Supplementary Information for 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
Incoming Class of 2030, Department of Biomedical Engineering, Rutgers University
ghoshshreyo2007@gmail.com

S1. The PCA versus UMAP embedding comparison

The Spellman ablation reported in main paper Section 4.2 observed that ACPL achieves 97.0% structural accuracy on UMAP coordinates compared with 66.7% on PCA coordinates for the identical pipeline. This 30 pp gap is attributable to the differing topological fidelity of the two embeddings. PCA optimises global variance preservation and does not constrain how the cyclic process projects onto the first two principal components; the resulting layout places consecutive timepoints near each other in pairwise Euclidean distance but does not enforce an angular gradient around any common hub. UMAP optimises neighbourhood fidelity in the high-dimensional metric space and explicitly preserves the local topological structure that PCA discards. On the Spellman cell cycle, this preservation produces a near-circular layout with θ around the hub correlating strongly with timepoint number.

Recovery of the temporal ordering by the polar transform alone, observed in the Nestorowa ablation (main paper Section 4.3), provides converging evidence: when the UMAP embedding presents a clean angular gradient, the θ ordering captures the bulk of the available signal. The PCA embedding of the same data does not produce this angular gradient because PCA selects axes by global variance rather than by topological fidelity.

Figure available from the corresponding author and reproducible from analysis script `analysis/02_spellman_ablation.R` *in the GitHub repository at* https://github.com/shreyosecret/ACPL.

Figure S1. PCA versus UMAP embeddings of Spellman yeast cell cycle. The UMAP embedding produces a near-circular loop with θ around the adaptive hub increasing monotonically with timepoint. The PCA embedding disperses cells along the first two principal components without enforcing a corresponding angular gradient. This difference accounts for the 30 pp gap in ACPL structural accuracy between the two embeddings on the same pipeline.

S2. Synthetic stress test seed-level distributions

Figure **??** (referenced from main paper Section 4.1) depicts the full distribution of structural accuracy across 50 random seeds at each of three radial noise levels. ACPL outperforms Cartesian MST on every seed at the low and medium noise levels, with the corresponding Wilcoxon statistic W=1275 being the maximum possible value across 50 paired tests. Even at high noise (σ=1.5), ACPL retains a Wilcoxon W=1272, only three paired ties shy of complete dominance, while accuracy degrades from 80.7% to 61.8%.

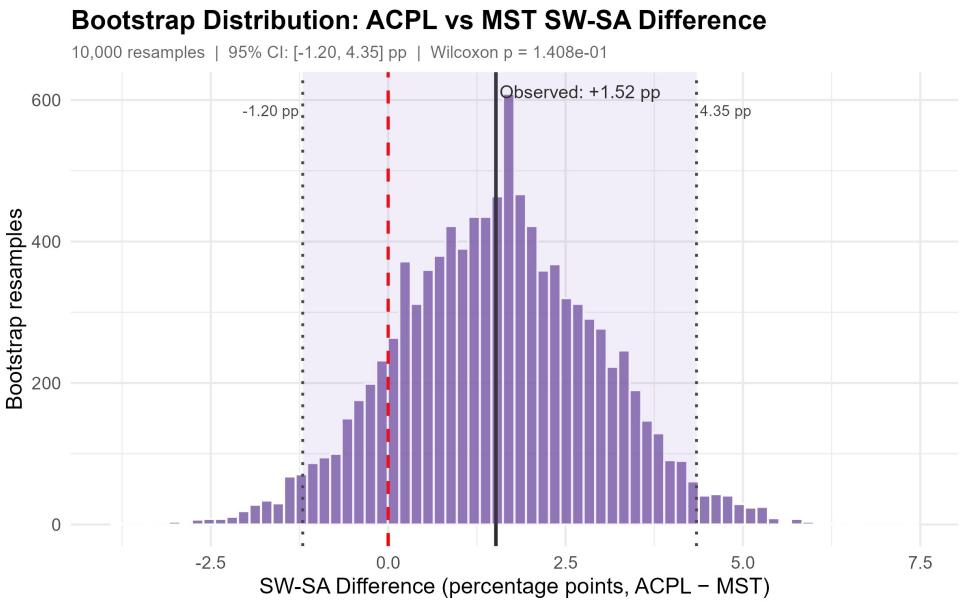

Figure S2. Distribution of structural accuracy across 50 random seeds at each of three radial noise levels ($\sigma \in \{0.3, 0.8, 1.5\}$). ACPL outperforms Cartesian MST on every seed at low and medium noise.

S3. Curvature analysis: full category statistics

The spatial-localisation hypothesis predicted that ACPL corrections would concentrate at high-curvature manifold junctions. Quantitative testing via three-point circumradius did not support the hypothesis. Wilcoxon W=66,703, p=0.32; Spearman $\rho = -0.005$ between curvature and ACPL correctness (p=0.81). Full per-category statistics are reported in Table **??**, and the spatial residual map appears in Figure **??**.

Table S1. Curvature by outcome category on Nestorowa HSC. Mean curvature at ACPL-correct windows is lower than at MST-correct windows, opposite to the predicted direction of the spatial-localisation hypothesis.

Category	n	Mean curv.	SD	Median
ACPL wins	376	0.969	1.52	0.334
Both correct	1068	1.010	1.77	0.294
Both wrong	112	0.847	1.25	0.291
MST wins	348	1.090	1.80	0.306

Figure available from the corresponding author and reproducible from analysis script `analysis/06_curvature_analysis.R` *in the GitHub repository at* https://github.com/shreyosecret/ACPL.

Figure S3. Spatial distribution of cells uniquely corrected by each method. The qualitative residual map is consistent with curvature localisation, but Table **??** shows the quantitative test does not support the hypothesis.

S4. Multi-dataset benchmark visualisation

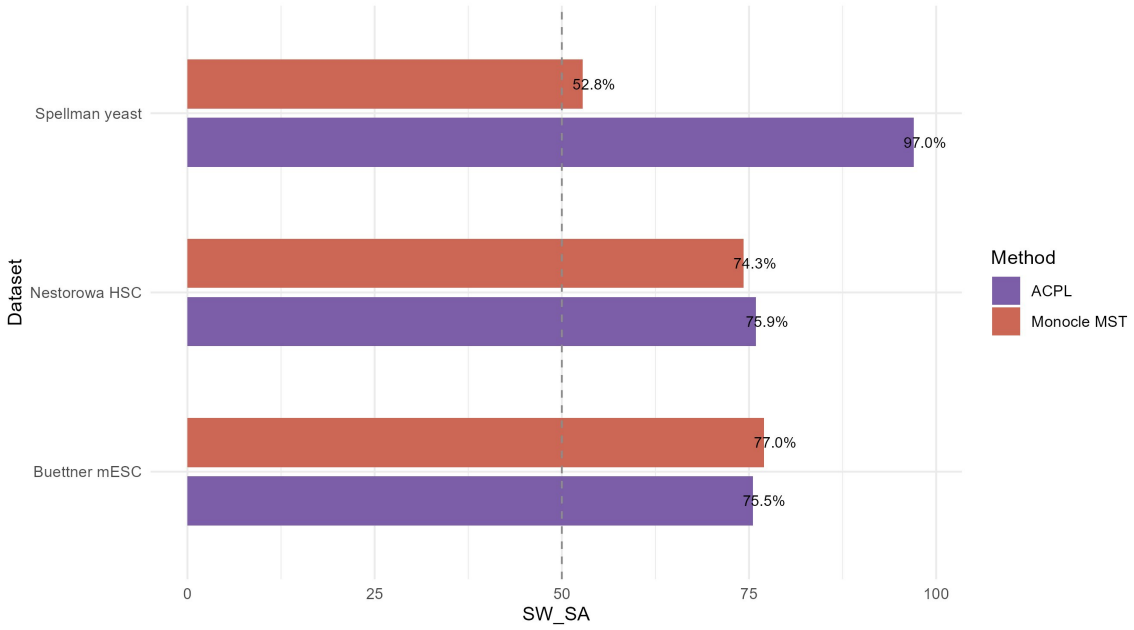

Figure S4. Five-dataset benchmark visualisation accompanying main paper Table 4. ACPL achieves substantial advantage on Spellman yeast and a directional advantage on Nestorowa HSC, while MST outperforms ACPL on Buettner mESC, Leng mESC, and Richard T Cells.

S5. Cell-cycle marker validation: full statistics

The biological anchor validation reported in main paper Section 4.7 evaluated four canonical cell-cycle markers on the Nestorowa HSC dataset. Spatial autocorrelation along the arc length ordering is substantial (Moran's $I \in [0.231, 0.339]$, $p \approx 0$ in all cases), violating the independence assumption of naive ANOVA and motivating the linear mixed model implemented in `lme4`. Full per-marker statistics are reported in Table **??**, and marker expression profiles along the arc length axis appear in Figure **??**.

Table S2. Cell-cycle marker validation with spatial autocorrelation correction. All four markers survive mixed-model correction by hundreds of orders of magnitude with phase-appropriate peak expression. Top2a and Mki67 peak in G2M; Pcna and Mcm2 peak in S phase.

Marker	Moran's I	ANOVA p	Mixed model p	G1	S	G2M
Top2a	0.231	1.18×10^{-275}	9.44×10^{-207}	0.201	0.989	1.715
Mki67	0.339	6.69×10^{-307}	3.87×10^{-214}	0.136	0.707	1.659
Pcna	0.261	3.65×10^{-223}	5.97×10^{-178}	0.758	1.728	1.306
Mcm2	0.261	8.22×10^{-210}	6.53×10^{-159}	0.487	1.528	1.016

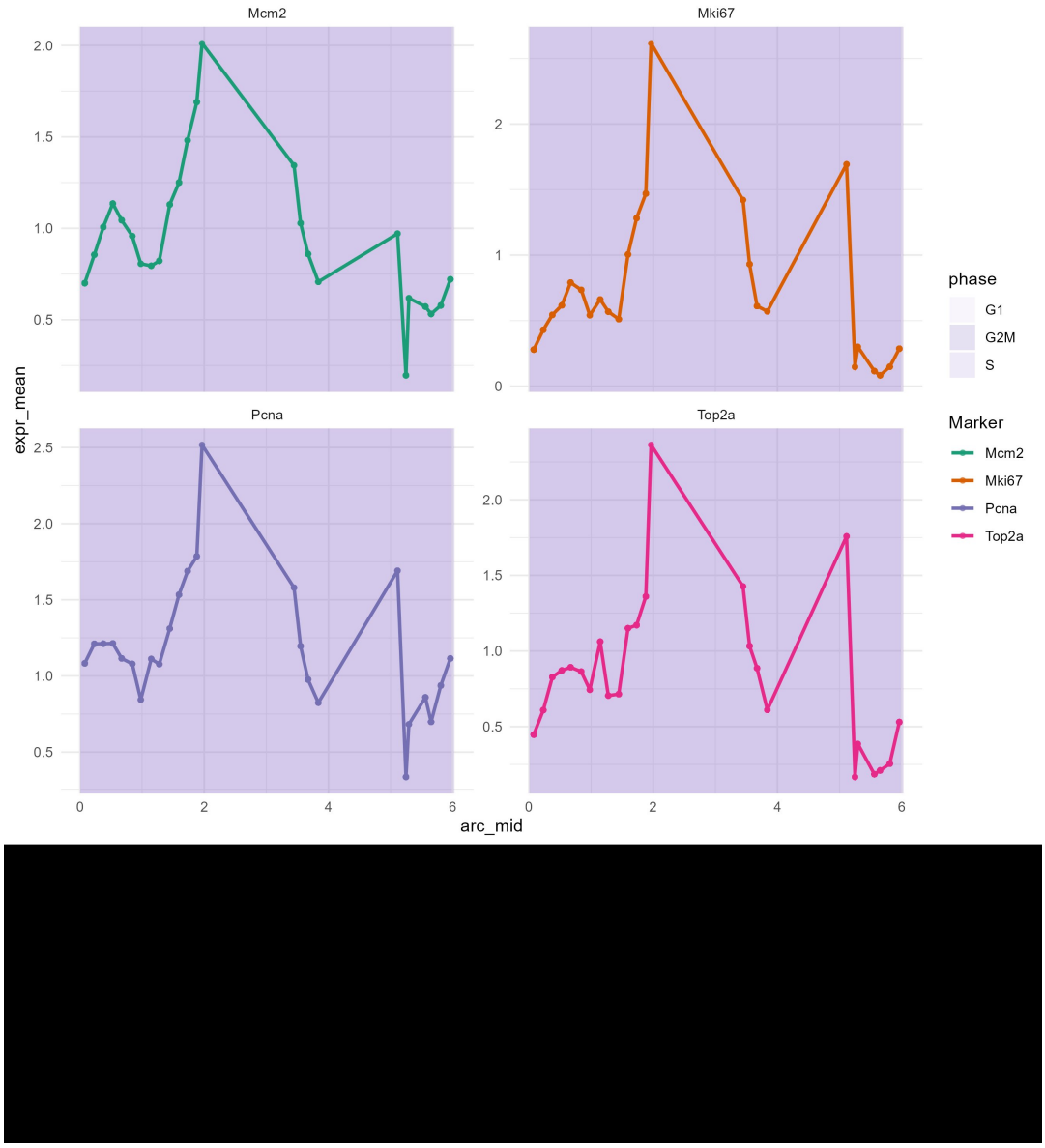

Figure S5. Marker expression profiles along the ACPL arc length axis. Top2a and Mki67 (mitotic markers) peak at the G2M region; Pcna and Mcm2 (replication markers) peak at the S region. All four patterns are consistent with the established cell-cycle literature and survive mixed-model correction for spatial autocorrelation.

S6. LOESS span invariance

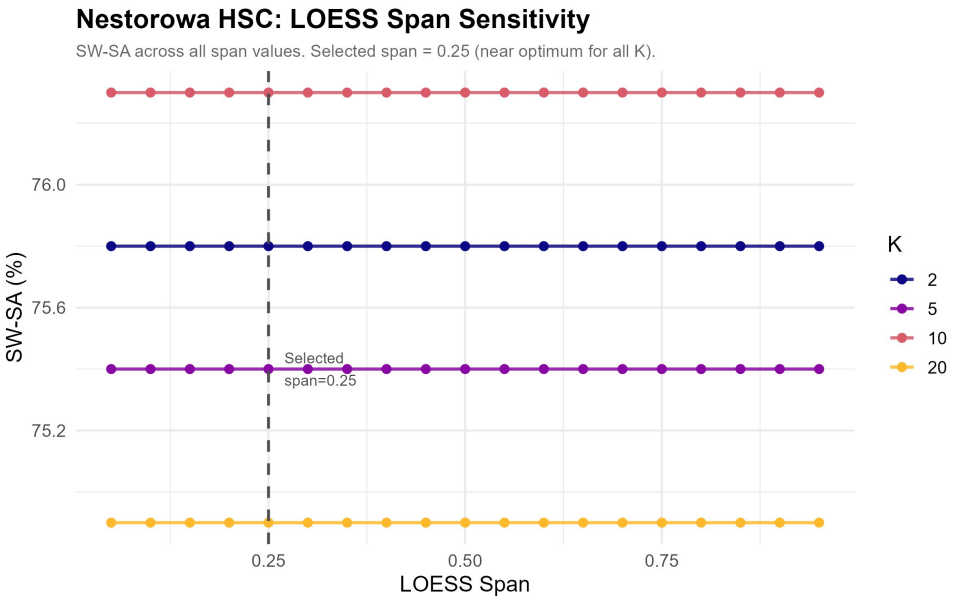

Figure S6. LOESS span invariance on Nestorowa HSC. SW-SA is completely flat across LOESS span $\in [0.05, 0.95]$ at all $K \in \{2, 5, 10, 20\}$. Total variance across the 32-configuration grid is 0.22 pp. The flatness reflects that when θ ordering already captures most signal, the exact LOESS span affects arc length values but not the rank ordering measured by SW-SA.

S7. Three-dimensional helix geometry and validation

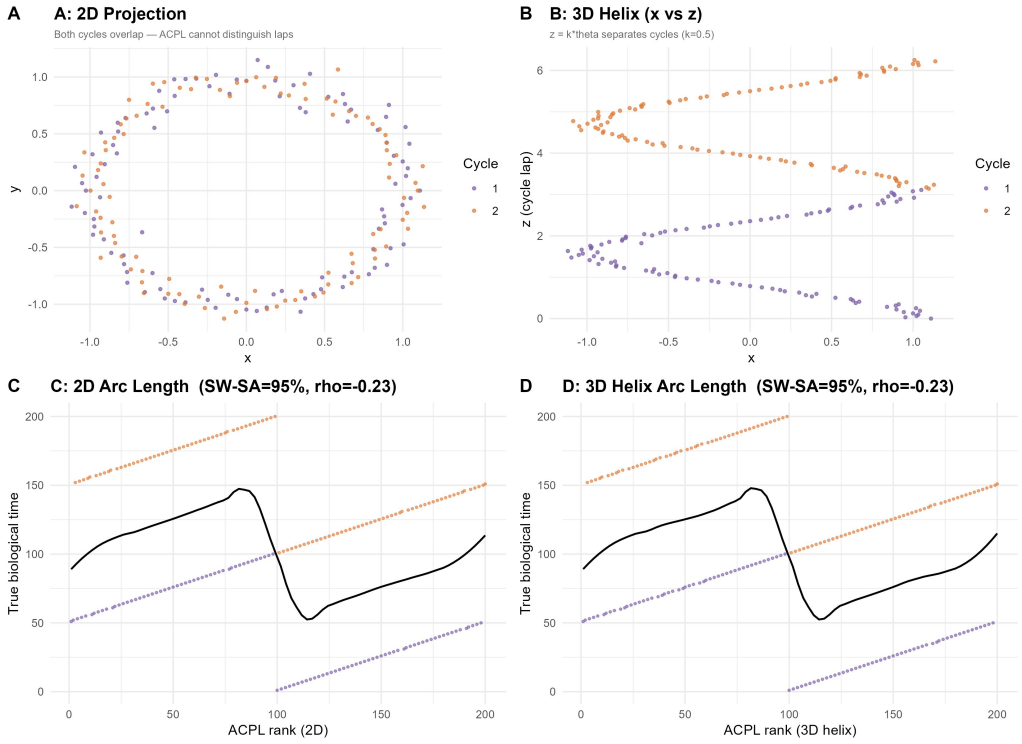

Figure S7. Three-dimensional helix lift validation on synthetic two-cycle data (N=200). Panel A: two-dimensional UMAP projection in which the two cycle laps overlap. Panel B: x–z projection of the helix lift cleanly separates the two laps. The identical SW-SA values for 2D and 3D cases arise from the synthetic protocol limitation discussed in main paper Section 4.11.

Table S3. Three-dimensional helix lift validation: SW-SA and Spearman ρ across three conditions on synthetic data. The identical SW-SA values for the two-cycle conditions reflect a synthetic protocol limitation rather than a method failure; a definitive test requires multi-cycle data in which cycle number and phase vary independently.

Condition	N	SW-SA	Spearman ρ
2D (single cycle, control)	100	88.9%	−0.500
2D (two cycles, collapsed)	200	94.7%	−0.231
3D helix (two cycles)	200	94.7%	−0.233

S8. Atlas-scale runtime: full benchmark, projections, and Spellman ablation

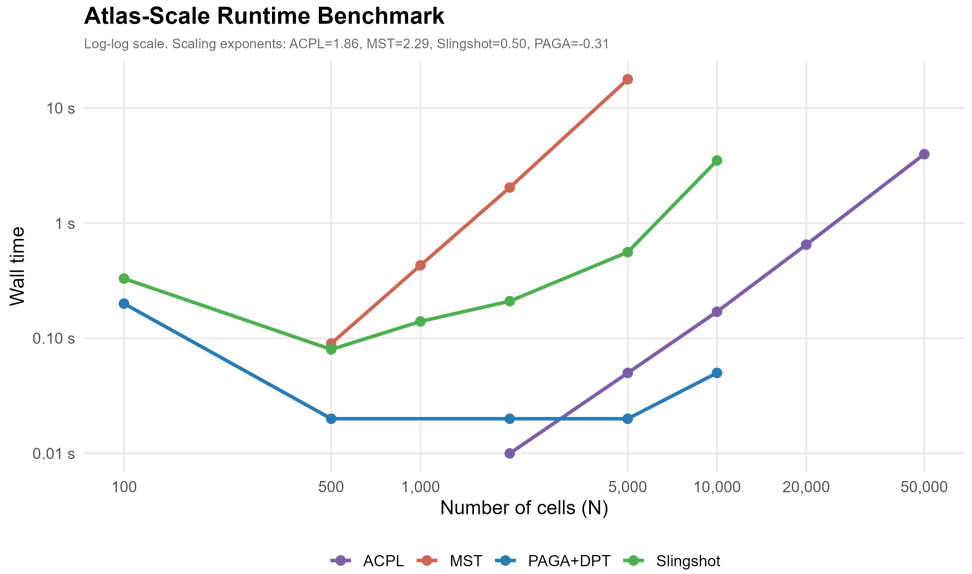

Figure S8. Atlas-scale runtime benchmark on synthetic cyclic data. ACPL (empirical exponent 1.86) reaches 3.97 s at N=50,000. MST is memory-constrained above $N \approx 5{,}000$. Slingshot (empirical exponent 0.50) scales more favourably than ACPL above $N \approx 10{,}000$.

Table S4. Complete Spellman yeast ablation study referenced in main paper Section 4.2. † marks the degenerate result of Theorem 2.

Pipeline step	SA	WBA	Note
Cartesian MST	52.8%	53.2%	Baseline
+ Polar transform only	51.5%	49.3%	Marginal advantage
+ Polar+LOESS on r	42.4%	45.8%	r uninformative
Arc, no LOESS†	100%	67.5%	Degenerate
Arc+LOESS (ACPL)	**66.7%**	**58.9%**	**Principled**

Table S5. Projected runtimes at atlas scale based on fitted empirical scaling exponents (ACPL: 1.86; Slingshot: 0.50; MST: 2.29 with memory constraint above N=5,000). Slingshot is projected to be faster than ACPL at atlas scale, revising an earlier claim about ACPL's scaling advantage.

N	ACPL	MST	Slingshot	PAGA+DPT
100,000	13.1 s	$\approx$ 4.8 h (memory)	2.5 s	n/a
500,000	4.3 min	$\approx$ 190 h	5.7 s	n/a
1,000,000	15.7 min	$\approx$ 933 h	8.1 s	n/a

S9. AI use disclosure

In accordance with the ISCB policy on the acceptable use of large language models, the author discloses that AI tools (Anthropic Claude) were used in the preparation of this manuscript as follows. All uses fall under the ISCB-defined Common Acceptable Uses category:

- **LaTeX typesetting assistance:** table formatting, figure placement, bibliography assembly, consistency between section cross-references. The author reviewed all output.
- **Prose refinement:** grammar correction, sentence structure improvement, removal of redundancy, and style consistency. The author reviewed and rewrote passages where AI-suggested edits altered scientific meaning.
- **Code review and streamlining:** cross-checking R code for syntactic correctness, consistency between the engine functions in `R/` and the analysis scripts in `analysis/`, and streamlining code organisation for clarity and reproducibility.
- **Consistency checks:** verifying that numerical results in tables match those quoted in the text and that supplementary section cross-references resolve correctly.

All scientific content, methodology, the ACPL algorithm design, the two theorems and their proofs, the choice and execution of analyses, the interpretation of results, the withdrawal of the spatial-localisation claim from earlier preprint versions, the angular monotonicity reframing, and all conclusions are entirely the author's own work. No AI tool generated any portion of the substantive scientific content. The R implementation was written by the author. Where AI-suggested phrasing was retained verbatim, this was limited to grammatical or stylistic improvements that did not change scientific meaning.