ORIENTATION-AUGMENTED ROUTE TOPOLOGICAL REPRESENTATION FOR TRAJECTORY CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

For traffic behavior analysis, trajectory clustering requires representations that make raw movement sequences comparable while suppressing irrelevant geometric variation and retaining the spatial structure needed to distinguish travel routes. Topological trajectory representations describe how paths wind around obstacles or landmarks, but flattened harmonic embeddings may not distinguish routes with similar cumulative responses and different landmark-relative orientations. We propose **Azimuth-Resolved Harmonic Representation** (ARHR), which supplements the cumulative harmonic embedding with a compact summary of landmark-relative angular activity. Conditional analysis gives sufficient conditions for this angular summary to distinguish trajectories with equal cumulative harmonic responses. Experiments on real-world traffic data show dataset-dependent improvements over the evaluated flattened-harmonic baseline. Weighting controls and order ablations show that clustering performance depends on the activity definition, while higher angular orders do not consistently improve clustering across datasets. These findings clarify when angular activity helps harmonic trajectory representations and support retaining this complementary structure according to the route distinctions required by the task.

1 INTRODUCTION

Trajectory clustering groups mobility traces into recurring route patterns and is fundamental to traffic behavior analysis. Representations used for clustering should attenuate the effects of sampling differences and local geometric fluctuations, while retaining the spatial structure needed to distinguish different travel routes. The same route can yield markedly different position sequences under different observation regimes, and spatially nearby trajectories may still belong to different route classes. The key for trajectory representation, therefore, is to decide which variations can be ignored and which path differences must be retained, so that subsequent analysis can focus on task-relevant structure (Li et al., 2018).

Prior work addresses trajectory comparability through geometric similarity methods and deep trajectory representation learning. The former directly measure differences among sampled points, curves, or local segments—for example, TRACLUS discovers shared subpaths via partition-and-group aggregation (Lee et al., 2007)—whereas the latter map movement sequences to vectors through training; for instance, t2vec uses sequence representation learning to cope with low sampling rates and noise, and TrajCL improves robustness via trajectory augmentation and contrastive learning (Li et al., 2018; Chang et al., 2023). For tasks that concern global winding patterns, coordinate differences induced by local deformations need not imply a change of path class. What is needed, therefore, is a representation that can downweight such variations while preserving overall path structure.

Topological trajectory representations organize similarity by how paths wind relative to obstacles, holes, or landmarks, and thereby provide a direct route to ignoring local deformations that do not alter the relevant topological relations. Accumulating topological responses along a path yields compact vectors that are convenient for comparison; on discrete maps, Hodge-based methods further construct flattened embeddings via harmonic projection and develop the choice of landmark

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

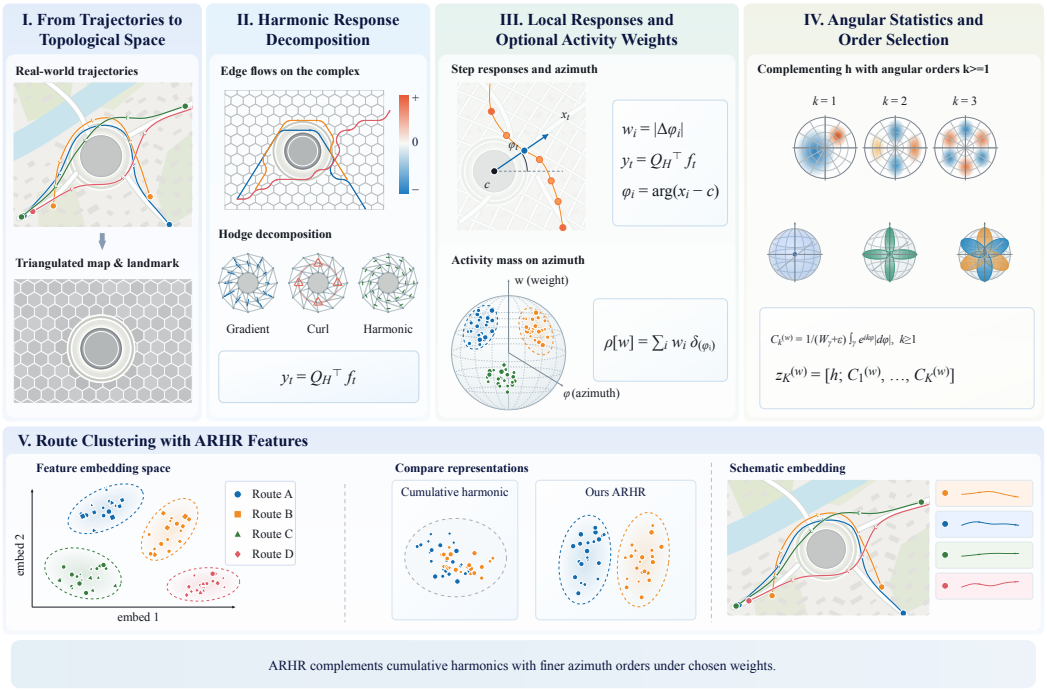

Figure 1: Overview of orientation ambiguity and ARHR. I–II: trajectories are mapped to a triangulated complex and yield local harmonic responses y_t. III–IV: landmark-relative azimuth ϕ_i and optional activity weights w_i (main setting $w_i = |\delta_i|$) define angular coefficients $C_k^{(w)}$ that characterize angular structure for $k \geq 1$, which are combined with the cumulative harmonic embedding h in the truncated descriptor. V: the schematic embedding illustrates how the augmented descriptor can separate route distinctions that are unresolved by cumulative harmonics alone.

references (Yin et al., 2017; Ghosh et al., 2018; Frantzen et al., 2021; Grande et al., 2024). These studies advance the construction, aggregation, and reference selection of topological responses. Yet when the map and local harmonic responses are given, cumulative responses may still fail to separate trajectories that share similar winding magnitudes but differ in azimuthal configuration.

We call this representational collision *orientation ambiguity*. Trajectory motion occurs at specific azimuths around a landmark, whereas the cumulative harmonic vector does not explicitly represent how angular activity is distributed over azimuth. Consequently, trajectories with similar cumulative harmonic responses can still differ in where angular activity occurs around the landmark. The issue is not that harmonic responses misdescribe winding, but that the resolving power of a fixed aggregation level is insufficient for the path distinctions required by the task. We therefore take the landmark-relative distribution of angular activity as complementary information for describing these route differences, and study how angular statistics can supplement cumulative harmonic responses for the path distinctions required by the task.

Accordingly, we propose **Azimuth-Resolved Harmonic Representation** (ARHR), which enriches harmonic trajectory representations with the landmark-relative distribution of angular activity. Circular Fourier statistics describe where this activity occurs around a landmark, complementing cumulative responses with azimuthal structure relevant to route discrimination. Finite-order truncation maps each trajectory to a fixed-dimensional, semantically explicit real vector; when the map and local response computation are held fixed, the unnormalized coefficients support incremental updates and segment composition. Figure 1 summarizes orientation ambiguity, the construction of ARHR, and the role of angular statistics in route discrimination.

We examine the identifiability and practical utility of the representation through conditional analysis, controlled simulations, and real trajectory data. Experiments take an angular-activity–weighted

low-order descriptor and its order extensions as the evaluation objects, and compare them with topological baselines, geometric methods, and deep representations. Results show that retaining landmark-relative azimuthal structure mitigates the examined trajectory collisions and, without additional representation-learning training, supports standard clustering at low per-trajectory feature-extraction cost. Further analyses of weights, angular order, and harmonic configuration indicate that discriminative gains depend on the correspondence between activity and occurrence azimuth; raising the angular order need not increase clustering gains; and when the original harmonic representation already separates classes well, the benefit of adding azimuthal information diminishes. These findings distinguish increases in representation capacity from improvements in task discrimination.

Contributions.

1. **Azimuth-resolved topological trajectory representation.** Building on the landmark-relative distribution of angular activity, we develop order-adjustable ARHR to complement cumulative harmonic responses with angular statistics, providing a finer description of the spatial organization of angular activity.
2. **Resolving properties.** We derive azimuth-shift equivariance and a turning-point reduction that clarifies what the angular statistics retain beyond cumulative harmonic responses.
3. **Practical discrimination and computational utility.** Through controlled simulations and comparisons on real data, we test whether low-order descriptors resolve trajectory collisions and convert that separation into clustering gains, while assessing computational cost.
4. **Capacity versus effective discriminative information.** By comparing different angular orders and harmonic configurations, we study when added information improves task discrimination and provide a basis for choosing an appropriate representation level.

2 Related Work

2.1 Trajectory Similarity and Representation

One line of work compares trajectories through sampled points or local segments. Classical geometric distances include Fréchet (Alt & Godau, 1995; Eiter et al., 1994), Hausdorff (Huttenlocher et al., 1993), dynamic time warping (Berndt & Clifford, 1994), and longest common subsequence (Vlachos et al., 2002). Lee et al. (2007) propose TRACLUS, a partition-and-group framework that discovers common sub-trajectory patterns via density-based clustering of representative line segments (Ester et al., 1996). Another line learns vector embeddings so that similarity is less sensitive to non-uniform sampling and noise, including seq2seq trajectory encoders (Li et al., 2018), contrastive dual-attention encoders (Chang et al., 2023), and end-to-end deep trajectory clustering (Fang et al., 2021; Yue et al., 2019). Related unsupervised temporal clustering methods jointly learn representations and partitions for multivariate series (Madiraju, 2018; Ma et al., 2019; Liu & Chen, 2024). These approaches address geometric or learned comparability. We focus on a narrower representation gap: when a path is compressed into a cumulative winding response relative to map landmarks, how to retain the landmark-relative orientation needed for absolute-route separation.

2.2 Topological Trajectory Representations

Topology-based methods use holes or obstacles as references and accumulate winding-related quantities along a path (Yin et al., 2017; Ghosh et al., 2018). Most directly related is the flow-embedding approach of Frantzen et al. (2021), which models trajectories as signed edge flows on a simplicial complex and projects them with harmonic eigenvectors of the Hodge 1-Laplacian (Jiang et al., 2011; Schaub et al., 2020; Barbarossa & Sardellitti, 2020; Lim, 2020). The framework distinguishes a stepwise harmonic embedding from a fixed-length *flattened* cumulative embedding; the latter is commonly used for downstream tasks such as anomaly detection. Broader geometric and topological deep learning studies signals on graphs and higher-order complexes (Bronstein et al., 2017; 2021; Bruna et al., 2013; Kipf & Welling, 2017; Veličković et al., 2018; Xu et al., 2019; Chen et al., 2020; Defferrard et al., 2016; Gilmer et al., 2017; Morris et al., 2019; Ebli et al., 2020; Bunch et al., 2020; Bodnar et al., 2021b;a; Edelsbrunner & Harer, 2010), but typically do not target fixed-length absolute-route descriptors under orientation ambiguity.

Prior work also restores discriminability by keeping endpoints or redesigning landmarks. Yin et al. (2017) use (s, t, h) with start/end locations together with the harmonic integral. Grande et al. (2024) search for landmark configurations that improve harmonic separation when holes are insufficient. We instead keep a fixed map and the original harmonic computation, and apply ARHR with an angular-activity–weighted landmark azimuth readout. The goal is a fixed-length descriptor that jointly retains cumulative winding and where angular motion concentrates around a chosen landmark, without storing the full position sequence.

3 METHOD

ARHR complements cumulative harmonic responses with angular statistics that describe the landmark-relative distribution of trajectory activity. The following sections define these quantities and their combination into a trajectory representation. Figure 1 illustrates the construction.

3.1 LOCAL RESPONSES AND LANDMARK REFERENCE

We discretize the motion domain as a two-dimensional simplicial complex $\mathcal{K} = (V, E, F)$ and fix a landmark c inside a chosen hole. Let $Q_H \in \mathbb{R}^{|E|\times m}$ be an orthonormal basis of the kernel of the map Hodge 1-Laplacian. After map matching, a trajectory comprises T directed edge traversals. The t-th traversal is a sparse edge flow f_t, with $+1$ or -1 on the traversed edge according to whether the travel direction agrees with the reference orientation. The corresponding local harmonic response is

$$y_t = Q_H^\top f_t \in \mathbb{R}^m. \tag{1}$$

Trajectories on the same map share the harmonic basis and the landmark, so that representation coordinates remain consistent.

For an input polyline $\gamma = (x_0, \ldots, x_N)$, define the landmark-relative azimuth $\phi_i = \arg(x_i - c)$ and unwrap angles along travel. This azimuth describes the spatial bearing of a position relative to the landmark. It is not the tangential heading of the moving object. On segment i, the unwrapped endpoints determine $\delta_i = \phi_i - \phi_{i-1}$ and the mid-azimuth $m_i = (\phi_i + \phi_{i-1})/2$. For segments that do not pass through the landmark c, the angular integral depends only on these endpoint angles and is invariant to collinear subdivision of the same polyline segment. Harmonic responses are computed on map-matched edge traversals. Azimuthal activity is computed on the input polyline. We therefore index the former by t and the latter by i.

3.2 ANGULAR STATISTICS AND TRAJECTORY REPRESENTATION

The cumulative harmonic embedding aggregates the local responses,

$$h(\gamma) = \sum_{t=1}^{T} y_t = Q_H^\top f \in \mathbb{R}^m, \tag{2}$$

where $f = \sum_t f_t$ is the accumulated edge flow. This is the flattened harmonic baseline used throughout the paper.

For azimuthal activity, a nonnegative segment weight w_i assigns the contribution of local motion in the aggregate. Weights may be defined from sampling, time gaps, arc length, or angular activity. We use the absolute unwrapped increment $w_i = |\delta_i|$, so that both clockwise and counterclockwise angular motion produce nonnegative mass, while radial motion with little azimuth change is down-weighted. Write $W_\gamma = \sum_i w_i$ for the total activity.

For orders $k \geq 1$, the ARHR coefficients are the normalized segment-exact integrals of $e^{ik\phi}\,|d\phi|$ along the polyline,

$$C_k^{(w)}(\gamma) = \frac{1}{W_\gamma + \varepsilon} \sum_i w_i\, e^{ikm_i}\, \mathrm{sinc}\big(k\delta_i/2\big), \qquad k \geq 1, \tag{3}$$

where $\mathrm{sinc}(x) = \sin(x)/x$ with $\mathrm{sinc}(0) = 1$, and $\varepsilon > 0$ for numerical stability. A midpoint discretization of the same weights is used only as an ablation (Appendix A.2). When $W_\gamma = 0$, all nonzero-order coefficients are set to zero while h is retained.

Under the default weight $w_i = |\delta_i|$, segment contributions telescope within each monotone-azimuth run. In particular, if the unwrapped azimuth is monotone over the whole trajectory and $D = \phi_N - \phi_0$,

$$C_k^{(w)} = \frac{|D|}{|D| + \varepsilon} e^{ik(\phi_0 + D/2)} \operatorname{sinc}\left(\frac{kD}{2}\right), \tag{4}$$

Thus, in the monotone case the angular coefficient, denoted P_k, is determined by the endpoint azimuths and net angular span. More generally, the telescoping structure depends on endpoint and azimuth-turning-point angles together with total angular variation; information beyond the monotone endpoint-span control therefore arises from azimuth reversals. The segment-wise derivation is given in Appendix A.1.

The phase of the first-order coefficient $C_1^{(w)}$ encodes the resultant azimuth of activity, and its modulus reflects concentration versus cancellation under that statistic. Activity on different azimuths can cancel, so a near-zero modulus does not mean the trajectory has no angular motion. We keep the real and imaginary parts of each complex coefficient and do not force unit-modulus normalization.

Given a truncation order $K \geq 1$, the trajectory representation concatenates the harmonic embedding with the first K angular blocks,

$$z_K^{(w)}(\gamma) = \left[h(\gamma);\ \operatorname{Re} C_1^{(w)}(\gamma),\ \operatorname{Im} C_1^{(w)}(\gamma),\ \ldots,\ \operatorname{Re} C_K^{(w)}(\gamma),\ \operatorname{Im} C_K^{(w)}(\gamma)\right] \in \mathbb{R}^{m+2K}. \tag{5}$$

We use the low-order setting $K = 1$, which yields an $m+2$ representation. The number of available harmonic channels m is fixed by the map and landmark. The angular order K sets how finely the retained activity distribution is resolved.

3.3 Resolvability and Computation

Azimuthal resolvability. Consider a pair of trajectories γ and γ_α that share the same cumulative harmonic response, whose corresponding activity occurrence azimuths differ by a common shift α, and whose weights are unchanged. Equation (3) gives

$$h(\gamma_\alpha) = h(\gamma), \qquad C_k^{(w)}(\gamma_\alpha) = e^{ik\alpha} C_k^{(w)}(\gamma), \quad k \geq 1. \tag{6}$$

Thus, when a retained order satisfies $C_k^{(w)}(\gamma) \neq 0$ and $e^{ik\alpha} \neq 1$, the two trajectories are separated in $z_K^{(w)}$. Angular-activity weights depend on adjacent azimuth differences, so a common azimuth translation leaves their values unchanged. Identical harmonic responses can arise under corresponding map and harmonic weight symmetries; the analysis does not require every map rotation to satisfy this condition.

Figure 2 illustrates this distinction on a controlled two-hole map ($\beta_1 = 2$). The four corridor classes share one origin–destination pair and homologous winding around the landmark. Their complete harmonic coordinates strongly overlap in $\mathbb{R}^2$ (panel (b)), whereas the first-order angular coefficients separate the same trajectories by landmark-relative azimuth (panel (c)). Construction details appear in Appendix A.7.

Order and identifiability. Higher angular orders can distinguish distributions that collide at lower orders (Appendix A.1), but the additional coordinates may also capture within-class variation. Increasing K therefore need not improve clustering, and a finite-order representation does not uniquely determine the full activity distribution or event sequence.

Incremental computation. With the map, landmark, and local recording convention held fixed, h, the total activity W_γ, and the unnormalized angular statistics are all accumulated from local contributions. They support step-wise updates and merging of already segmented trajectory pieces. Normalization is applied when the representation is emitted; normalized nonzero-order coefficients themselves do not add directly. After the harmonic basis is precomputed, feature extraction costs $O(mT + KN)$, and the accumulated state stores $O(m + K)$ numbers. Map construction and harmonic-basis computation are preprocessing costs. Per-trajectory map matching is a separate stage and is excluded from the feature-extraction complexity reported here.

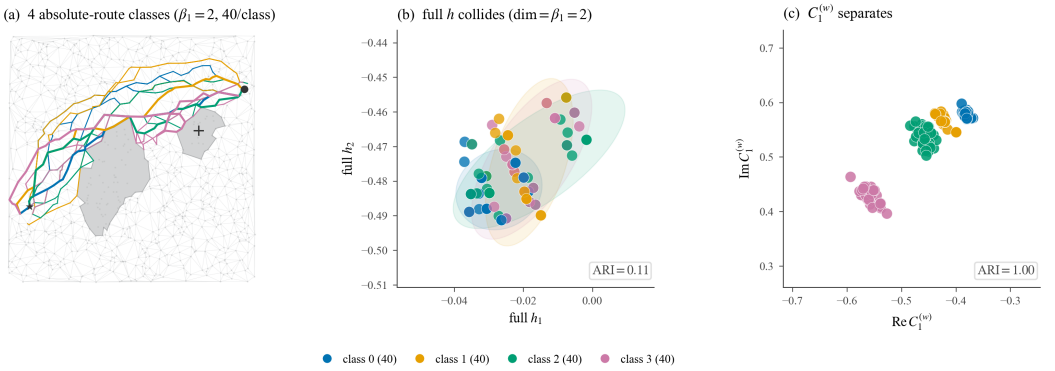

Figure 2: Controlled resolvability probe on a Frantzen-style two-hole complex (β_1=2; 40 paths per class; shared OD). (a) Four absolute-route corridor classes homologous about the landmark. (b) Full harmonic coordinates (h_1, h_2) strongly overlap across the four classes (ARI$\approx$ 0.11). (c) The same trajectories separate in the real–imaginary plane of $C_1^{(w)}$ (ARI= 1.00).

4 Experiments

We evaluate ARHR along three complementary axes: route discrimination, the role of the activity measure, and the effect of representation capacity. Section 3 analyzes the resolvability mechanism; here we examine whether the resulting representation remains useful on real trajectory collections and how its behavior changes with the representation choices defined there. The default azimuthal block is the segment-exact integral in Eq. (3), which, for segments that do not pass through the landmark, does not change under collinear subdivision of the same polyline segment. Appendix A.2 reports the sampling density of the data, the effect of collinear densification, and the effect of deleting interior vertices while keeping h fixed. Appendix A also reports coefficient properties, extended activity-measure variants, a position-noise sensitivity check, and implementation timing details. ARHR uses linear per-trajectory extraction; details in Appendix A.7.

4.1 Experimental Setup

Datasets. We use two complementary evaluation regimes. The absolute-route benchmark is built from openDD (Breuer et al., 2020), Swiss Roundabouts (Espadaler Clapés et al., 2025), 3DJunction1,[1] and DLR Urban Traffic (Schicktanz et al., 2025), covering roundabouts and signalized or multi-hole junctions. Dataset-specific entry/exit and route-selection rules define the evaluated route classes; consequently, endpoint information can be highly informative in this benchmark. We therefore include explicit endpoint baselines and use these experiments to assess absolute-route clustering rather than endpoint-independent discrimination. Section 4.2.2 separately evaluates matched-endpoint settings in which labels are defined by interior route structure. Sample counts, class construction, landmarks, and museum ethics notes appear in Appendix A.6 (Table 12).

Representations and baselines. Unless otherwise stated, ARHR uses the first-order segment-exact angular coefficient in Eq. (3) with $w_i = |\delta_i|$, concatenated with the cumulative harmonic embedding h to form $z_1^{(w)}$. We compare against the flattened harmonic representation of Frantzen et al. (Frantzen et al., 2021), endpoint coordinates E_{xy}, the Yin-style endpoint-augmented descriptor $[h; E_{xy}]$ (Yin et al., 2017), Ghosh et al. (Ghosh et al., 2018), geometric baselines (ResampledXY40, Fréchet (Alt & Godau, 1995; Eiter et al., 1994), Hausdorff (Huttenlocher et al., 1993), and TRACLUS (Lee et al., 2007)), and learned representations (DTC (Madiraju, 2018), TimesURL (Liu & Chen, 2024), E^{2}DTC (Fang et al., 2021), t2vec (Li et al., 2018), and TrajCL (Chang et al., 2023)).

[1] Partner multi-hole junction cohort (N=643, 27 `label_v4` classes); released with the paper code (Appendix A.6).

Table 1: Absolute-route clustering ARI and NMI (mean over k-means seeds $\{0, 1, 2\}$). Bold marks the best entry in each metric column. openDD is the unweighted mean over rdb1–rdb7. Avg is the mean over openDD, Swiss, 3DJ, and DLR. Yin (s, t, h) is implemented as $[h; E_{xy}]$ with landmark-centered start/end coordinates.

		openDD		Swiss		3DJ		DLR		Avg	
Category	Method	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
Top.	ARHR (ours)	**0.907**	0.941	0.880	0.904	0.996	0.992	0.919	0.945	0.926	0.946
	Yin (s, t, h)	0.906	0.942	0.861	0.904	**1.000**	**1.000**	**0.924**	0.954	0.923	0.950
	Endpoint E_{xy}	0.871	0.934	0.782	0.888	0.991	0.992	0.855	0.926	0.875	0.935
	Frantzen-flat	0.678	0.766	0.457	0.564	**1.000**	**1.000**	0.872	0.887	0.752	0.804
	Ghosh	0.634	0.759	0.695	0.806	0.961	0.954	**0.924**	**0.955**	0.804	0.869
Trad.	ResampledXY40	0.772	0.878	0.875	0.921	**1.000**	**1.000**	0.844	0.880	0.873	0.920
	Fréchet	0.825	**0.945**	**0.992**	**0.992**	**1.000**	**1.000**	0.899	0.936	**0.929**	**0.968**
	Hausdorff	0.764	0.851	0.591	0.780	0.635	0.906	0.871	0.915	0.715	0.863
	TRACLUS	0.115	0.301	0.260	0.607	0.271	0.564	0.338	0.645	0.246	0.529
Deep	DTC	0.820	0.905	0.877	0.924	**1.000**	0.999	0.803	0.889	0.875	0.929
	TimesURL	0.793	0.891	0.751	0.850	0.960	0.987	0.812	0.889	0.829	0.904
	E2DTC	0.612	0.814	0.530	0.644	0.119	0.352	0.810	0.901	0.518	0.678
	t2vec	0.656	0.806	0.601	0.681	0.259	0.593	0.629	0.747	0.536	0.707
	TrajCL	0.724	0.840	0.765	0.862	0.746	0.893	0.655	0.799	0.723	0.849

Protocol. Fixed-dimensional representations are standardized and clustered with k-means using n_{init}=10; reported results are means over clustering seeds $\{0, 1, 2\}$. The number of clusters is set to the number of route classes. Pairwise-distance baselines use the corresponding agglomerative or medoid procedure. We report adjusted Rand index (ARI) (Hubert & Arabie, 1985) and normalized mutual information (NMI). For openDD, results are averaged uniformly over rdb1–rdb7.

4.2 Route Discrimination

4.2.1 Absolute-route benchmark

Table 1 evaluates absolute-route clustering across four traffic datasets. ARHR substantially improves over the flattened harmonic baseline on openDD ($0.678 \rightarrow 0.907$ ARI), Swiss ($0.457 \rightarrow 0.880$), and DLR ($0.872 \rightarrow 0.919$), while the harmonic baseline is already saturated on 3DJunction1. Across datasets, ARHR reaches mean ARI 0.926, close to the Yin-style endpoint–harmonic representation (0.923) and Fréchet distance (0.929).

Endpoint features are themselves strong in this benchmark because absolute-route identity is closely associated with route entry and exit. Table 1 should therefore not be interpreted as evidence that ARHR captures information unavailable from endpoints. Rather, it shows that angular statistics can recover route distinctions that are under-resolved by the fixed harmonic representation while remaining competitive with explicit endpoint augmentation. Section 4.2.2 isolates the complementary question of whether the angular statistics remain informative when endpoint variation is controlled.

4.2.2 Matched-endpoint discrimination

The absolute-route benchmark does not isolate information beyond route endpoints. We therefore evaluate endpoint-controlled settings whose labels depend on interior route structure. The controlled same-OD experiment uses three interior templates (n=24), while the Lyon synthetic benchmark averages twelve same-OD route-choice tasks. We additionally report a museum-visitor setting with manually annotated local circulation modes; because its endpoint context is defined at the tour level, we treat it as an auxiliary real-data result rather than a strict endpoint-controlled test.

Table 2 compares h, start/end azimuths E, the monotone closed-form control P_1 from Eq. (4), the first-order activity block a, and selected concatenations. This answers whether discrimination reduces to "endpoint + net span." We do *not* treat endpoint+turning-point features as a rival baseline: under default weights they are an equivalent structural description of C_k (Sec. 3.2).

On the controlled same-OD probe, h+P_1 equals h while a=1: class separation is not explained by endpoint azimuths or net span. On Lyon, E remains null, P_1 is only moderate, and a leads both P_1 and h+P_1. The museum-visitor setting is complementary rather than a pure matched-endpoint null:

Table 2: Matched-endpoint route discrimination (ARI / NMI; mean over k-means seeds $\{0, 1, 2\}$). E is the start/end azimuth baseline; P_1 is the monotone endpoint-span control (4); a is the first-order activity block C_1. Lyon is the mean over twelve OD buckets. Museum: E is tour-level; h, P_1, and a use the local near-exhibit segment.

		h		E		$h{+}E$		P_1		$h{+}P_1$		a		$h{+}a$	
Setting	n	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
Controlled same-OD	24	0.410	0.647	0.000	0.000	0.410	0.647	0.114	0.253	0.410	0.647	**1.000**	**1.000**	**1.000**	**1.000**
Lyon corridor	968	0.402	0.445	0.000	0.000	0.402	0.445	0.510	0.566	0.523	0.569	**0.734**	**0.753**	0.536	0.580
Museum approach	108	0.445	0.553	0.034	0.049	0.422	0.457	0.538	0.563	0.920	0.893	0.463	0.528	0.691	0.756

tour-level E is near chance, while local P_1 partially tracks approach direction, so $h{+}P_1$ is strong; h and a are each only partially informative and $h{+}a$ rises above either alone. These experiments support a narrower claim than superiority over every endpoint description: when endpoints and the monotone chord P_1 under-resolve interior structure, angular activity retains distinctions that require turning information. Concatenation with h remains task dependent.

4.3 Activity Measure and Azimuthal Correspondence

The first-order representation depends on the activity measure used to aggregate landmark-relative azimuth. We therefore evaluate $z_1^{(w)}$ under alternative nonnegative segment weights while keeping the remaining representation and clustering protocol unchanged. Table 3 compares angular activity $w_i = |\delta_i|$, equal weighting, arc length, and a permutation control that preserves the angular-activity weights within each trajectory but reassigns them to different occurrence azimuths.

Table 3: Activity-measure diagnostics for the first-order ARHR representation $z_1^{(w)}$ (ARI). Best per column in **bold**.

Activity measure	openDD	Swiss	3DJunction1	DLR
Angular activity	**0.907**	**0.880**	**0.996**	**0.919**
Equal weights	0.724	0.672	0.848	0.841
Arc-length weights	0.818	0.672	0.879	0.842
Permuted angular activity	0.716	0.491	0.558	0.526

Angular-activity weighting gives the strongest result on all four datasets. Equal and arc-length weighting retain the same first-order angular form but perform substantially worse, indicating that the choice of activity measure matters. Permuting the activity weights reduces ARI while preserving their marginal values, showing that the representation is sensitive to the pairing between angular increments and their occurrence azimuths. This diagnostic does not by itself establish information beyond endpoint- or turning-point-derived angular structure. Time weights and a mid-trajectory azimuth baseline appear in Appendix A.3.

4.4 Representation Capacity

We examine two representation choices: the angular truncation order K and the harmonic configuration summarized by the number of available harmonic channels d_H. The d_H settings in Table 5 arise from the tested map/landmark configurations; they should not be interpreted as a freely tunable width of a fixed one-hole harmonic kernel.

We first fix the harmonic representation and angular-activity measure and vary K. Table 4 compares the first-order representation z_1, truncations through $K{=}2$ and $K{=}4$, and a phase-only first-order variant that removes the modulus of C_1. Under the evaluated standardization and clustering protocol, z_1 matches or exceeds the higher-order truncations on openDD, 3DJunction1, and DLR. On Swiss-F1, $K{=}2$ is higher than $K{=}1$ under the segment-exact readout, while $K{=}4$ again falls behind; we therefore keep $K{=}1$ as the default low-order setting and treat higher K as a capacity check rather than a mandatory upgrade. The phase-only variant remains competitive on openDD but is weaker on Swiss and comparable or slightly stronger on DLR, so the modulus of the first-order statistic is not uniformly dispensable.

Table 4: Angular-order truncations of ARHR under a fixed harmonic configuration and activity measure (ARI). openDD is the mean over rdb1–rdb7.

Representation	dim	openDD	Swiss	3DJ	DLR
ARHR, K=1 (z_1)	m+2	0.907	0.880	0.996	0.919
ARHR, K=2 (z_2)	m+4	0.896	0.953	0.837	0.890
ARHR, K=4 (z_4)	m+8	0.880	0.753	0.949	0.894
ARHR, K=1, phase-only	m+2	0.908	0.847	0.983	0.923

These results should not be read as an information-theoretic claim that higher angular orders are redundant. Section 3 shows that higher orders can distinguish distributions that collide at lower orders. Rather, the experiment shows that the additional distinctions exposed by $K>1$ do not uniformly improve the evaluated route partitions under the present feature scaling and clustering protocol.

We next compare harmonic configurations summarized by d_H. Table 5 compares the harmonic coordinate h and its first-order ARHR extension at the same d_H. When fewer harmonic channels are available, the first nonzero-order statistic produces large gains—for example on openDD rdb5 and rdb7 and on Swiss-F1. As d_H grows and the harmonic representation itself becomes more discriminative, the gain shrinks toward zero and can become slightly negative in saturated settings such as 3DJunction1.

Table 5: Harmonic-configuration diagnostic (ARI). For each site, the same d_H-channel harmonic feature is reported alone (h) and with the first-order ARHR extension (z_1); Δ is the absolute lift.

	d_H=1			d_H=2			d_H=3			d_H=4		
Dataset	h	z_1	Δ	h	z_1	Δ	h	z_1	Δ	h	z_1	Δ
openDD rdb5	0.679	0.999	+0.320	0.919	0.999	+0.080	0.942	0.999	+0.057	0.999	0.999	+0.000
openDD rdb7	0.411	0.995	+0.584	0.708	0.998	+0.290	0.980	0.998	+0.018	0.998	0.998	+0.000
Swiss-F1	0.310	0.880	+0.570	0.505	0.883	+0.378	0.906	0.898	−0.008	0.922	0.917	−0.005
3DJunction1	0.882	0.981	+0.099	1.000	0.989	−0.011	1.000	0.996	−0.004	1.000	0.996	−0.004
DLR-UT	0.872	0.919	+0.047	0.914	0.922	+0.008	0.920	0.922	+0.002	0.921	0.924	+0.003

Together, the two diagnostics compare two ways of enriching the representation: retaining additional angular moments and changing the available harmonic configuration. Additional distinctions improve clustering only when they align with the target route structure.

5 Discussion

ARHR complements cumulative harmonic responses with landmark-relative angular statistics, and the experiments show that the usefulness of this augmentation is task dependent. Activity weighting affects discrimination, higher angular orders do not consistently improve clustering, and richer harmonic configurations can reduce the gain from angular information. Endpoint features are already strong for absolute-route labels, whereas the matched-endpoint experiments isolate settings in which angular statistics remain informative beyond endpoint variation. ARHR remains a compact, incrementally computable descriptor, but it is sensitive to observation quality and does not retain the complete trajectory sequence. Our current evidence is limited to route clustering with known class counts.

6 Conclusion

We introduced ARHR, a compact trajectory representation that complements cumulative harmonic responses with landmark-relative angular statistics. Controlled and real-data experiments show that this information can resolve route distinctions under-resolved by the harmonic baseline, while its benefit depends on the task and representation configuration. These results support angular statistics as a task-dependent complement to cumulative harmonic trajectory representations.

AI USE STATEMENT

In this work, we used generative AI tools for designing research methodology or experiments and implementing methods. We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, assisting with translation, cleaning and reformatting datasets, supporting qualitative and thematic data analysis, or interpreting results, and the remaining required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools for creating or editing software code and editing a research paper to improve readability. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Helmut Alt and Michael Godau. Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(01n02):75–91, 1995.

Sergio Barbarossa and Stefania Sardellitti. Topological signal processing over simplicial complexes. *IEEE Transactions on Signal Processing*, 68:2992–3007, 2020.

Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd international conference on knowledge discovery and data mining*, pp. 359–370, 1994.

Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Lio, Guido F Montufar, and Michael Bronstein. Weisfeiler and lehman go cellular: Cw networks. *Advances in neural information processing systems*, 34:2625–2640, 2021a.

Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montufar, Pietro Lio, and Michael Bronstein. Weisfeiler and lehman go topological: Message passing simplicial networks. In *International conference on machine learning*, pp. 1026–1037. PMLR, 2021b.

Antonia Breuer, Jan-Aike Termöhlen, Silviu Homoceanu, and Tim Fingscheidt. opendd: A large-scale roundabout drone dataset. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6. IEEE, 2020.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

Eric Bunch, Qian You, Glenn Fung, and Vikas Singh. Simplicial 2-complex convolutional neural nets. *arXiv preprint arXiv:2012.06010*, 2020.

Yanchuan Chang, Jianzhong Qi, Yuxuan Liang, and Egemen Tanin. Contrastive trajectory similarity learning with dual-feature attention. In *2023 IEEE 39th International conference on data engineering (ICDE)*, pp. 2933–2945. IEEE, 2023.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *International conference on machine learning*, pp. 1725–1735. PMLR, 2020.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Stefania Ebli, Michaël Defferrard, and Gard Spreemann. Simplicial neural networks. *arXiv preprint arXiv:2010.03633*, 2020.

Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

Thomas Eiter, Heikki Mannila, et al. Computing discrete fréchet distance. 1994.

Jasso Espadaler Clapés, Robert Fonod, Emmanouil Manos Barmpounakis, and Nikolaos Geroliminis. Vehicle trajectory dataset from drone-collected data at three swiss roundabouts. In *104th Annual Meeting of the Transportation Research Board*. Zenodo, 2025.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pp. 226–231, 1996.

Ziquan Fang, Yuntao Du, Lu Chen, Yujia Hu, Yunjun Gao, and Gang Chen. E 2 dtc: An end to end deep trajectory clustering framework via self-training. In *2021 IEEE 37th international conference on data engineering (ICDE)*, pp. 696–707. IEEE, 2021.

Florian Frantzen, Jean-Baptiste Seby, and Michael T Schaub. Outlier detection for trajectories via flow-embeddings. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pp. 1568–1572. IEEE, 2021.

Abhirup Ghosh, Benedek Rozemberczki, Subramanian Ramamoorthy, and Rik Sarkar. Topological signatures for fast mobility analysis. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 159–168, 2018.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. Pmlr, 2017.

Vincent P Grande, Josef Hoppe, Florian Frantzen, and Michael T Schaub. Topological trajectory classification and landmark inference on simplicial complexes. In *2024 58th Asilomar Conference on Signals, Systems, and Computers*, pp. 44–48. IEEE, 2024.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

Daniel P Huttenlocher, Gregory A Klanderman, William Rucklidge, et al. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9): 850–863, 1993.

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Jae-Gil Lee, Jiawei Han, and Kyu-Young Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pp. 593–604, 2007.

Xiucheng Li, Kaiqi Zhao, Gao Cong, Christian S Jensen, and Wei Wei. Deep representation learning for trajectory similarity computation. In *2018 IEEE 34th international conference on data engineering (ICDE)*, pp. 617–628. IEEE, 2018.

Lek-Heng Lim. Hodge laplacians on graphs. *Siam Review*, 62(3):685–715, 2020.

Jiexi Liu and Songcan Chen. Timesurl: Self-supervised contrastive learning for universal time series representation learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 13918–13926, 2024.

Qianli Ma, Jiawei Zheng, Sen Li, and Gary W Cottrell. Learning representations for time series clustering. *Advances in neural information processing systems*, 32, 2019.

Naveen Sai Madiraju. Deep temporal clustering: Fully unsupervised learning of time-domain features. Master's thesis, Arizona State University, 2018.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 4602–4609, 2019.

Michael T Schaub, Austin R Benson, Paul Horn, Gabor Lippner, and Ali Jadbabaie. Random walks on simplicial complexes and the normalized hodge 1-laplacian. *SIAM Review*, 62(2):353–391, 2020.

Clemens Schicktanz, Lars Klitzke, Kay Gimm, Giancarlo Rizzo, Karsten Liesner, Henning Hajo Mosebach, and Sascha Knake-Langhorst. The dlr urban traffic dataset (dlr-ut): A comprehensive traffic dataset from an urban research intersection. In *2025 IEEE Intelligent Vehicles Symposium (IV)*, pp. 398–405. IEEE, 2025.

Davide Vecchia, Fatima Hachem, Davide Molteni, Maria Luisa Damiani, and Gian Pietro Picco. Unleashing the power of UWB for indoor mobility analytics: A museum case study. *Data Science and Engineering*, 2026. doi: 10.1007/s41019-026-00358-6. Dataset: Zenodo 10.5281/zenodo.19681005, CC BY 4.0.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Michail Vlachos, George Kollios, and Dimitrios Gunopulos. Discovering similar multidimensional trajectories. In *Proceedings 18th international conference on data engineering*, pp. 673–684. IEEE, 2002.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

Xiaotian Yin, Yu-Yao Lin, Chien-Chun Ni, Jiaxin Ding, Wei Han, Dengpan Zhou, Jie Gao, and Xianfeng Gu. Decentralized trajectory tracking using homology and hodge decomposition in sensor networks. *arXiv preprint arXiv:1708.09129*, 2017.

Mingxuan Yue, Yaguang Li, Haoze Yang, Ritesh Ahuja, Yao-Yi Chiang, and Cyrus Shahabi. Detect: Deep trajectory clustering for mobility-behavior analysis. In *2019 IEEE International conference on big data (Big Data)*, pp. 988–997. IEEE, 2019.

A Supplementary Analysis

This appendix provides derivations and extended diagnostics that complement the main analysis. Section A.1 expands properties of the staged angular coefficients used in Section 3. Section A.2 reports sampling density, collinear densification, and interior-vertex deletion for the angular readout. Section A.3 reports activity-measure variants omitted from the main comparison. Section A.4 provides additional harmonic-configuration diagnostics beyond the controlled comparisons in Section 4.4. Section A.5 examines degradation under normalized coordinate perturbations in synthetic map units, reported as σ/R. Section A.7 records the host and timing conventions for Table 13 and the controlled map behind Figure 2. Section A.6 lists evaluation cohorts, label construction, and museum ethics notes. Section A.1 also records a measure-level example in which higher-order coefficients separate distributions that collide at first order.

A.1 Properties of ARHR Coefficients

For completeness, we expand properties used in Section 3.3. For azimuthal activity the default $k \geq 1$ implementation is the segment-exact integral in Eq. (3): with unwrapped increments $\delta_i = \phi_i - \phi_{i-1}$, mid-azimuths $m_i = (\phi_i + \phi_{i-1})/2$, and weights $w_i = |\delta_i|$,

$$S_k^{(w)}(\gamma) = \sum_i w_i \, e^{ikm_i} \, \mathrm{sinc}\big(k\delta_i/2\big), \qquad C_k^{(w)} = S_k^{(w)}/(W_\gamma + \varepsilon), \tag{7}$$

where $W_\gamma = \sum_i w_i$ and $\mathrm{sinc}(x) = \sin(x)/x$, $\mathrm{sinc}(0) = 1$. A discrete midpoint sum $\sum_i w_i e^{ik\phi_i^{\mathrm{mid}}}$ is used only as an ablation of the same weights, not as the default coefficient. The cumulative harmonic embedding is $h(\gamma) = \sum_t y_t$, kept separate from the angular coefficients.

Segment-wise telescoping. Under the default weight $w_i = |\delta_i|$ and $\delta_i \neq 0$,

$$|\delta|\, e^{ikm}\, \mathrm{sinc}(k\delta/2) = \frac{\mathrm{sgn}(\delta)}{ik}\big(e^{ik\phi_i} - e^{ik\phi_{i-1}}\big).$$

Interior terms therefore cancel inside each monotone-azimuth run. If the unwrapped azimuth is monotone on the whole trajectory, write $D = \phi_N - \phi_0$. Then $W_\gamma = |D|$ and

$$C_k^{(w)} = \frac{\mathrm{sgn}(D)}{ik(|D| + \varepsilon)}\big(e^{ik\phi_N} - e^{ik\phi_0}\big) = \frac{|D|}{|D| + \varepsilon}\, e^{ik(\phi_0 + D/2)}\, \mathrm{sinc}\left(\frac{kD}{2}\right), \tag{8}$$

which is Eq. (4). The second equality uses $(e^{ik\phi_N} - e^{ik\phi_0})/(ik) = D\, e^{ik(\phi_0 + D/2)}\, \mathrm{sinc}(kD/2)$ together with $\mathrm{sgn}(D)\, D = |D|$. The prefactor equals one only in the limit $\varepsilon \to 0$. When the azimuth reverses, the same cancellation retains the endpoint and turning-point azimuths together with the total variation $W_\gamma > |D|$.

Higher-order measure distinction. As a measure-level example, place the same total mass equally at 0 and π, or equally at $\pi/2$ and $3\pi/2$. Both distributions have vanishing first-order coefficients, while their second-order coefficients have opposite signs. First-order cancellation can therefore leave azimuthal differences that higher orders still make visible. The example describes a distinction at the measure level and does not assume that every such distribution is realized by a feasible trajectory on every map.

Azimuth-shift equivariance. Suppose the occurrence azimuths of corresponding local activities are shifted by a common angle α, while the associated weights remain unchanged. Then δ_i and w_i are unchanged, m_i shifts by α, and

$$S_k^{(w)}(\gamma_\alpha) = \sum_i w_i\, e^{ik(m_i + \alpha)}\ \mathrm{sinc}\big(k\delta_i/2\big) = e^{ik\alpha} S_k^{(w)}(\gamma), \tag{9}$$

and therefore

$$C_k^{(w)}(\gamma_\alpha) = e^{ik\alpha} C_k^{(w)}(\gamma), \qquad k \geq 1. \tag{10}$$

Under the additional assumption that the cumulative harmonic response is unchanged, $h(\gamma_\alpha) = h(\gamma)$. Whenever a retained order satisfies $C_k^{(w)}(\gamma) \neq 0$ and $e^{ik\alpha} \neq 1$, the shifted trajectory is separated in $z_K^{(w)}$. This is the resolvability relation summarized by Eq. (6). Angular-activity weights $w_i = |\delta_i|$ are invariant to a common azimuth translation, so the weight hypothesis holds for that default measure.

Additivity before normalization. For two already segmented trajectory pieces γ_1 and γ_2 that share the same map, landmark, and local recording convention,

$$S_k^{(w)}(\gamma_1 * \gamma_2) = S_k^{(w)}(\gamma_1) + S_k^{(w)}(\gamma_2), \tag{11}$$

and the same additivity holds for the total activity W and for the cumulative harmonic embedding $h = \sum_t y_t$. Normalized nonzero-order coefficients are obtained only after these sufficient statistics are merged; the normalized values themselves do not add directly. This justifies the incremental updates and segment composition stated in Section 3.3.

A.2 Sampling and collinear subdivision

We record the sampling density of openDD rdb5 (Table 6), then change the vertex set in two ways while keeping the underlying geometry or the matched harmonic coordinate h under control. Table 7 inserts collinear points on a fixed polyline and compares discrete versus segment-exact $K{=}1$ coefficients. Table 8 deletes interior vertices with h held fixed and reports absolute-route ARI.

A.2.1 SAMPLING DENSITY

Table 6 summarizes vertex counts and per-segment unwrapped azimuth increments on openDD rdb5 under the absolute-route benchmark. Most segments change azimuth by well under one degree, so the discrete midpoint sum is already a fine angular partition; the segment-exact default matches that limit analytically.

Table 6: Vertex counts and per-segment unwrapped azimuth increments $|\Delta\phi|$ on openDD rdb5 (absolute-route benchmark, n=3814).

Quantity	min	q25	q50	q75	q95	max	mean
#vertices / traj.	86	331	381	431	546	1051	377.4
$\|\Delta\phi\|$ (deg), all segs.	0	0.073	0.384	0.837	1.058	1.681	0.462
$\max\|\Delta\phi\|$ (deg) / traj.	0.527	0.960	1.049	1.132	1.260	1.681	1.043

A.2.2 COLLINEAR DENSIFICATION

Table 7 inserts collinear vertices on a fixed polyline ($8\times$ densification) and measures Euclidean drift of the K=1 angular block relative to the undensified polyline. The discrete midpoint readout drifts under densification; the segment-exact integral stays at machine precision on both the synthetic coarse paths and rdb5, matching the definitional invariance claimed in the main text.

Table 7: Euclidean change of the K=1 angular block under collinear $8\times$ densification of a fixed polyline, relative to $1\times$.

Cohort	Readout (K=1)	median L_2	q95 L_2
Synthetic coarse	Discrete	5.79×10^{-2}	1.30×10^{-1}
Synthetic coarse	Exact	6.94×10^{-17}	1.11×10^{-16}
rdb5 (n=3814)	Discrete	2.27×10^{-6}	4.73×10^{-6}
rdb5 (n=3814)	Exact	6.94×10^{-17}	2.22×10^{-16}

A.2.3 INTERIOR-VERTEX DELETION

Table 8 deletes interior vertices while holding the matched harmonic coordinate h fixed, then re-clusters under the absolute-route protocol. ARI remains saturated for both exact and discrete K=1 blocks; higher K shows only small RMS drift of the discrete block under nonuniform deletion. This is a frozen-h diagnostic of angular readout stability, not an end-to-end robustness claim.

Table 8: Absolute-route ARI on rdb5 (n=3814, 13 classes) after deleting interior vertices with h held fixed. Median vertex counts: 381 (original), 192 (half), 97 (quarter), 194 (nonuniform).

K	Method	original	half	quarter	nonunif.
1	$[h;\, a_{\mathrm{exact}}]$	0.999	0.999	0.999	0.999
1	$[h;\, E_{\mathrm{az}}]$	0.999	0.999	0.999	0.999
4	$[h;\, a_{\mathrm{exact}}]$	0.981	0.981	0.981	0.981
4	$[h;\, E_{\mathrm{az}}]$	0.999	0.999	0.999	0.999

Under nonuniform deletion, the median RMS drift of the discrete angular block is 1.04×10^{-4} for K=1 and 7.79×10^{-3} for K=4. The median drift of the exact block remains at machine precision.

A.3 EXTENDED ACTIVITY-MEASURE DIAGNOSTICS

Table 3 in the main text compares angular activity, equal weights, arc length, and the permutation control for $z_1^{(w)}$. Table 9 extends that comparison with two additional controls under the same absolute-route protocol. Time weighting uses the inter-sample interval Δt; under the pseudo-time

sampling adopted here, it coincides with equal weighting, so the two rows match. Mid-trajectory azimuth replaces the integrated statistic by a single landmark-relative bearing at the path midpoint. Neither variant improves over the angular-activity measure used by the default first-order representation. The permutation control remains weak, indicating that alignment between activity weights and occurrence azimuth—rather than the marginal weight values alone—is required.

Table 9: Additional activity-measure variants for $z_1^{(w)}$ (ARI). openDD is the mean over rdb1–rdb7. The main-text Table 3 already reports angular / equal / arc-length / permuted.

Weighting	openDD	Swiss	3DJ	DLR
Angular activity $\lvert d\phi \rvert$	0.907	0.880	0.996	0.919
Equal	0.724	0.672	0.848	0.841
Time Δt	0.724	0.672	0.848	0.841
Arc length Δs	0.818	0.672	0.879	0.842
Permuted $\lvert d\phi \rvert$	0.716	0.491	0.558	0.526
Mid-trajectory azimuth	0.728	0.688	0.840	0.678

A.4 Additional Harmonic-Configuration Diagnostics

Section 4.4 compares angular order and harmonic configurations under matched feature settings. Here we report an additional exploratory punch search on two openDD sites that remain unsaturated under the thin harmonic representation. The search modifies peripheral holes and records the strongest pure-harmonic ARI observed among the tested map configurations. These values are empirical search outcomes rather than optima over all possible complexes.

Table 10: Punch hunt on unsaturated openDD sites (ARI). "best tested harm." is the strongest pure-harmonic ARI observed under the listed punch protocols, not a mathematical maximum over all complexes.

Site	thin harm.	ARHR	best tested harm.	gap	Punch configuration
rdb2	0.482	0.715	0.610	+0.105	wedge+`never+wedge`; β_1 stuck at 1
rdb6	0.344	0.648	0.655	−0.007	`never+wedge`, β_1=8 (catches)

On rdb2, the tested punch protocols leave a gap relative to ARHR. On rdb6, a configuration with β_1=8 essentially matches ARHR under the reported search. The contrast illustrates that enlarging topological capacity can close some, but not necessarily all, orientation-related gaps without angular readouts.

A.5 Sensitivity to Position Noise

The main experiments evaluate clean route representations. We additionally test how landmark-relative angular statistics degrade under normalized coordinate perturbations on a synthetic map. σ is expressed in the coordinate units of the controlled site; these levels are not calibrated GPS meters. Table 11 reports ARI on an orientation-collision site for the harmonic baseline and first-order ARHR with activity weights $\lvert d\phi \rvert$. Labels are absolute-route classes; clustering uses Scaler + k-means with N=288 trajectories and 24 classes. The site places the landmark at the origin with landmark radius R=4 in the same synthetic map units used for the injected noise. Bypass arms extend to radius 12.

Table 11: Coordinate-perturbation stress test (ARI). Rows are isotropic noise levels relative to landmark radius R; columns are pure harmonics and first-order ARHR.

σ	σ/R	harm.	ARHR
0	0	0.089	1.000
0.25	0.0625	0.100	0.784
0.5	0.125	0.100	0.597
0.75	0.1875	0.096	0.569
1.0	0.25	0.094	0.548
1.5	0.375	0.094	0.481
2.0	0.5	0.032	0.393

Isotropic Gaussian noise with $\sigma \in \{0, 0.25, 0.5, 0.75, 1.0, 1.5, 2.0\}$ in synthetic map units is added to interior coordinates. Endpoints remain fixed. Relative to the landmark radius these levels are $\sigma/R \in \{0, 0.0625, 0.125, 0.1875, 0.25, 0.375, 0.5\}$. Thus σ=0.25 is 6.25% of R. On this site the median vertex–landmark distance coincides with R, so the same percentage applies to that median as well. At σ=0, ARHR recovers perfect absolute-route ARI of 1.000, while pure harmonics stay near chance at 0.089. As σ increases, the first-order representation degrades gradually from 1.000 to 0.784, 0.597, 0.569, 0.548, 0.481, and 0.393, revealing an explicit sensitivity of the azimuthal statistics to coordinate perturbations. Pure harmonics never recover the orientation cue and drop further to 0.032 at σ/R=0.5.

A.6 Datasets, Labels, and Ethics

Table 12 records the evaluation cohorts used in the absolute-route and matched-endpoint experiments: sample counts, class counts, sensing, label construction, and the landmark / map convention used for harmonic and angular features. openDD results in the main tables are unweighted means over rdb1–rdb7; the per-site rows below make that average inspectable. Rare absolute-route classes are removed according to the site-specific preprocessing rules used to construct each cohort.

Table 12: Dataset statistics for the absolute-route benchmark and the museum matched-endpoint cohort. N and #classes are after the listed filters. Landmark / map is the reference used for h and angular coefficients.

Dataset	Site	N	#cl.	Sampling	Label construction	Landmark / map
openDD	rdb1	27078	9	Drone / HD map	Official entry→exit; site-specific rare classes removed	NEVER-island centroid
openDD	rdb2	4337	8	Drone / HD map	Official entry→exit; site-specific rare classes removed	NEVER-island centroid
openDD	rdb3	7727	13	Drone / HD map	Official entry→exit; site-specific rare classes removed	NEVER-island centroid
openDD	rdb4	1034	3	Drone / HD map	Official entry→exit; site-specific rare classes removed	NEVER-island centroid
openDD	rdb5	3814	13	Drone / HD map	Official entry→exit; site-specific rare classes removed	NEVER-island centroid
openDD	rdb6	10577	9	Drone / HD map	Official entry→exit; site-specific rare classes removed	NEVER-island centroid
openDD	rdb7	7691	13	Drone / HD map	Official entry→exit; site-specific rare classes removed	NEVER-island centroid
Swiss	F1	3310	8	Drone tracks	Heuristic arm OD sectors (not official)	Fitted island center
3DJunction1	—	643	27	Junction polylines	`label_v4`; keep count≥3	Vertex-cloud median
DLR-UT	AIM	7626	9	Roadside / daytime	OpenDRIVE arms→entry/exit; min≥5	Junction dense-core median
Museum	approach	108	4	Indoor UWB	Manually annotated local circulation modes	Exhibit furniture centroid

3DJunction1 provenance. 3DJunction1 is a partner multi-hole junction shapefile used in prior topological trajectory benchmarks at this site; cleaned routes with `label_v4` route identities (N=643, 27 classes with count ≥ 3). It is not a public DOI release at submission time; the cleaned cohort and preprocessing scripts ship with the paper code.

Museum data source and labels. The museum cohort uses indoor visitor polylines from the public MUSE UWB museum release (Vecchia et al., 2026) (Zenodo record `10.5281/zenodo.19681005`, CC BY 4.0). Table 2 reports this museum-visitor setting with manually annotated local circulation modes (4 classes, N=108). Its endpoint context is defined at the tour level, whereas h and the angular statistics are computed on the local near-exhibit segment, so the row is an auxiliary real-data result rather than a strict endpoint-controlled test.

Ethics and privacy (museum). MUSE trajectories are released as UWB tag coordinates without names or other direct personal identifiers; tag IDs may be reused across days and are not stable person IDs. We use only the publicly released anonymized spatial trajectories, do not attempt re-identification, and follow the original dataset license and site policies.

Reproducibility. Fixed-dimensional results average k-means seeds $\{0, 1, 2\}$ with n_{init}=10 and k equal to the number of route classes. openDD averages are unweighted over rdb1–rdb7. Swiss-F1 uses parse seed 20260918 (main-table / DTC cohort). Code, feature caches, and table-building scripts will be released with the camera-ready version; public datasets are cited above, and 3DJunction1 ships with the release pack.

A.7 Implementation and Timing Details

Wall-clock cost on openDD rdb5. Table 13 reports same-host timings for ARHR and baselines on openDD rdb5 (N=3814). ARHR, Endpoint E_{xy}, and Yin $[h; E_{xy}]$ are analytic, linear per-trajectory extractors with no representation-learning stage. On this absolute-route site, Yin and ARHR are near-saturated; Endpoint alone is already a strong simple baseline. Pairwise geometric methods can match or exceed that ARI at quadratic matrix cost; learned methods incur multi-epoch training before clustering.

Table 13: Cost and ARI on openDD rdb5 (N=3814). Train is representation-learning time. Feature is analytic extraction, pairwise-matrix construction, or post-training encoding. Cluster is Scaler plus k-means or agglomerative clustering.

Category	Method	Scaling	Train (s)	Feature (s)	Clust. (s)	ARI
Topological	ARHR (ours)	$O(NL)$	—	2.3	5.8	**0.999**
	Yin $[h; E_{xy}]$	$O(NL)$	—	0.03	2.4	**0.999**
	Endpoint E_{xy}	$O(N)$	—	0.02	2.6	0.827
	Frantzen-flat	$O(NL)$	—	0.11	8.8	0.679
	Ghosh	$O(NL)$	—	41.7	—	0.796
Traditional	Fréchet	$O(N^2L^2)$	—	12.3	0.23	1.000
	Hausdorff	$O(N^2L^2)$	—	4.57	0.21	0.962
	ResampledXY40	$O(NL)$	—	2.1	15.5	0.696
Deep	TrajCL (e=20)	$O(ENF)$	2724	62.4	42.2	0.857
	TimesURL (e=20)	$O(ENF)$	1050	—	10	0.827
	t2vec warm-start	$O(ENF)$	277	—	5	0.554
	DTC formal	$O(ENF)$	37.5	—	36	0.813

Hardware and software for Table 13. Table 13 was timed on one evaluation host. Analytic ARHR, Yin, Endpoint, and flattened-harmonic features and k-means ran on CPU. Deep methods used one GPU on the same host. Table 14 lists the host and the software stack used for those measurements.

Timing conventions. Pairwise Fréchet and Hausdorff Feature times are the first construction of the full distance matrix. For those pairwise baselines, trajectories are arc-length resampled to L=32 points before the matrix is built. ARHR and flattened harmonics use map-matched edge traversals and the input polyline azimuths, without that resampling. Endpoint E_{xy} uses landmark-centered

Table 14: Host and software used for Table 13.

Item	Specification
CPU	Intel Xeon Gold 6338 @ 2.00 GHz, 64 cores
Memory	1007 GB
GPU	NVIDIA A800-SXM4 80 GB
Python	3.11
NumPy	2.4
SciPy	1.17
scikit-learn	1.8
PyTorch	2.3 with CUDA 12.1

start and end coordinates only; Yin $[h; E_{xy}]$ stacks the already-available harmonic coordinate with E_{xy}. Feature time for ARHR, Yin, and flattened harmonics counts extraction after the harmonic basis is available. Map construction and harmonic-basis computation are preprocessing costs. Per-trajectory map matching is a separate stage and is excluded from the feature-extraction complexity reported here. Train time for deep methods is representation-learning wall-clock before encoding. Cluster time covers StandardScaler and k-means with $n_{\text{init}}=10$, or the agglomerative or medoid step used for distance baselines. Reported ARI and NMI for fixed-dimensional methods are means over clustering seeds $\{0, 1, 2\}$, matching Table 1.

Controlled map for Figure 2. Figure 2 uses a Frantzen two-hole complex (`random_graph`, 1000 points, two holes, seed 512; $\beta_1=2$). Same-OD corridors are sampled as weight-perturbed shortest paths on that mesh. Four absolute-route classes (40 paths each) are homologous about the landmark and share one origin–destination pair, so the full harmonic coordinate $h = Q_H^\top f \in \mathbb{R}^2$ strongly overlaps across classes (panel (b); ARI≈ 0.11), while the first-order coefficient $C_1^{(w)}$ separates the same trajectories (panel (c); ARI$= 1.00$). Panel (a) plots the discrete mesh corridors used for the features.