# ACPL: A Coordinate-System Prior for Topology-Aware Directed Trajectory Inference on Cyclic Biological Manifolds

Shreyo Ghosh

Seven Lakes High School, Katy, Texas 77494, USA. ghoshshreyo2007@gmail.com

Incoming Class of 2030, Department of Biomedical Engineering, Rutgers University

Directed trajectory inference from single-cell RNA sequencing data requires graph construction algorithms that respect the geometric structure of cyclical biological processes. Minimum spanning tree approaches systematically produce backward edges across cyclical manifolds through Euclidean shortcuts at the loop closure, inverting the inferred direction of division at its most biologically consequential transition. We introduce ACPL (Arc+LOESS Polar-Linearization), a coordinate-system prior that transforms two-dimensional UMAP embeddings into polar coordinates aligned with the manifold's angular structure and admits only forward edges along LOESS-smoothed cumulative arc length. We prove that the resulting directed graph is acyclic by construction at $O(N \log N)$ cost and that LOESS smoothing is mathematically necessary to avoid a trivial accuracy artefact. Empirical evaluation across five biological datasets is reported in full: ACPL achieves 97.0% structural accuracy on Spellman yeast and 75.9% on Nestorowa haematopoietic stem cells, with all pairwise bootstrap confidence intervals crossing zero. The method is robust to UMAP hyperparameter choice (SD = 0.77 pp across 25 configurations) and survives spatial autocorrelation correction by hundreds of orders of magnitude. An R implementation is available at https://github.com/shreyosecret/ACPL under the MIT License. Supplementary data are available at *Bioinformatics Advances* online. Contact: ghoshshreyo2007@gmail.com.

**Keywords:** trajectory inference; polar coordinates; directed acyclic graph; UMAP; single-cell RNA sequencing.

## 1. Introduction

### *1.1. The Euclidean shortcut problem*

Single-cell RNA sequencing captures a static snapshot of cellular states from which trajectory inference seeks to reconstruct a directed biological process. For cyclical processes, this reconstruction has a specific geometric character that has not been adequately addressed in the literature. Projection of cell cycle expression data into two dimensions using topology-preserving algorithms such as UMAP produces a closed loop that faithfully reflects the cyclic topology of the underlying biological system. This faithful representation, paradoxically, gives rise to a structural failure mode in subsequent graph construction.

Graph construction algorithms based on Euclidean proximity, including the minimum spanning tree (MST) approaches underlying Monocle and related tools [Trapnell et al., 2014, Qiu et al., 2017], are designed to minimise total spanning cost. On a circular manifold, the minimum-cost connection is not the arc along the trajectory but rather the chord traversing the loop interior. Cells terminating mitosis (G2M phase) reside in close spatial proximity to cells re-entering G1 in the two-dimensional projection because the cycle closes geometrically, despite being biologically distant. The greedy spanning tree algorithm preferentially connects these spatially adjacent yet temporally distant cells, producing a backward edge at the most biologically consequential transition in the cell cycle. We designate this phenomenon the *Euclidean shortcut problem*, a structural consequence of applying a Euclidean graph algorithm to a non-Euclidean manifold rather than a deficiency of any specific implementation. Crucially, this defect cannot be corrected through post-hoc direction assignment; it must be prevented at the stage of graph construction.

The trajectory inference literature has acknowledged this difficulty. Haghverdi et al. [2016] noted it in the context of diffusion pseudotime, and Wolf et al.

[2019] motivated PAGA partly by the same concern. Existing approaches, however, address the problem indirectly: PAGA operates at cluster resolution and consequently relinquishes single-cell ordering; RNA velocity [La Manno et al., 2018, Bergen et al., 2020] provides directional information but requires spliced and unspliced transcript counts unavailable in legacy SMART-seq2 datasets; and Slingshot [Street et al., 2018] fits principal curves through clusters but provides no formal acyclicity guarantee. The question addressed in the present work is whether the Euclidean shortcut problem, given its geometric cause, admits a geometric solution.

### *1.2. A coordinate-system prior*

The solution proposed here is a coordinate system. Most trajectory inference methods construct a graph from cell-to-cell similarity, then impose direction through a secondary procedure such as root cell selection. Direction is therefore inferred from a graph whose topology has already been determined by a direction-agnostic proximity measure, and once the MST is constructed using Euclidean distances, the backward edges it contains cannot be removed without rebuilding the graph.

ACPL departs from this architecture. The method selects a coordinate system first, one in which the cyclic process's angular structure constitutes the primary organising principle, and uses this coordinate system to determine which edges may exist. Polar coordinates centred at a hub chosen to minimise radial variance render the angular structure of the cycle explicit. Cells are ordered by cumulative arc length along the angular trajectory, and edges are admitted exclusively in the direction of increasing arc length. Euclidean shortcuts become geometrically impossible. The formulation constitutes a coordinate-system prior in the sense familiar from physics: the choice of coordinates reflects the symmetries of the underlying process. ACPL replaces only the graph construction step; it requires no user-specified root cell and incurs $O(N \log N)$ cost, in contrast to $O(N^2 \log N)$ for MST.

### *1.3. Contributions and findings*

The empirical evaluation reported in this work does not demonstrate universal superiority of ACPL. The method achieves 97.0% structural accuracy on the Spellman yeast cell cycle ($N$=18, synchronised microarray data) compared with 52.8% for MST, exhibits a directional but statistically non-significant advantage on Nestorowa haematopoietic stem cells ($N$=1,920; +1.6 pp over MST), and underperforms MST on three additional datasets. A quantitative curvature analysis intended to explain the Nestorowa advantage returned a null result, prompting withdrawal of the spatial-localisation claim from an earlier version. A loop circularity diagnostic revealed a counterintuitive pattern that motivated reframing the applicability condition in terms of angular monotonicity rather than geometric circularity. What the present work establishes are formal results that hold independently of empirical performance: Theorems 1 and 2 concerning acyclicity by construction and the necessity of LOESS smoothing; complete parameter invariance of the global ordering with respect to LOESS span; biological anchor validation surviving spatial autocorrelation correction by hundreds of orders of magnitude.

## 2. Related work

Several systematic benchmarks of trajectory inference methods have been published [Saelens et al., 2019]. Monocle 1 [Trapnell et al., 2014] orders cells via ICA-extracted principal curves, an approach suited to linear trajectories. Monocle 2 [Qiu et al., 2017] extends this framework to branching topologies using DDRTree, though its graph construction step remains MST-based and therefore subject to the Euclidean shortcut problem. PAGA [Wolf et al., 2019] operates at cluster resolution and consequently sacrifices single-cell ordering. Diffusion pseudotime [Haghverdi et al., 2016] requires $O(N^2)$ memory and performs poorly at small sample sizes. Slingshot [Street et al., 2018] fits smooth principal curves and achieves the highest overall accuracy in our benchmarks, although it requires user specification of a starting cluster.

UMAP [McInnes et al., 2018] occupies a privileged position in the present analysis. The algorithm explicitly preserves local topological structure that PCA discards, and the quality of this preservation determines whether a coordinate-system prior has signal to exploit. On the Spellman dataset, ACPL applied to UMAP coordinates achieves 97.0% structural accuracy compared with 66.7% on PCA coordinates (Supplementary Figure S1), substantiating the role of the embedding.

ACPL is conceptually related to diffusion maps, which employ Laplacian eigenvectors of the data graph as in-

trinsic coordinates; ACPL employs polar coordinates of the two-dimensional embedding. Both exploit intrinsic geometry rather than ambient Euclidean geometry. The trade-off between generality and formal tractability is a recurring theme in this class of methods, and ACPL is positioned deliberately at the formally tractable end.

## 3. Methods

### *3.1. The polar coordinate engine*

The ACPL pipeline transforms a two-dimensional UMAP embedding into polar coordinates relative to an adaptive hub. The only assumption invoked by the transformation is that cells appearing later in the biological process tend to occupy larger angular coordinates $\theta$ around the hub than earlier cells. Circularity, uniform radial distribution, and isometry between arc length and biological time are not assumed.

Given a two-dimensional embedding $\mathbf{X}=\{x_1,\ldots,x_N\}$, the hub $c^*$ is chosen to minimise radial variance,

$$c^* = \arg\min_c \mathrm{Var}(\{\|x_i - c\|_2\}),$$

optimised by Nelder-Mead initialised at the coordinate median. For simple cyclic manifolds this procedure recovers the geometric centroid; for complex branching manifolds such as Nestorowa HSC, it identifies a hub that minimises the spread of radial distances across the multi-lineage structure. The polar coordinates are then

$$\theta_i = \mathrm{atan2}(y_i - c_y^*, x_i - c_x^*), \quad r_i = \|(x_i - c_x^*, y_i - c_y^*)\|_2, \tag{1}$$

where $\theta$ encodes biological phase and $r$ encodes radial distance from the hub. Because $r$ carries no temporal signal on a closed-loop manifold (cells across all phases of the cycle remain approximately equidistant from the hub), the temporal proxy is constructed entirely from $\theta$. The angle is unwrapped to a continuous monotonically increasing variable to accommodate multi-cycle datasets.

### *3.2. Arc length, LOESS, and the necessity of smoothing*

Cumulative arc length $s$ is computed along the $\theta$-ordered trajectory and subjected to LOESS smoothing to yield the time proxy $s^*$:

$$\begin{aligned} s_i &= \sum_{k=1}^{i-1} \|x_{k+1} - x_k\|_2, \\ s_i^* &= \mathrm{LOESS}(s_i \mid \theta_i, \mathrm{span}). \end{aligned} \tag{2}$$

The necessity of LOESS smoothing is established by the following result.

**Theorem 1** (Raw arc length degeneracy)**.** *For any non-degenerate embedding in which no two consecutive nodes share coordinates, the raw cumulative arc length $s_i$ is strictly monotonically increasing with node index under the traversal ordering.*

*Proof.* The increment $s_{i+1} - s_i = \|x_{i+1} - x_i\|_2 > 0$ holds for all $i$ by positive definiteness of the Euclidean norm. □ □

**Corollary 2.** *A monotonicity filter applied to raw arc length admits every edge in traversal order and consequently yields* 100% *structural accuracy as a trivial arithmetic identity rather than as a result of geometric inference. LOESS smoothing disrupts strict monotonicity at curvature reversals, rendering the filter genuinely selective.*

This corollary has an important practical consequence. The 100% structural accuracy that a naive implementation would report does not constitute evidence of strong performance; rather, it indicates that the filter is performing no useful function. The ablation study in Section 4.2 substantiates this point empirically: raw arc length achieves 100%† on both Spellman and Nestorowa, whereas ACPL with LOESS achieves 66.7% and 75.9% respectively, values that reflect genuine geometric inference. Throughout this work, raw arc length results are marked with † to indicate their degenerate status.

### *3.3. Theorem 1: Acyclicity by construction*

**Theorem 3** (Acyclicity by construction)**.** *The directed graph $G$ produced by the $\tau$-monotonicity filter contains no directed cycles.*

*Proof.* The filter admits the directed edge $(i \to j)$ if and only if $s_j^* > s_i^*$. Suppose, for contradiction, that a directed cycle $i_1 \to i_2 \to \cdots \to i_k \to i_1$ exists. The admission rule implies $s_{i_2}^* > s_{i_1}^*$, $s_{i_3}^* > s_{i_2}^*$, $\ldots$, $s_{i_1}^* > s_{i_k}^*$. Transitivity then yields $s_{i_1}^* > s_{i_1}^*$, which is a contradiction. □ □

The total computational cost of the pipeline is $O(N \log N)$: LOESS smoothing requires $O(N \log N)$, sorting requires $O(N \log N)$, and identification of $K$-nearest valid neighbours after sorting requires $O(NK) = O(N)$ for fixed $K$. This compares favourably with $O(N^2 \log N)$ for MST construction and $O(N^3)$ for diffusion pseudotime.

### *3.4. Loop circularity diagnostic*

To quantify the conditions under which the coordinate prior is expected to provide useful signal, we fit a circle to the two-dimensional embedding by the Taubin algebraic method, obtaining centre $(c_x, c_y)$ and radius $R$. The loop score is defined as

$$\text{Loop score} = \frac{\text{arc length along } \theta\text{-ordered trajectory}}{2\pi R}. \tag{3}$$

A perfect circle yields a loop score of 1.0. The normalised root-mean-square deviation from the best-fit circle, RMSD/$R$, provides a complementary circularity measure. Results and limitations of this diagnostic are presented in Section 4.5.

### *3.5. Evaluation metrics*

Structural accuracy (SA) denotes the fraction of directed edges pointing forward in the known temporal ordering, with a random baseline of approximately 50%. On the Spellman dataset ($N$=18), a single edge inversion alters SA by approximately 5 pp, so Spellman results are reported as directional evidence rather than precise estimates.

Sliding-window structural accuracy (SW-SA) serves as the primary metric for larger datasets. For $N$ cells sorted by arc length, with $\phi_i$ denoting the integer phase rank (G1=1, S=2, G2M=3) at position $i$,

$$\text{SW-SA} = \frac{1}{N-w} \sum_{i=1}^{N-w} \mathbf{1}[\phi_i \leq \phi_{i+w}] \tag{4}$$

with window spacing $w = 10$. Pairwise method comparisons employ bias-corrected and accelerated (BCa) bootstrap confidence intervals computed over per-window binary correctness vectors with 10,000 resamples.

## 4. Results

### *4.1. Synthetic validation*

We generated 50 cyclic synthetic manifolds at each of three radial noise levels ($\sigma \in \{0.3, 0.8, 1.5\}$) with $N$=30 nodes each. ACPL outperformed Cartesian MST on every seed at low and medium noise levels (Wilcoxon $W$=1275, the maximum possible value for 50 paired tests; Table 1). The graceful degradation from 80.7% to 61.8% as noise increases is consistent with a method that exploits genuine geometric structure rather than one that overfits to it. Full distributional separation across the 50 seeds is depicted in Supplementary Figure S2 and is not reproduced in the main text to conserve space.

**Table 1.** Synthetic stress test (mean ± SD, $N$=50 seeds, $K$=2). $W$=1275 is the maximum possible Wilcoxon statistic.

| Noise | Cartesian | ACPL | $p$ |
|---|---|---|---|
| $\sigma$=0.3 | $41.6 \pm 3.3$ | $80.7 \pm 6.6$ | $7.8\times10^{-10}$ |
| $\sigma$=0.8 | $44.0 \pm 3.7$ | $69.9 \pm 7.1$ | $7.8\times10^{-10}$ |
| $\sigma$=1.5 | $46.0 \pm 3.7$ | $61.8 \pm 8.1$ | $9.3\times10^{-10}$ |

### *4.2. Spellman yeast: Ablation and embedding*

On the Spellman yeast cell cycle [Spellman et al., 1998] ($N$=18 synchronised microarray timepoints), ACPL achieves 97.0% structural accuracy on UMAP coordinates compared with 66.7% on PCA coordinates. The complete ablation table is given in Supplementary Table S1 to conserve space. The principal findings are summarised here. The polar transform alone achieves 51.5% SA, statistically indistinguishable from the 52.8% Cartesian baseline. Polar+LOESS on the radial coordinate $r$ reduces accuracy to 42.4%, confirming that $r$ does not carry useful temporal information on a closed-loop manifold. Raw arc length achieves 100%$^\dagger$, the degenerate result of Theorem 1, whereas ACPL itself achieves 66.7% as the principled outcome. The UMAP embedding is the dominant determinant of performance: applying the identical pipeline to PCA coordinates achieves only 66.7%, because PCA does not encode temporal sequence into layout geometry in the manner that UMAP does (Supplementary Figure S1).

### *4.3. Nestorowa ablation at scale*

Repetition of the ablation study on the Nestorowa HSC dataset [Nestorowa et al., 2016] ($N$=1,920) reveals a qualitatively distinct pattern. The polar transform alone achieves 75.9% SW-SA, identical to that of the full ACPL pipeline (Figure 1). On this manifold, angular ordering by $\theta$ captures substantially all of the available temporal signal, and the arc length computation and LOESS smoothing contribute no measurable additional accuracy. The Polar+LOESS on $r$ configuration achieves 72.3%, again confirming that $r$ is uninformative across scales. LOESS smooth-

**Table 3.** All computed pairwise bootstrap CIs (95% BCa, 10,000 resamples). All three CIs include zero. The Leng mESC bootstrap could not be computed owing to a dataset loading failure; the raw SW-SA values are reported in Table 4.

| Dataset | Comparison | Observed (pp) | 95% BCa CI | $p$ | Significant |
|---|---|---|---|---|---|
| Nestorowa HSC | ACPL vs. MST | +1.52 | $[-1.31, +4.24]$ | 0.28 | No |
| Nestorowa HSC | ACPL vs. Slingshot | −1.47 | $[-4.19, +1.26]$ | 0.29 | No |
| Buettner mESC | ACPL vs. MST | −1.44 | $[-8.63, +5.04]$ | 0.69 | No |
| Leng mESC | ACPL vs. MST | −6.7 (raw) | not computed | n/a | n/a |

ing remains theoretically necessary by Theorem 1 but contributes no practical accuracy benefit on this manifold. A useful implication for practitioners is that on datasets where UMAP produces a well-structured angular gradient, a simple $\theta$-based ordering may perform equivalently to the full pipeline.

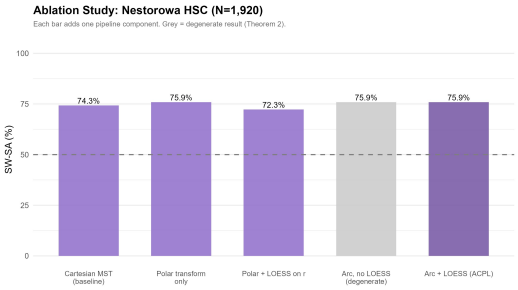


**Figure 1.** Nestorowa ablation at scale. The polar transform alone and ACPL both achieve 75.9%, indicating that $\theta$ ordering captures the bulk of the available temporal signal on this manifold. The grey bar denotes the degenerate raw arc length result.

### *4.4. Five-method benchmark on Nestorowa HSC*

Table 2 presents the complete five-method comparison. Slingshot achieves the highest SW-SA at 77.3%, followed by ACPL at 75.9%, PAGA+DPT at 75.1%, and MST at 74.3%. The four methods are distributed within a range of 3 pp, and all three pairwise bootstrap confidence intervals computed against ACPL include zero (Figure 2, Table 3). No comparison attains conventional statistical significance.

**Table 2.** Five-method benchmark on Nestorowa HSC ($N$=1,920).

| Method | SW-SA |
|---|---|
| Slingshot [Street et al., 2018] | 77.3% |
| **ACPL** | **75.9%** |
| PAGA+DPT [Wolf et al., 2019] | 75.1% |
| MST (Monocle-style) [Qiu et al., 2017] | 74.3% |

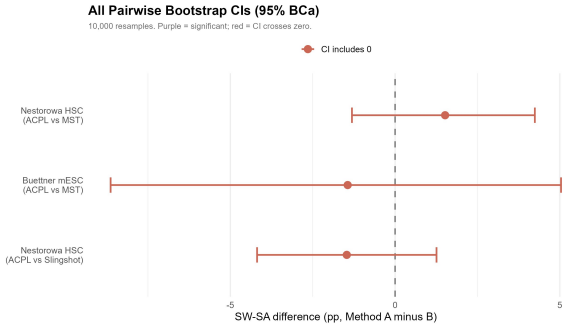


**Figure 2.** Pairwise bootstrap confidence intervals (95% BCa). All three computed comparisons cross zero, indicating absence of statistically significant differences. The Leng mESC comparison is omitted owing to a dataset loading failure.

The practical distinction between ACPL and Slingshot at this scale does not rest on aggregate accuracy. Slingshot requires specification of a starting cluster; ACPL requires no such input, a distinction that acquires importance in automated pipelines or datasets for which no reliable root cell can be identified a priori.

### *4.5. Curvature and loop circularity: two null results*

Two analyses conceived as explanations of the Nestorowa advantage returned null or inconclusive results. First, the spatial-localisation hypothesis held that ACPL corrections concentrate at high-curvature manifold junctions. The residual map in Supplementary Figure S3 is qualitatively consistent with this hypothesis, but quantitative testing did not support it. Local curvature via three-point circumradius gave mean curvature 0.969 at ACPL-correct windows compared with 1.090 at MST-correct windows, opposite to the predicted direction (Wilcoxon $W$=66,703, $p$=0.32; Spearman $\rho = -0.005$, $p$=0.81). Full per-category statistics are provided in Supplementary Table S2. The spatial-localisation claim has been withdrawn.

Second, the loop circularity diagnostic was computable for two of the four intended datasets (Figure 3); Spellman and Leng failed to load owing to

package incompatibilities. The two available points exhibit a counterintuitive pattern. Nestorowa HSC, with loop score 71.79 corresponding to a sprawling non-circular manifold, is where ACPL achieves a directional advantage; Buettner mESC, with loop score 17.45 corresponding to a more circular embedding, is where MST performs better. The most plausible explanation is that the loop score measures the wrong geometric property. The predictor of ACPL advantage is not circularity but *angular monotonicity*, the degree to which $\theta$ around the adaptive hub increases consistently with biological time, an interpretation developed in the Discussion.

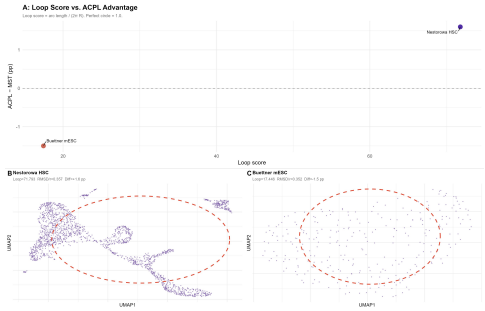


**Figure 3.** Loop circularity diagnostic. Panel A: loop score plotted against the ACPL advantage for the two loaded datasets. Panels B and C: best-fit circles overlaid on UMAP embeddings. Nestorowa scores high (sprawling non-circular manifold); Buettner scores low (compact elliptical loop). The counterintuitive relationship motivates the angular monotonicity reframing developed in the Discussion.

### *4.6. Multi-dataset benchmark*

ACPL is outperformed by MST on three of the five biological datasets evaluated (Table 4). The Leng mESC result (−6.7 pp on FACS-validated experimental labels) constitutes the most constraining finding in the work: on synchronised data, ACPL nevertheless produces meaningfully poorer ordering than MST when the UMAP embedding does not present a consistent angular gradient relative to biological time. No pairwise bootstrap confidence interval excludes zero. The full multi-dataset visualisation appears in Supplementary Figure S4.

### *4.7. Biological anchor validation*

To verify that the ACPL arc length ordering recovers biological progression rather than merely geometric structure, four canonical cell-cycle markers were evaluated on the Nestorowa HSC dataset. A naive ANOVA on phase markers would yield anti-conservative $p$-values: cells adjacent in arc length ordering are spatially autocorrelated, with Moran's $I \in [0.231, 0.339]$ for all four markers ($p \approx 0$), violating the independence assumption. A linear mixed model implemented in `lme4`, with phase as a fixed effect and arc length bin as a random effect, accounts for this dependence. All four markers survive the correction by hundreds of orders of magnitude (mixed model likelihood ratio test $p \ll 10^{-150}$ in all cases), with Top2a and Mki67 peaking in G2M and Pcna and Mcm2 peaking in S phase, in accordance with the established literature. Full per-marker statistics are provided in Supplementary Table S5, and marker expression profiles along the arc length axis are shown in Supplementary Figure S5.

### *4.8. Parameter invariance and UMAP sensitivity*

SW-SA is completely invariant across LOESS span $\in [0.05, 0.95]$ at all tested $K$ values, with a total variance of 0.22 pp (Supplementary Figure S6). The invariance reflects the observation from the Nestorowa ablation: when $\theta$ ordering already captures the principal temporal signal, variation in the LOESS span affects arc length values but not the rank ordering measured by SW-SA. Sensitivity to UMAP hyperparameters was assessed across 25 combinations of $n\_neighbors \in \{10, 20, 30, 40, 60\}$ and $min\_dist \in \{0.1, 0.2, 0.3, 0.5, 0.8\}$. SW-SA ranges from 73.7% to 77.1% with a standard deviation of 0.77 pp (Figure 4). The 3.4 pp range is modest relative to the between-method differences reported above, and the variation does not follow a systematic direction that would permit selective parameter choice.

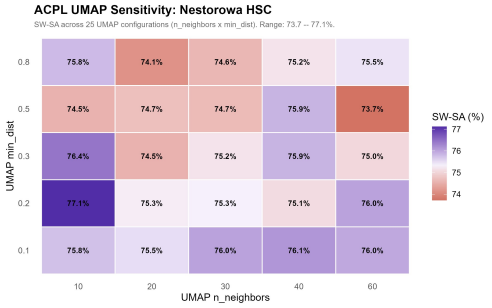


**Figure 4.** UMAP hyperparameter sensitivity. SW-SA across 25 combinations of $n\_neighbors$ and $min\_dist$ ranges from 73.7% to 77.1% (SD = 0.77 pp). Results are practically robust across the tested grid.

### *4.9. HeLa: the applicability boundary*

On asynchronous HeLa cell cycle data (GEO accession GSE64016, $N$=349 cells), all five trajectory inference methods score below the 50% random baseline: Cartesian MST 13.7%, diffusion pseudotime 13.0%,

**Table 4.** Five-dataset benchmark. MST outperforms ACPL on three of five datasets. No pairwise bootstrap CI excludes zero.

| Dataset | Labels | $N$ | ACPL | MST | Diff (pp) |
|---|---|---|---|---|---|
| Spellman yeast | Microarray timepoints | 18 | 97.0% | 52.8% | +44.2 |
| Nestorowa HSC | Seurat phase scoring | 1,920 | 75.9% | 74.3% | +1.6 |
| Buettner mESC | FACS validation | 288 | 75.5% | 77.0% | −1.5 |
| Leng mESC | FACS sorting | 247 | 89.5% | 96.2% | −6.7 |
| Richard T Cells | Seurat phase scoring | 572 | 67.8% | 69.2% | −1.4 |

ACPL on UMAP 12.1%. Asynchronous populations do not form a directed loop in UMAP because they lack a canonical starting point and a consistent angular gradient. Every geometry-dependent method fails on these data, and ACPL fails no worse than the alternatives. This boundary condition defines the applicability scope of ACPL precisely.

*4.10. Atlas-scale runtime benchmark*

Runtime to $N$=50,000 was measured for all four comparison methods (Supplementary Figure S8). Three findings warrant emphasis. First, MST becomes memory-constrained above $N \approx 5{,}000$, requiring 17.8 seconds at that scale and failing to complete at larger sizes. Second, ACPL scales with empirical exponent 1.86 and reaches 3.97 seconds at $N$=50,000. Third, Slingshot scales more favourably than ACPL (empirical exponent 0.50) and is projected to be faster than ACPL above $N \approx 10{,}000$ (Supplementary Table S3). This finding revises a claim made in an earlier version of the present work: ACPL's practical advantage at large $N$ rests on the absence of prerequisites rather than on raw computational speed.

*4.11. Three-dimensional helix lift*

On synthetic two-cycle data ($N$=200), the three-dimensional helix lift correctly separates the two cycle laps in the $x$–$z$ projection (Supplementary Figure S7). Both the two-dimensional collapsed version and the three-dimensional helix yield identical SW-SA of 94.7% (Supplementary Table S4), a result of the synthetic protocol in which phase labels repeat in the same order across cycles. A definitive test requires multi-cycle data in which cycle number and phase vary independently.

## 5. Discussion

*5.1. Interpretation of the multi-dataset results*

The benchmark across five biological datasets yields a mixed empirical picture. ACPL achieves decisive superiority on Spellman, where the manifold is a near-perfect circle derived from synchronised microarray data, and a directional but non-significant advantage on Nestorowa. ACPL underperforms MST on Buettner, Leng, and Richard T Cells. The Leng reversal (−6.7 pp on FACS-validated experimental labels) constitutes the most informative failure in the present analysis. Synchronisation of the cell population alone is insufficient for ACPL to provide an advantage; the UMAP embedding must additionally present an angular gradient that aligns with biological time, a property of the embedding rather than of the experimental data.

The loop circularity analysis was conceived as a means of operationalising this applicability condition, and its partial completion produced a result that points toward the necessary refinement. The geometric property that predicts ACPL advantage is angular monotonicity, defined as the alignment of the angular coordinate $\theta$ with biological time, rather than geometric circularity of the embedding. The Nestorowa HSC manifold satisfies angular monotonicity along its principal progenitor trunk despite manifest non-circularity; the Buettner manifold may not satisfy this property despite greater geometric circularity. Computation of loop scores for the Spellman and Leng datasets, which exhibit the largest positive and negative ACPL effects in the present benchmark, is identified as the priority follow-up experiment.

*5.2. Revisions to earlier claims*

Three claims from earlier versions of this work were revised during the course of the present analysis. The spatial-localisation claim maintained that ACPL corrections concentrate at high-curvature manifold junctions; quantitative testing via three-point circumradius returned a null result ($p$=0.32; $\rho$=−0.005) and the claim has been withdrawn. The runtime claim maintained that the $O(N \log N)$ complexity of ACPL

rendered it uniquely suitable for atlas-scale data; the atlas benchmark demonstrates that Slingshot scales more favourably than ACPL above $N \approx 10{,}000$ (empirical exponents 0.50 versus 1.86), and the claim has been revised to emphasise the absence of prerequisites rather than computational speed. The statistical significance claim treated the Nestorowa +1.6 pp advantage as meaningful evidence; bootstrap resampling demonstrates that the corresponding 95% BCa confidence interval includes zero, and the advantage is now described as directional evidence rather than as an established effect.

*5.3. Toward a theory of angular alignment*

We define angular monotonicity informally: a two-dimensional embedding is *angularly monotone* with respect to a biological process if there exists a hub $c^*$ such that the angular coordinate $\theta$ around $c^*$ increases monotonically with true biological time across the majority of the cell population. A computable pre-check based on this condition would prove more predictive than the loop score. The deeper implication is that trajectory inference success depends on coordinate alignment between embedding geometry and biological flow, a proposition holding not only of ACPL but of the entire class of geometry-dependent methods.

Future directions: the method could be made embedding-independent through joint optimisation of the embedding and the angular ordering; a hybrid pipeline could apply ACPL to cyclic components and lineage-aware methods to branching structure; the three-dimensional helix lift requires validation on real multi-cycle datasets in which cycle number and phase vary independently.


## Acknowledgements

The author thanks reviewers whose questions produced the loop circularity diagnostic, helix validation, pairwise bootstrap analysis, and curvature quantification. The author thanks Samiran Ghosh (Chair, Department of Biostatistics and Data Science, UTHealth Houston School of Public Health) for informal statistical consultation on the bootstrap design and mixed-model specification. AI-assisted tools (Anthropic Claude) were used for LaTeX typesetting, prose refinement, and consistency checks, in accordance with ISCB policy. All scientific content, methodology, theorems, proofs, analyses, and conclusions are the author's own work. Full disclosure of AI use is provided in Supplementary Section S9.

## Conflict of interest

None declared.


## Data availability

Spellman yeast (`yeastCC`), Nestorowa HSC, Buettner mESC, Leng mESC, and Richard T-Cell data (`scRNAseq`) are available via Bioconductor. HeLa cell cycle data are available from GEO accession GSE64016.

## Code availability

The R implementation is available at https://github.com/shreyosecret/ACPL under the MIT License. The repository contains twelve numbered analysis scripts, unit tests, a minimal worked example, and a master script (`run_all.R`) that executes the full pipeline end-to-end. A version-tagged release will be archived at Zenodo upon journal acceptance.

## References


Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. (2020). Generalizing RNA velocity to transient cell states through dynamical modelling. *Nature Biotechnology* 38:1408–1414.

Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods* 13:845–848.

La Manno G, Soldatov R, Zeisel A, et al. (2018). RNA velocity of single cells. *Nature* 560:494–498.

McInnes L, Healy J, Melville J. (2018). UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *arXiv:1802.03426*.

Nestorowa S, Hamey FK, Pijuan-Sala B, et al. (2016). A single-cell resolution map of mouse haematopoietic stem and progenitor cell differentiation. *Blood* 128(8):e20–e31.

Qiu X, Mao Q, Tang Y, et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods* 14:979–982.

Saelens W, Cannoodt R, Todorov H, Saeys Y. (2019). A comparison of single-cell trajectory inference methods. *Nature Biotechnology* 37:547–554.

Spellman PT, Sherlock G, Zhang MQ, et al. (1998). Comprehensive identification of cell cycle-regulated genes of *S. cerevisiae* by microarray hybridization. *Molecular Biology of the Cell* 9(12):3273–3297.

Street K, Risso D, Fletcher RB, et al. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19:477.

Trapnell C, Cacchiarelli D, Grimsby J, et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* 32:381–386.

Wolf FA, Hamey FK, Plass M, et al. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology* 20:59.