripser computation tractable and to standardize the point count of the filtration across encodings. The seed is set once at script entry (np.random.default_rng(2026)) and the RNG is then consumed sequentially by every subsample draw and every resampling iteration within the script, so that no two iterations within a run share 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_SUB ∈ {400, 600, 800, 1000} for each encoding.

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 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 and cylinder both add a fourth coordinate to accommodate the periodic longitude. Differences in test outcome between linear and the two 4-D encodings

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

therefore confound longitude representation with ambient-dimension choice. We do not separately rerun linear in a 4-D form (for example by appending a constant zero coordinate or a Gaussian-noise nuisance dimension), which would be the natural follow-up to isolate the encoding effect from the dimensionality effect. A partial mitigation is that the H_0 false-positive rate 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, the confound is real 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 $S^1 \times R$ has non-trivial first homology, $H_1(S^1 \times R) = \mathbb{Z}$. 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 from the observed H_1 features. A more demanding analysis would compute persistence on a null embedding (e.g. random longitudes preserving latitude 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 of a single real draw.

These caveats notwithstanding, 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 a factor 2.9× nominal; and the dimension-asymmetric pattern of H_0 calibrated versus H_1 inflated recurs across encodings, stratification cells (§4.3), and subsample sizes (§4.4). Any successor pipeline must reckon with the diagnostic regardless of whether the specific ambient-dimension and embedding-induced-homology contributions can be cleanly separated from the encoding choice in future work.

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,

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

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 TDA in atmospheric and cyclone studies: prior applications

Within atmospheric science, persistent homology has been used to recognise atmospheric river patterns in satellite-derived water-vapour fields (Muszynski et al., 2019); to quantify a diurnal cycle in tropical-cyclone cloud fields from geostationary infrared imagery (Tymochko et al., 2020); and to characterise weather regimes through the topology of the atmospheric attractor (Strommen et al., 2022). The present application differs from each. Muszynski et al. operate on gridded fields rather than point clouds. Tymochko et al. operate on single-storm imagery in image space. Strommen et al. operate on attractor reconstructions in phase space. Our scope is the joint point cloud of multiple storms' fixes in geographical–intensity space. We are not aware of prior work that has tested the encoding-invariance question explicitly at the cyclone-ensemble TDA scope, which is the gap the present paper addresses.

3.6 Test statistic

For each encoding, the same downstream statistic is used. The event ensemble's persistence diagram is compared against a bootstrap distribution of control PDs by mean Wasserstein-2 distance from the event PD to a reference set of 30–50 control PDs drawn at random from the full bootstrap pool. The null distribution is the same statistic computed with each control PD substituted for the event PD (jackknifing the reference set). The permutation p-value is the fraction of null draws whose distance equals or exceeds the event distance.

3.7 Pipeline runs

For each encoding we report four pipeline outputs. The D1 pool fuses all three events: 25 event storms vs 1,020 control storms, with 1,000 bootstrap iterations. The 50-placebo calibration uses the same calendar windows (Oct 29 / Mar 17 / May 10) at random years between 1985 and 2020 outside the real event years ±2, with 100 bootstrap iterations per placebo (one placebo is skipped for too few storms). Stratified attribution disaggregates the D1 pool into two basin cells (D2_WP, D2_SI) and three single-event cells (D3_Halloween, D3_StPatrick, D3_Gannon), with 500 bootstrap iterations each and Benjamini–Hochberg multiple-comparison correction. Subsample sensitivity reruns the D1 pool at $N_{SUB} \in \{400, 600, 800, 1000\}$ 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

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

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

Python 3.11; numpy 1.26, scipy 1.13, pandas 2.2, matplotlib 3.9, ripser 0.6.14 (Bauer, 2021), persim 0.3.8 (Saul et al., 2019). Vietoris–Rips persistent homology is computed to dimension 1 via ripser. Otter et al. (2017) survey alternative computational pipelines. Source scripts (scripts/11_event_ensemble_tda.py, 12_placebo_test.py, 13_stratified_attribution.py, 14_subsample_sensitivity.py) with the three encodings as switchable variants, and run logs derived/{11,12,13,14}_{4d,sph,cyl}_run.log containing every numerical result reported below, are catalogued in paper_claims_verified.md (parent directory).

---

4. Results

4.1 D1 pool: signal flips with encoding

Table 1 reports the D1 pool result under all three encodings on the same event and control sets.

Table 1. D1 pool ensemble TDA under three longitude encodings. Each row: event-PD Wasserstein-2 distance to the mean control PD; bootstrap null mean over 1,000 control draws of 25 storms each; permutation p-value as 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.

Encoding	event H0 dist	null mean H0	perm-p H0	event H1 dist	null mean H1	**perm-p H1**
Linear modular	2.140	1.641	0.088	0.538	0.492	**0.130**
Unit-sphere	2.256	1.719	0.082	0.646	0.503	**0.009**
Cylinder	1.857	1.613	0.190	0.529	0.500	**0.214**

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),

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

with the event H_1 distance sitting essentially at the null mean, a pattern consistent with the unit-sphere positive being tied to properties of the sphere metric rather than 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.

Encoding	H_0 false-positive	H_0 verdict	H_1 false-positive	H_1 verdict
Linear modular	4 / 49 = 8.2 %	within Wilson 95 %	4 / 49 = 8.2 %	within Wilson 95 %
Unit-sphere	2 / 49 = 4.1 %	within Wilson 95 %	5 / 49 = 10.2 %	within Wilson 95 %, upper
Cylinder	3 / 49 = 6.1 %	within Wilson 95 %	7 / 49 = 14.3 %	at Wilson 95 % upper bound

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×) to unit-sphere (2.0×) to cylinder (2.9×). 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.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

This has direct implications for the §4.1 D1 verdict. The unit-sphere D1 perm-p_H1 = 0.009 must be read against an H_1 calibration that produces false positives at ≈ 2× nominal. Adjusting upward by the calibration factor (0.009 × 10.2/5.0 ≈ 0.018) still leaves a value below 0.05, so the qualitative signal does not vanish, but the marginal-significance reading is degraded. More importantly, the calibration finding undermines the precondition 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×–2.9× over nominal.

Three further diagnostics. First, across the 49 cylinder placebos, the H_1 event-control distance has CV ≈ 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 ≈ 0.48 across placebos, versus CV ≈ 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 α = 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 ≥ 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_H1 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.

Cell	n event	n control	Linear H1	Sphere H1	Cylinder H1
D2_WP	9	240	0.046 (BH 0.140)	0.124 (BH 0.310)	0.058 (BH 0.145)
D2_SI	9	194	0.722	0.894	0.488
D3_Halloween	11	519	0.302	0.238	0.400
D3_StPatrick	9	314	0.056 (BH 0.140)	0.112 (BH 0.310)	0.044 (BH 0.145)
D3_Gannon	5	187	0.704	0.926	0.736

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

Two cells (D2_WP, D3_StPatrick) sit near or below the uncorrected α = 0.05 threshold under linear and cylinder encodings, and well above α under sphere. After BH correction across the five cells, no cell crosses α in any encoding. The cell that returns 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 under every encoding — H_0 is uniformly null at the cell scale. H_1 cell-level perm-p values straddle α at the uncorrected level in some encoding-cell combinations, but reach BH significance in none.

The pooled D1 H_1 verdict under any single encoding (0.130 / 0.009 / 0.214) is not localizable to any cell under that encoding's own disaggregation. Combined with the encoding sensitivity 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_SUB ∈ {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_4d_run.log, derived/14_sph_run.log, derived/14_cyl_run.log.

N_SUB	Linear modular	Unit-sphere	Cylinder
400	0.420	0.288	0.302
600	0.124	0.008	0.182
800	0.144	0.004	0.192
1000	0.590	0.050	0.424

No encoding yields a monotonic dependence on N_SUB. Under linear, the smallest perm-p is at N_SUB = 600 (0.124, still null); under sphere, 800 (0.004); under cylinder, 600 (0.182). The N = 1000 row weakens substantially in every encoding, including sphere where it sits at the boundary of α = 0.05. The pattern is consistent with a noise channel whose realisation depends on which subsample the seed happens to draw, rather than a stable signal as a function 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) and the sphere N_SUB ∈ {600, 800} sub-sample p-values {0.008, 0.004} (Table 4) jointly invite a positive narrative: low p-values robust to sub-sampling at mid-range N. Read in the context of Tables 1, 2, 3, 4 jointly, the same numbers read very differently. The linear 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 to a single cell; and the N_SUB = 1000 row degrades sharply

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

in every encoding including sphere. The sphere sub-table is a coherent island inside an otherwise incoherent multi-encoding picture. A single-encoding ensemble TDA pipeline that selected sphere a priori would have reported a positive signal that the broader picture does not support.

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 H_1: a defensible variation in how longitude is represented can flip the D1 verdict. And it is dimension-asymmetric in calibration: H_0 holds nominal alpha across all three encodings tested in §4.2, while H_1 over-rejects by 1.6× to 2.9× 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 alpha to begin with.

5.1 The D1 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 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, would have produced a paper claiming the answer to the same question with three different verdicts and three different sets of consequences for 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 D1 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, and without doing so under transparent caveat that the answer is encoding-conditional.

This may seem like a methodological house-keeping observation. We argue it is more than that. In TDA the stability theorem of Cohen-Steiner et al. (2007) is sometimes read as guaranteeing that small input perturbations produce small output perturbations. That is true under the input metric the theorem refers to. The theorem says nothing about the choice of metric itself. Re-encoding a periodic coordinate (linear vs sphere vs cylinder) is

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

not a “small perturbation of the input” — it is a choice of which metric to use. The stability theorem is silent on this choice. The encoding-flip we observe here 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× nominal at the friendliest end (linear, 8.2 %) and at 2.9× 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, 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 a population of short-lived clusters whose null distribution is broader and less easily pierced.

The same reading would apply to any pipeline that compares Wasserstein-2 distances between event and control persistence diagrams on this class of data: the relevant statistic is what fraction of the bootstrap null distribution lies below the event value, and the H_1 null is too narrow to support nominal calibration. The encoding choice (linear / sphere / cylinder) modulates the geometry of the loops but does not eliminate the narrowness of the null. Hence the H_1 over-rejection persists across encodings.

If this reading is correct, then it is not a defect specific to one encoding or one bug to be patched. It is a structural property of the test as applied to cyclone-trajectory data at the present sample sizes. Any future positive H_1 claim from this pipeline must either (a) demonstrate H_1 calibration to nominal alpha under the specific encoding being used, by an explicit placebo test that the authors include in the published analysis, or (b) use a test statistic whose null distribution is independently shown to be wide enough to support the calibration. We are not aware of prior ensemble-TDA cyclone-trajectory work that has met either bar.

5.3 The solar–cyclone case study at the ensemble-TDA scope

What follows from this for the substantive physical question? Under the linear and cylinder encodings, D1 H_1 is null. Under the sphere encoding, D1 H_1 is nominally below α but the H_1 placebo channel from which that verdict reads is 2.0× over nominal, and the result does not survive stratification or, on a strict reading, survive sub-sample sensitivity past N_SUB = 800. Across encodings, no cell-level stratification reaches BH-corrected significance, and Halloween 2003 — the canonical extreme — returns perm-p_H1 between 0.238 and 0.400 in every encoding, indistinguishable from the null.

The cleanest reading we can defend is that the ensemble-TDA pipeline in its present form does not provide adequate evidence either for or against a solar–cyclone perturbation

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

effect at the ensemble-topology scope. That is a weaker statement than an earlier single-encoding version of this analysis at the same scope (Dai, in revision) reported under a single encoding choice. The multi-encoding analysis appears to require it.

We emphasise the scope qualifier. This null verdict at the ensemble-TDA scope is consistent with both the older positive literature at the per-storm intensity-correlation scope (Elsner & Kavlakov, 2001; Elsner & Jagger, 2008; Hodges & Elsner, 2011) and the null literature at the microphysical CRC scope (Kristjánsson et al., 2008; Calogovic et al., 2010; Laken et al., 2012; IPCC AR6 WGI, 2021). The three scopes interrogate different observational signatures, and a null at one scope does not constitute evidence either way at another. Our finding is specifically that the ensemble-TDA scope, as implemented in the current pipeline, is not yet a tool that can answer the question. It can answer it once an encoding-invariant H_1-calibrated protocol is in place.

This conclusion rests on observational convergence across three encoding variants rather than on a controlled experiment. Causal isolation of any physical solar–cyclone effect at the ensemble-topology scope would require an independent experimental or quasi-experimental design — for example, a synthetic-event injection study with known perturbation strength applied to surrogate IBTrACS data — beyond the scope of the present methodological study.

5.4 A minimum protocol for future ensemble-TDA on cyclone trajectories

We propose three pre-registration requirements for future ensemble-TDA studies that aim to test perturbation hypotheses on cyclone-trajectory data.

First, multi-encoding agreement testing. Any positive ensemble-TDA H_1 claim should be reported under at least two principled encodings of the periodic geographical coordinates, with the verdicts compared. A single-encoding claim is not sufficient. Two encodings agreeing is necessary but not sufficient; three encodings agreeing is stronger evidence. Disagreement across encodings is direct evidence that the claim is encoding-dependent and should not be advanced as a positive physical finding without further investigation of why the encodings disagree.

Second, dimension-resolved placebo calibration. The placebo set should be large enough to estimate false-positive rate per dimension (H_0 separately from H_1) to within reasonable confidence at the alpha being claimed. At $n = 49$ placebos and $\alpha = 0.05$, the Wilson 95 % CI for nominal calibration is approximately [1.4 %, 14.7 %], wide enough that values inside it are formally compatible with nominal. A more demanding standard would use $n = 200$ placebos to narrow the Wilson CI and resolve H_1-specific inflation more sharply.

Third, stratified attribution at the natural physical scale. Pooled-significant-but-stratified-null is a recurring failure mode in climate-statistics composites. It belongs to the same family as Simpson's-paradox effects in conventional regression (Simpson, 1951), where a marginal aggregate signal disaggregates into subgroup signals of opposite sign. Pre-registration of a stratification scheme (basin × event × season) before the pooled test result is read protects against post-hoc carving of the pool to support whichever cell happens to come out significant.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

5.5 A note on topological aggregation instability

A further point sharpens why the stratification step in our protocol is more essential, not less, than in regression-based composite analyses. Topology, as encoded in a persistence diagram, is a *global* summary statistic of the entire point cloud: the birth and death scales of every connected component and every loop are functions of the joint geometry of all points simultaneously. Pooling two point clouds is therefore not equivalent to summing two persistence diagrams. Inserting subgroup points into a shared filtration can create new loops at scales where neither subgroup alone supported them, and can destroy loops that each subgroup carried independently. Subgroup cancellation, in the topological setting, is not linearly inspectable from the pooled diagram alone. We label this failure mode *topological aggregation instability* (a working term introduced here): a small misalignment between subgroup geometries, combined with the presence of sparse cross-subgroup fixes that bridge the gap, can produce a pooled persistence diagram with features that have no consistent attribution to any subgroup, and that vanish when the pool is dissected. Figure 1 illustrates the geometry schematically.

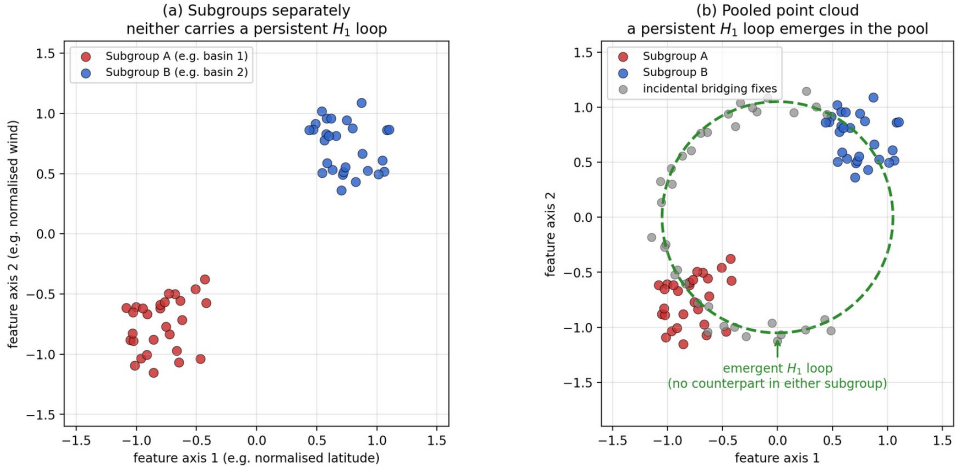

Figure 1

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, 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 no counterpart in either subgroup considered separately, and disappears under disaggregation. Pooling is not a linear operation on persistence diagrams; subgroup cancellation in the topological setting is not inspectable from the pooled diagram alone.

This figure shows why H_1 specifically 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 in the

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

H_0 channel rather than opening one, so the H_0 null distribution is broader and less easily pierced by a modest outlier event ensemble. H_1 carries 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 the label *topological aggregation instability* as a working descriptive term for 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 on the real line, 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 et al., 2015), with the kernel bandwidth offering a tunable handle on null variance. A simpler first-pass attempt would normalize the H_1 statistic itself: dividing the event H_1 distance by an estimate of the control bootstrap standard deviation transforms a raw distance into a z-score and may widen the null without changing 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 a $Kp \geq 8$ threshold (approximately 50 events). Sample size alone is not the cure. Narrowness of the H_1 null is a shape property; a larger 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× sample increase would tighten the Wilson 95 % CI on the placebo false-positive rate by approximately $\sqrt{2}$, allowing finer resolution of any residual encoding-specific calibration bias.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

The third pathway replaces 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 to a single computational choice rather than to ensemble TDA on cyclone trajectories in general.

We frame the present paper as 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 the same data, 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 and motivates the search for a fix by us or by other groups.

6. Conclusions

We applied an ensemble-TDA pipeline to test whether three benchmark extreme geomagnetic events (Halloween 2003, St Patrick 2015, Gannon 2024) perturb the topology of the cyclone activity in their immediate ±15-day windows. The result is mixed. Three principled longitude encodings of the (lat, lon, wind) point cloud — linear modular, unit-sphere, and lat-linear-plus-longitude-cylinder — give three qualitatively different D1 H_1 perm-p values (0.130, 0.009, 0.214) on the same event set and the same control pool, demonstrating that the pooled-ensemble test verdict in this pipeline depends on the choice of representation 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 %), while the H_1 false-positive rate 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 at n = 49). Stratified attribution and subsample-size sensitivity recur the same pattern: H_1 results that disagree across encodings, no cell-level signal surviving BH correction in any encoding, no monotonic dependence 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 point-cloud TDA on geospatial trajectories with periodic coordinates is more fragile than its formal stability theorems suggest. We propose 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 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.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

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) with the three encodings as switchable variants (normalize_pc_sphere, normalize_pc_cylinder, and the in-line linear modular form), and derived run logs (derived/{11,12,13,14}_{4d,sph,cyl}_run.log) reproducing every numerical value reported in Tables 1–4 and Table E1, are available from the corresponding author on reasonable request.

Author contributions

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

Competing interests

The author declares no competing interests.

Acknowledgements

Analysis was performed using Anthropic Claude Code as a coding assistant, with all design decisions, statistical interpretations, encoding-comparison framing, and writing by 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. IBTrACS data are provided by NOAA NCEI.

References

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., & Ziegelmeier, L. (2017). Persistence images: a stable vector representation of persistent homology. *J. Mach. Learn. Res.*, 18, 1–35.

Bauer, U. (2021). Ripser: efficient computation of Vietoris-Rips persistence barcodes. *J. Appl. Comput. Topol.*, 5, 391–423.

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16, 77–102.

Calogovic, J., Albert, C., Arnold, F., Beer, J., Desorgher, L., & Flueckiger, E. O. (2010). Sudden cosmic ray decreases: no change of global cloud cover. *Geophys. Res. Lett.*, 37, L03802.

Carlsson, G. (2009). Topology and data. *Bull. Amer. Math. Soc.*, 46(2), 255–308.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1), 103–120.

Edelsbrunner, H., & Harer, J. L. (2010). *Computational Topology: An Introduction*. American Mathematical Society, Providence, RI.

Elsner, J. B., & Jagger, T. H. (2008). United States and Caribbean tropical cyclone activity related to the solar cycle. *Geophys. Res. Lett.*, 35, L18705.

Elsner, J. B., & Kavlakov, S. P. (2001). Hurricane intensity changes associated with geomagnetic variation. *Atmos. Sci. Lett.*, 2, 86–93.

Ghrist, R. (2008). Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.*, 45(1), 61–75.

Ver Hoef, L., Adams, H., King, E. J., & Ebert-Uphoff, I. (2023). A primer on topological data analysis to support image analysis tasks in environmental science. *Artif. Intell. Earth Syst.*, 2(1), AIES-D-22-0039.1.

Hodges, R. E., & Elsner, J. B. (2011). Evidence linking solar variability with US hurricanes. *Int. J. Climatol.*, 31, 1897–1907.

Hodges, R. E., Jagger, T. H., & Elsner, J. B. (2014). The sun–hurricane connection: diagnosing the solar impacts on hurricane frequency over the North Atlantic basin using a space–time model. *Nat. Hazards*, 73, 1063–1081.

Huth, R., & Stryhal, J. (2024). Association of the 11-year solar cycle with correlation and teleconnection structures in tropospheric circulation. *J. Space Weather Space Clim.*, 14, 31.

IPCC AR6 WGI (2021). *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.)], Cambridge University Press, Chapter 7.

Kirkby, J., Curtius, J., Almeida, J., et al. (2011). Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation. *Nature*, 476, 429–433.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.*, 91(3), 363–376.

Kristjánsson, J. E., Stjern, C. W., Stordal, F., Fjæraa, A. M., Myhre, G., & Jónasson, K. (2008). Cosmic rays, cloud condensation nuclei and clouds — a reassessment using MODIS data. *Atmos. Chem. Phys.*, 8, 7373–7387.

Laken, B. A., Pallé, E., Calogovic, J., & Dunne, E. M. (2012). A cosmic ray–climate link and cloud observations. *J. Space Weather Space Clim.*, 2, A18.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

Li, D., Xiao, Z., Xu, J., & Zhao, L. (2023). Asymmetric modulation of solar activity on tropical cyclone frequency over the western North Pacific and the possible mechanism. *Front. Earth Sci.*, 11, 1169685.

Marsh, N. D., & Svensmark, H. (2000). Low cloud properties influenced by cosmic rays. *Phys. Rev. Lett.*, 85, 5004–5007.

Mendoza, B., & Pazos, M. (2009). A 22-yr hurricane cycle and its relation with geomagnetic activity. *J. Atmos. Sol.-Terr. Phys.*, 71, 2047–2054.

Mileyko, Y., Mukherjee, S., & Harer, J. (2011). Probability measures on the space of persistence diagrams. *Inverse Problems*, 27, 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.

Muszynski, G., Kashinath, K., Kurlin, V., Wehner, M., & Prabhat (2019). Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets. *Geosci. Model Dev.*, 12, 613–628.

Nayak, S. (2024). Effects of solar variability on tropical cyclone activity. *Earth Space Sci.*, 11, e2023EA003500.

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Sci.*, 6, 17.

Pérez-Peraza, J., Velasco, V., & Kavlakov, S. (2008). Wavelet coherence analysis of Atlantic hurricanes and cosmic rays. *Geofísica Internacional*, 47(3), 231–244.

Pierce, J. R. (2017). Cosmic rays, aerosols, clouds, and climate: recent findings from the CLOUD experiment. *J. Geophys. Res. Atmos.*, 122, 8051–8055.

Reininghaus, J., Huber, S., Bauer, U., & Kwitt, R. (2015). A stable multi-scale kernel for topological machine learning. *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 4741–4748. doi:10.1109/CVPR.2015.7299106.

Saul, N., Tralie, C., et al. (2019). Scikit-TDA: topological data analysis for Python. https://scikit-tda.org/.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *J. R. Stat. Soc. B*, 13(2), 238–241.

Strommen, K., Chantry, M., Dorrington, J., & Otter, N. (2022). A topological perspective on weather regimes. *Climate Dynamics*, 60, 1415–1434.

Svensmark, H. (2007). Cosmoclimatology: a new theory emerges. *Astron. Geophys.*, 48(1), 1.18–1.24.

Svensmark, J., Enghoff, M. B., Shaviv, N. J., & Svensmark, H. (2016). The response of clouds and aerosols to cosmic ray decreases. *J. Geophys. Res. Space Phys.*, 121, 8152–8181.

https://doi.org/10.5194/egusphere-2026-2975
Preprint. Discussion started: 23 June 2026

Tymochko, S., Munch, E., Dunion, J., Corbosiero, K., & Torn, R. (2020). Using persistent homology to quantify a diurnal cycle in Hurricane Felix. *Pattern Recognit. Lett.*, 133, 137–143.

Wasserman, L. (2018). Topological data analysis. *Annu. Rev. Stat. Appl.*, 5, 501–532.

Yu, F., & Luo, G. (2014). Effect of solar variations on particle formation and cloud condensation nuclei. *Environ. Res. Lett.*, 9, 045004.