scMomenta: A Second-Order Neural ODE Framework for Single-Cell Temporal Prediction

Anonymous authors
Paper under double-blind review

Abstract

Single-cell temporal prediction aims to reconstruct cell populations at unsampled time points from destructive measurements that provide only unpaired population snapshots. Most existing neural differential equation methods model latent cellular dynamics with a first-order equation, in which position and time alone determine the velocity. However, hidden kinetic variables, such as unspliced RNA, and dimensionality reduction can make latent dynamics depend on history that position does not record. We introduce **scMomenta**, a second-order neural ODE framework that augments the latent position with an explicit velocity state and learns a damped acceleration field. We also design an initial-velocity module that estimates each cell's velocity by optimal transport. On five time-course datasets, scMomenta achieves strong overall performance against six competing methods, improving the accuracy of single-cell temporal prediction. Experiments on latent dynamical order, trajectory crossing and velocity-space perturbation further show that an explicit velocity state captures temporal structure that position alone does not. Our code is available at `https://anonymous.4open.science/r/scMomenta-code-anonymous`.

1 Introduction

Single-cell RNA sequencing (scRNA-seq) resolves cellular states at single-cell resolution (Tang et al., 2009). However, the measurement is destructive. Because RNA capture requires cell lysis, each cell can be profiled only once. Sampling a biological process over time therefore yields unpaired population snapshots rather than cell trajectories. Constrained by experimental cost, a time course typically covers only a few time points. *Single-cell temporal prediction* aims to fill these gaps. Given the populations observed at the sampled time points, it predicts the populations at unsampled time points, either inside the observed range (*interpolation*) or beyond the last observation (*extrapolation*). Accurate prediction can reduce experimental cost and support the study of dynamic biological processes.

The prevailing approach models cellular dynamics with a differential equation in a low-dimensional latent space and parameterizes its drift with a neural network (Chen et al., 2018). Existing methods differ mainly in the constraints they add to this drift, such as structural priors, stochastic diffusion, cell–cell interactions and cell growth (Section 2). For example, PRESCIENT (Yeo et al., 2021) models the drift as the negative gradient of a potential, and scDiffEq (Vinyard et al., 2025) adds a state-dependent diffusion. Despite differences in latent representations and biological constraints, many existing methods use a first-order form,

$$\frac{\mathrm{d}z}{\mathrm{d}t} = f_\theta(z, t), \tag{1}$$

where z denotes the latent cellular state. Under this formulation, the instantaneous velocity is determined entirely by the current position and time. Consequently, two cells that occupy the same latent position are assigned the same instantaneous direction, regardless of how they arrived there.

This assumption can fail for single-cell data in three respects. *Biophysically*, RNA production involves hidden kinetic variables: in the transcription–splicing model, eliminating the unobserved

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.

unspliced abundance yields higher-order dynamics in the spliced abundance (La Manno et al., 2018; Bergen et al., 2020). *Mathematically*, even first-order gene-space dynamics generally loses first-order closure when projected onto a low-dimensional latent space (Proposition 1), so latent dynamics can depend on history that position does not record. *Empirically*, lineage tracing shows that progenitors in highly similar transcriptional states can give rise to different descendants (Weinreb et al., 2020), whereas a deterministic first-order flow assigns one velocity to each position and cannot represent crossing trajectories (Proposition 2). None of this implies that cellular dynamics is intrinsically second order, but it motivates asking whether an explicit temporal state beyond position helps model the observed populations.

We introduce **scMomenta**, a second-order neural ODE framework that augments the latent position with an explicit velocity. The velocity compactly represents temporal history, and a neural network learns the damped acceleration field that drives it. scMomenta is the first to bring second-order neural ODEs to single-cell temporal prediction. This step adds no new assumption; it removes an existing one, namely that velocity depends only on the current position. The initial velocity, however, is not observed, because cells are not paired across snapshots. We therefore design an initial-velocity module, a velocity encoder initialized with barycentric velocities from optimal transport (OT) between the first two observed snapshots and trained jointly with the dynamics. The OT coupling serves only as an initialization, not as the true lineage correspondence. We evaluate scMomenta on five time-course datasets of developmental and reprogramming processes and compare it with six temporal prediction methods under the same protocol for interpolation and extrapolation. Beyond accuracy, we examine latent dynamical order, trajectory crossing, and level- versus velocity-space interventions as complementary evidence for explicitly modeling velocity.

Our contributions are as follows.

- We introduce **scMomenta**, a second-order neural ODE framework for single-cell temporal prediction that explicitly models latent position and velocity, providing a compact state representation for history-dependent dynamics.
- We develop an OT-based initialization scheme for the otherwise unobserved initial velocity in unpaired population snapshots, using barycentric transport velocities to initialize a learnable velocity encoder.
- We conduct comprehensive evaluations on five time-course datasets, covering interpolation and extrapolation, and compare against six existing methods. Additional experiments investigate latent dynamical order, trajectory crossing, and velocity-space interventions.

2 Related Work

Single-cell temporal prediction. Neural differential equations are a common framework for predicting single-cell population dynamics from time courses. MIOFlow (Huguet et al., 2022) learns population flows in a manifold-preserving latent space, and scNODE (Zhang et al., 2024) uses a neural ODE in the latent space of a variational autoencoder. PRESCIENT (Yeo et al., 2021) and PI-SDE (Jiang & Wan, 2024) constrain the drift with a potential, scDiffEq (Vinyard et al., 2025) adds state-dependent diffusion, and scIMF (Jiang et al., 2026) models population-level interactions through a mean-field drift. TIGON (Sha et al., 2024) and DeepRUOT (Zhang et al., 2025) use unbalanced optimal transport to model growth together with motion, and Wang et al. (2025) parameterize growth explicitly. Despite different biological assumptions, these methods generally represent the cellular state by position alone, whereas scMomenta adds velocity and learns an acceleration field.

Waddington-OT (Schiebinger et al., 2019) uses optimal transport to infer ancestor–descendant couplings between consecutive snapshots, whereas we use it only to initialize the unobserved velocity. DMSB (Chen et al., 2023) and 3MSBM (Theodoropoulos et al., 2025) build Schrödinger bridges in position–velocity space, but they connect prescribed marginals and are not designed to predict beyond the last observation, so we do not include them as baselines.

Second-order dynamics in machine learning. Augmented neural ODEs enlarge the state space to overcome limitations of first-order flows (Dupont et al., 2019), but their extra coordinates have no prescribed meaning. Second-order neural ODEs identify the extra state with velocity, $\dot{z} = v$ and $\dot{v} = f_\theta(z, v, t)$, as in SONODE (Norcliffe et al., 2020); related second-order dynamics appear

in latent variational models (Yıldız et al., 2019), controlled Hamiltonian systems (Zhong et al., 2020) and graph models (Rusch et al., 2022). Momentum-based formulations strictly enlarge the representable set relative to first-order flows (Sander et al., 2021), mitigate vanishing gradients and reduce the number of function evaluations (Xia et al., 2021). Unlike these models, scMomenta learns from unpaired snapshots, in which velocity is unobserved (Section 3.3).

3 METHODOLOGY

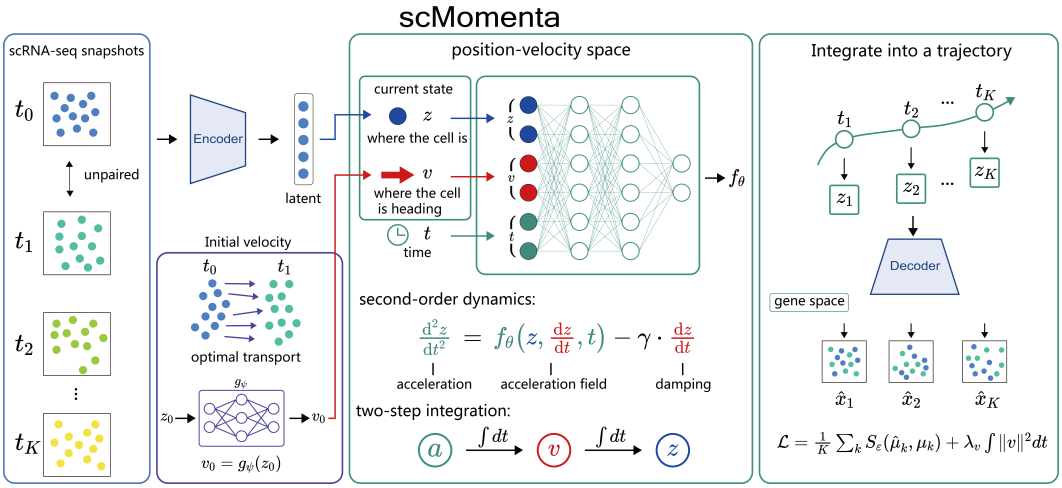

Figure 1: Overview of scMomenta. A frozen encoder maps unpaired snapshots to a latent space, optimal transport between the first two snapshots initializes the velocity encoder, and a second-order neural ODE evolves position and velocity under a learned acceleration field with damping.

3.1 PROBLEM SETTING AND THE ORDER OF THE DYNAMICS

A time-course experiment measures populations $\{\mathcal{X}_k\}_{k=0}^{K}$ at irregular sampling times $t_0 < \cdots < t_K$. Each $\mathcal{X}_k$ is a set of cells in gene space $\mathbb{R}^G$, unmatched across snapshots, which we regard as a sample from a population distribution μ_k. Given an observed index set $\mathcal{O}$ that contains 0, the task is to predict μ_k for $k \notin \mathcal{O}$; held-out indices inside the range of $\mathcal{O}$ define interpolation, and indices beyond $\max \mathcal{O}$ extrapolation. Most existing methods model the latent dynamics with the first-order form of Equation 1; the rest of this section explains why position alone may not be a sufficient state.

Higher-order dynamics from hidden kinetic variables. In the classical transcription–splicing model, the unspliced and spliced abundances u and s of a gene obey $\dot{u} = \alpha(t) - \beta u$ and $\dot{s} = \beta u - \gamma_{\deg} s$, where $\alpha(t)$ is the transcription rate, β the splicing rate and $\gamma_{\deg}$ the degradation rate (La Manno et al., 2018; Bergen et al., 2020). Standard expression matrices do not report u and s separately, so a model of s alone leaves u hidden. Eliminating u gives

$$\ddot{s} + (\beta + \gamma_{\deg})\,\dot{s} + \beta\gamma_{\deg}\, s \;=\; \beta\,\alpha(t), \tag{2}$$

which is second order in s: the level s alone does not determine its future, but s together with its rate $\dot{s}$ does. The rate carries exactly what the hidden variable adds, since at a given level $\dot{s} = \beta u - \gamma_{\deg} s$ is set by the unspliced pool u (Appendix F.3); it is also the quantity that RNA velocity estimates from unspliced and spliced counts (La Manno et al., 2018).

Order of projected dynamics. Even if gene-space dynamics were first order, the latent dynamics that a model learns need not be. Consider the simplest case, linear dynamics $\dot{x} = Ax$ with $x \in \mathbb{R}^G$ and a linear representation $z = Wx$ with $W \in \mathbb{R}^{d\times G}$ and $d \ll G$. Repeated differentiation gives

$$z^{(j)} \;=\; WA^j x, \qquad j = 0, 1, 2, \ldots. \tag{3}$$

We say that the projected dynamics *closes at order* m if $z^{(m)}$ is a linear function of $z, \dot{z}, \ldots, z^{(m-1)}$, and denote the smallest such order by m^*.

Proposition 1 (Order of projected linear dynamics). *Let $O_m = [W; WA; \ldots; WA^{m-1}] \in \mathbb{R}^{md\times G}$ be the observability matrix. Then z closes at order m if and only if $\mathrm{row}(WA^m) \subseteq \mathrm{row}(O_m)$. The minimal order m^* exists and satisfies $\lceil r/d \rceil \le m^* \le G$, where $r = \mathrm{rank}(O_G)$. Finally, $m^* = 1$ if and only if $\mathrm{row}(W)$ is an invariant subspace of A, which, when A has distinct eigenvalues, is a measure-zero condition on W.*

Intuitively, z reveals only d linear combinations of the genes, and each derivative reveals up to d more, $WA^j x$. The dynamics closes once the combinations revealed by z and its derivatives span the r-dimensional part of x on which the future of z depends (the row space of O_G), so its order is at least $\lceil r/d \rceil$; a compressed state can thus depend on information that only its derivatives carry. Beyond the linear case, the same history dependence is described by memory and delay-embedding formalisms (Mori, 1965; Zwanzig, 1973; Takens, 1981).

Distinct futures from a shared position. Cells in similar states can have different futures. Live imaging shows that PU.1 and GATA1 reinforce early myeloid lineage choice rather than initiate it (Hoppe et al., 2016), Nanog-low embryonic stem cells can differ in differentiation propensity (Filipczyk et al., 2015), and clonal barcoding gives the same picture at population scale (Weinreb et al., 2020). What distinguishes such cells must then be held by some state component other than position, and a first-order model has none. The sharpest case is a *trajectory crossing*, where two paths with different histories meet at one position and then separate, so that only the direction of arrival distinguishes them.

Proposition 2 (Injectivity of deterministic first-order flows). *Let $f(z,t)$ be locally Lipschitz in z, and let φ_t be the flow of $\dot z = f(z,t)$ from t_0 to t. Then φ_t is injective for every t. In particular, two solutions that agree at one time agree at all times.*

A deterministic first-order model therefore cannot represent a crossing, whereas a velocity state records the direction of arrival; Section 4.4 tests this on data with known coupling. A diffusion term can separate two paths from a common point, but only as fluctuation about a shared trend that does not depend on the direction of arrival.

3.2 The scMomenta Model

Latent space. We follow scNODE (Zhang et al., 2024), a first-order neural ODE for single-cell temporal prediction, and use its variational autoencoder (Kingma & Welling, 2014) to map cells to a latent space (Figure 1, left), with an encoder $(\eta(x), \sigma(x)) = \mathrm{Enc}(x)$ onto $\mathbb{R}^L$ and a linear decoder $\mathrm{Dec}(z) = z\,W_{\mathrm{dec}}^{\top} + b_{\mathrm{dec}}$. Latent positions are sampled as $z = \eta(x) + \sigma(x) \odot \epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$, the only source of stochasticity in the model. The autoencoder is pretrained on the cells of the observed time points and then frozen, so the dynamics are learned in a fixed latent geometry; from here on, we also write μ_k for the observed population in this space.

Second-order neural ODE. In each of the three observations of Section 3.1, the missing information is carried by how the state changes rather than by where it is. We therefore model the latent dynamics by a second-order equation,

$$\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} \;=\; f_\theta\!\left(z, \frac{\mathrm{d}z}{\mathrm{d}t}, t\right) \;-\; \gamma\,\frac{\mathrm{d}z}{\mathrm{d}t}, \tag{4}$$

in which the acceleration field f_θ is a neural network and $\gamma > 0$ is a damping coefficient. Introducing the velocity $v = \dot z$ as a state turns Equation 4 into a first-order system on $\mathbb{R}^{2L}$, which we integrate,

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} z \\ v \end{bmatrix} = \begin{bmatrix} v \\ f_\theta(z, v, t) - \gamma v \end{bmatrix}, \qquad \begin{bmatrix} z(t_0) \\ v(t_0) \end{bmatrix} = \begin{bmatrix} z_0 \\ g_\psi(z_0) \end{bmatrix}. \tag{5}$$

The acceleration $a = f_\theta(z, v, t) - \gamma v$ is integrated once to give v and again to give z (Figure 1, middle). Only the acceleration field is learned, while $\dot z = v$ holds by definition, so unlike in a generic augmented neural ODE (Section 2), the additional state has an explicit temporal meaning.

Damping and the first-order special case. The damping γ, a single learned scalar (Appendix C), sets the timescale $\tau = 1/\gamma$ on which the velocity relaxes towards a value set by position and time. When τ is short, velocity has no time to depart from that value.

Proposition 3 (First-order limit)**.** *Let f_θ be uniformly Lipschitz in v with constant Λ, and let $\gamma > \Lambda$. Then $\gamma v^* = f_\theta(z, v^*, t)$ has a unique solution $v^*(z,t)$ for every (z,t). Along solutions of Equation 5, $v(t) - v^*(z(t),t)$ decays at rate $\gamma - \Lambda$ to a residual of size $O(1/(\gamma - \Lambda))$, which is set by the drift of v^* along the trajectory.*

After a time long compared with τ, $\dot{z} \approx v^*(z,t)$, which is the form of Equation 1 (the scaling is given with the proof in Appendix B). In this sense the second-order model removes an assumption rather than adding one: it contains first-order dynamics as its overdamped regime, and learning γ lets the data decide how far to depart from it. We stop at second order because Proposition 1 does not fix how many orders the latent dynamics requires, and one additional derivative is the smallest step that separates cells sharing a position (Proposition 2); Section 4.3 tests whether it suffices at the latent dimension used here. The velocity is a compact representation of temporal history, not a claim that cells possess mechanical inertia.

3.3 Optimal-Transport Initialization of the Velocity Encoder

Non-identifiability of the initial velocity. A second-order system also needs an initial velocity, but a snapshot records only positions, and because measurement destroys cells, no cell at t_0 can be followed to its descendants. The initial velocity points towards the cells that a cell becomes, so it is fixed by the pairing between consecutive snapshots, not by any snapshot alone. We call such a pairing a *coupling*, the joint distribution of the latent state at two time points. The training objective of this class of methods, including ours, compares predicted and observed populations one time point at a time, $\mathcal{L}(\theta,\psi) = \sum_{k\in\mathcal{O}} S\big(\hat{\mu}_k^{\theta,\psi},\, \mu_k\big)$, where $\hat{\mu}_k^{\theta,\psi}$ is the population predicted at t_k and S compares population distributions.

Proposition 4 (Marginal matching does not identify coupling)**.** *If two parameter settings induce the same sequence of population marginals $\{\hat{\mu}_k\}_{k\in\mathcal{O}}$, then any loss that depends only on these marginals assigns them the same objective value, regardless of the couplings between their individual trajectories.*

Second-order neural ODEs usually learn the initial velocity end-to-end (Norcliffe et al., 2020). Since infinitely many couplings share the same marginals (Appendix B), learning g_ψ this way from such a loss determines the initial velocity only up to couplings that leave the marginals unchanged.

Initial velocity from optimal transport. scMomenta instead reads the initial velocity out of the transport between the first two snapshots (Figure 1, bottom left). Let $Z_0 = \{z_i\}_{i=1}^{n_0}$ and $Z_1 = \{z'_j\}_{j=1}^{n_1}$ be the latent codes of the cells at t_0 and at t_1. We solve exact optimal transport under uniform marginals and squared Euclidean cost,

$$P^* = \arg\min_{P\in\Pi(\mathbf{p},\mathbf{q})} \sum_{i,j} P_{ij}\|z_i - z'_j\|^2, \qquad \mathbf{p} = \tfrac{1}{n_0}\mathbf{1}, \quad \mathbf{q} = \tfrac{1}{n_1}\mathbf{1}, \tag{6}$$

and convert the barycentric projection of each cell into a velocity,

$$\widehat{T}(z_i) = \frac{\sum_j P^*_{ij}\, z'_j}{\sum_j P^*_{ij}}, \qquad v_0(z_i) = \frac{\widehat{T}(z_i) - z_i}{t_1 - t_0}. \tag{7}$$

We avoid entropic regularization, which would spread each cell's mass over more targets and shorten the inferred displacement. The optimal plan carries the first population onto the second with the least total squared displacement, and Equation 7 is the initial velocity of the resulting displacement interpolation, in which each cell moves in a straight line at constant velocity, tracing the Wasserstein-2 geodesic between the two populations (Villani, 2009). We warm-start g_ψ on these targets by minimizing $\frac{1}{n_0}\sum_{i=1}^{n_0}\|g_\psi(z_i) - v_0(z_i)\|^2$ and then train it jointly with the dynamics, so the OT velocity is a starting point rather than a constraint: training can move away from it wherever later snapshots require.

3.4 Training Objective and Protocol

Each trajectory starts from a cell observed at t_0 with initial velocity $g_\psi(z_0)$ and is integrated through Equation 5 to all training times, and the simulated cells at t_k form the predicted population $\hat{\mu}_k$

(Figure 1, right). With $K_{\mathcal{O}} = |\mathcal{O}| - 1$ observed time points after t_0, the training loss is

$$\mathcal{L} = \frac{1}{K_{\mathcal{O}}} \sum_{k \in \mathcal{O} \setminus \{0\}} S_\varepsilon\big(\hat{\mu}_k,\ \mu_k\big) + \lambda_v \int \|v\|^2 \,\mathrm{d}t, \tag{8}$$

where S_ε is the debiased Sinkhorn divergence (Cuturi, 2013; Genevay et al., 2018; Feydy et al., 2019); the second term keeps the velocity bounded under long integration. Training has three stages: the VAE is pretrained on the observed cells and frozen, g_ψ is warm-started on the OT targets of Equation 7, and f_θ, γ and g_ψ are then optimized jointly under Equation 8. Held-out time points are excluded from both the loss and the OT initialization.

4 Results

4.1 Experimental Setup

We use five published time courses (8–19 time points, 10,895–62,840 cells; Appendix D): zebrafish (ZB) and *Drosophila* (DR) embryogenesis, mouse fibroblast-to-iPSC reprogramming (SC), and human germ cell (GA) and embryonic pancreas (MA) development. Mechanistic experiments also use *C. elegans* embryogenesis with its fully mapped lineage (CE) (Sulston et al., 1983) and human pluripotent stem cells differentiating into pancreatic β cells (VE).

We compare with six methods from Section 2: MIOFlow, PRESCIENT, scDiffEq, scNODE, PI-SDE and scIMF. All methods use identical splits and one evaluation pipeline, with predictions decoded to gene space, and scMomenta uses a single architecture and hyperparameter setting throughout (Appendix C). We report W_2, a debiased Sinkhorn divergence in a 50-dimensional PCA space that measures distributional agreement, and Pearson, the correlation between the per-gene mean expression of predicted and observed populations, over five random seeds (Appendix E).

4.2 Temporal Prediction

scMomenta achieves the best result in nine of the ten interpolation entries and six of the ten extrapolation entries of Table 1. Extrapolation is harder for every method, since no later observation constrains the prediction.

The advantage of scMomenta is largest in extrapolation on SC, the longest time course, where it reduces W_2 by about one third relative to the best competitor. From interpolation to extrapolation, its W_2 on SC grows by about half, whereas that of scIMF, the closest competitor in SC interpolation, more than doubles; the ablation in Section 4.6 finds the largest cost of removing the velocity state in the same setting. On DR and GA extrapolation, where scIMF and scNODE are best, scMomenta is within 7% of their W_2 and ahead of the remaining methods.

Figure 2 shows the held-out ZB time points in a UMAP embedding (McInnes et al., 2018). scMomenta reproduces both their temporal ordering and the narrow branch at the lower right, whereas

Table 1: Temporal prediction results. Values are mean ± standard deviation over five seeds. The **best** and second best results are marked.

Method	ZB		SC		DR		GA		MA	
	$W_2\downarrow$	Pearson↑	$W_2\downarrow$	Pearson↑	$W_2\downarrow$	Pearson↑	$W_2\downarrow$	Pearson↑	$W_2\downarrow$	Pearson↑
Interpolation										
MIOFlow	$35.29_{\pm 1.11}$	$.9892_{\pm .0007}$	$40.68_{\pm 1.83}$	$.9660_{\pm .0028}$	$69.73_{\pm 0.74}$	$.9572_{\pm .0054}$	$79.75_{\pm 4.64}$	$.9267_{\pm .0156}$	$25.81_{\pm 2.47}$	$.9693_{\pm .0042}$
PRESCIENT	$45.53_{\pm 0.92}$	$.9693_{\pm .0017}$	$43.65_{\pm 1.16}$	$.9511_{\pm .0024}$	$67.43_{\pm 0.43}$	$.8808_{\pm .0075}$	$67.47_{\pm 0.26}$	$.9451_{\pm .0011}$	$24.99_{\pm 0.74}$	$.9648_{\pm .0021}$
scDiffEq	$42.38_{\pm 0.73}$	$.9756_{\pm .0038}$	$48.69_{\pm 2.58}$	$.9390_{\pm .0051}$	$62.73_{\pm 1.95}$	$.9477_{\pm .0019}$	$66.04_{\pm 1.87}$	$.9411_{\pm .0053}$	$20.45_{\pm 0.79}$	$.9693_{\pm .0015}$
scNODE	$28.89_{\pm 1.08}$	$.9856_{\pm .0018}$	$33.33_{\pm 0.98}$	$.9690_{\pm .0018}$	$56.96_{\pm 0.80}$	$.9539_{\pm .0041}$	$54.54_{\pm 2.25}$	$.9533_{\pm .0084}$	$21.60_{\pm 0.61}$	$.9662_{\pm .0023}$
PI-SDE	$26.94_{\pm 1.13}$	$.9901_{\pm .0018}$	$29.22_{\pm 1.21}$	$.9725_{\pm .0014}$	$53.46_{\pm 0.34}$	$\mathbf{.9725_{\pm .0018}}$	$58.34_{\pm 0.92}$	$.9581_{\pm .0022}$	$20.63_{\pm 0.29}$	$.9655_{\pm .0019}$
scIMF	$23.53_{\pm 0.16}$	$.9895_{\pm .0003}$	$28.76_{\pm 0.32}$	$.9726_{\pm .0004}$	$54.08_{\pm 0.21}$	$.9588_{\pm .0008}$	$56.08_{\pm 0.17}$	$.9506_{\pm .0011}$	$20.81_{\pm 0.19}$	$.9619_{\pm .0004}$
scMomenta	$\mathbf{23.24_{\pm 0.14}}$	$\mathbf{.9913_{\pm .0001}}$	$\mathbf{27.48_{\pm 0.24}}$	$\mathbf{.9744_{\pm .0006}}$	$\mathbf{50.51_{\pm 0.25}}$	$.9626_{\pm .0015}$	$\mathbf{51.43_{\pm 0.31}}$	$\mathbf{.9630_{\pm .0019}}$	$\mathbf{17.33_{\pm 0.36}}$	$\mathbf{.9734_{\pm .0009}}$
Extrapolation										
MIOFlow	$50.99_{\pm 1.71}$	$.9591_{\pm .0069}$	$99.24_{\pm 16.6}$	$.9061_{\pm .0372}$	$107.5_{\pm 13.4}$	$.8249_{\pm .0591}$	$87.68_{\pm 15.4}$	$.8669_{\pm .0542}$	$72.81_{\pm 4.17}$	$.8870_{\pm .0216}$
PRESCIENT	$59.94_{\pm 2.66}$	$.9303_{\pm .0034}$	$61.12_{\pm 2.08}$	$.9478_{\pm .0036}$	$89.38_{\pm 0.36}$	$.8562_{\pm .0045}$	$67.96_{\pm 1.82}$	$.9118_{\pm .0068}$	$53.94_{\pm 0.61}$	$.9064_{\pm .0021}$
scDiffEq	$74.09_{\pm 10.6}$	$.9462_{\pm .0085}$	$85.26_{\pm 7.09}$	$.9229_{\pm .0148}$	$107.5_{\pm 10.3}$	$.8741_{\pm .0147}$	$68.54_{\pm 7.19}$	$.9284_{\pm .0087}$	$43.09_{\pm 2.18}$	$.9212_{\pm .0074}$
scNODE	$48.68_{\pm 1.04}$	$.9565_{\pm .0025}$	$88.75_{\pm 22.5}$	$.9090_{\pm .0298}$	$102.6_{\pm 8.26}$	$.8491_{\pm .0327}$	$\mathbf{60.80_{\pm 1.99}}$	$\mathbf{.9342_{\pm .0060}}$	$34.81_{\pm 3.20}$	$.9388_{\pm .0156}$
PI-SDE	$57.54_{\pm 3.68}$	$.9685_{\pm .0030}$	$105.5_{\pm 8.48}$	$.9096_{\pm .0121}$	$89.63_{\pm 4.19}$	$.9034_{\pm .0154}$	$84.44_{\pm 4.28}$	$.9037_{\pm .0063}$	$32.18_{\pm 2.78}$	$.9485_{\pm .0112}$
scIMF	$36.19_{\pm 0.23}$	$.9665_{\pm .0009}$	$67.37_{\pm 4.76}$	$.9470_{\pm .0033}$	$\mathbf{72.85_{\pm 1.14}}$	$\mathbf{.9366_{\pm .0034}}$	$65.56_{\pm 1.35}$	$.9297_{\pm .0010}$	$33.34_{\pm 0.56}$	$.9412_{\pm .0016}$
scMomenta	$\mathbf{35.79_{\pm 0.63}}$	$\mathbf{.9768_{\pm .0033}}$	$\mathbf{40.05_{\pm 2.20}}$	$\mathbf{.9718_{\pm .0046}}$	$77.82_{\pm 1.54}$	$.9084_{\pm .0129}$	$64.21_{\pm 1.73}$	$.9332_{\pm .0068}$	$\mathbf{30.19_{\pm 1.16}}$	$\mathbf{.9495_{\pm .0053}}$

scDiffEq and PRESCIENT place part of the early populations away from real cells, scNODE leaves part of the main body uncovered, and MIOFlow covers only a narrow band of it. On SC, scMomenta likewise reproduces the ordering of the held-out time points (Appendix Figure 11).

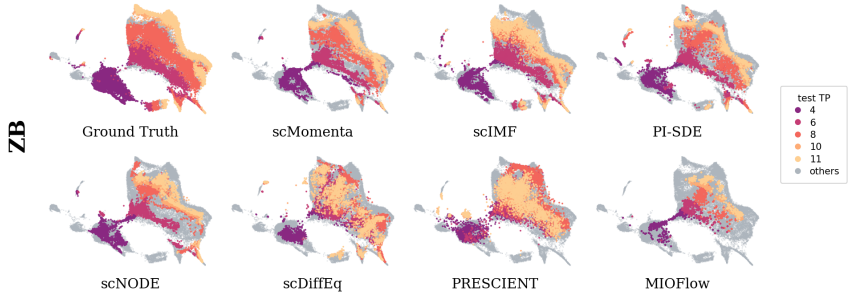

Figure 2: Held-out time points on ZB. Grey points show all real cells. Coloured points show the held-out time points, as ground truth and as generated by each method.

4.3 The Order Cost of Dimensionality Reduction

Sections 4.3 to 4.5 test the three observations of Section 3.1: the order cost of projection, trajectory crossings, and interventions on rate. We first quantify Proposition 1 in a controlled setting. Rather than claiming that the biological dynamics is first order, we fit a first-order linear field $\dot{x} = Ax$ to optimal-transport velocities in gene space, take it as the ground truth, and ask whether its projection stays closed at first order. Because the full-space dynamics is first order by construction, any higher order found after projection is caused by dimensionality reduction alone.

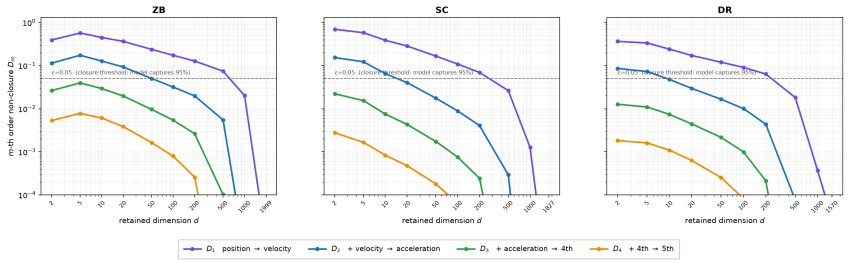

Figure 3: Closure error D_m against the retained dimension d, on logarithmic axes. The rightmost point of each panel is the no-reduction anchor, at which d equals the numerical rank of the data matrix.

With W the first d principal components, the derivatives $z^{(j)} = WA^j x$ of $z = Wx$ (Equation 3) are exact, requiring neither finite differences nor cell-to-cell matching. The closure error D_m is one minus the out-of-sample R^2 of a regression of the m-th derivative on the lower derivatives and time, and we regard the dynamics as closed at order m when $D_m < 0.05$. We sweep d from 2 to 1,000 on ZB, SC and DR, with a no-reduction anchor at the numerical rank of the data matrix (Appendix F.1).

At the no-reduction anchor, the rightmost point of each panel in Figure 3, D_1 falls below 10^{-4} on all three datasets, so the probe correctly reports first order when no reduction is applied. As d decreases, the required order rises, as Proposition 1 anticipates (Appendix Table 7). The order exceeds one for every $d \leq 200$, and at the latent dimension of scMomenta, $d = L = 50$, order two suffices on SC and DR, while D_2 of ZB (0.0502) lies just above the threshold.

Compressing a first-order system can thus make the retained coordinates non-Markovian at first order, and one additional order, carried by an explicit velocity state, largely restores closure. The question is therefore not whether cellular dynamics has one fixed physical order, but whether position alone is a sufficient state at the latent resolution used for prediction.

4.4 Trajectory Crossings under Second-Order Dynamics

A trajectory crossing directly tests whether position alone determines what a cell becomes, since a deterministic first-order model cannot represent two paths with different histories that pass through a common position (Proposition 2). On synthetic data with known pairing, scMomenta recovers the crossing (coupling accuracy 1.00), whereas the same model reduced to first order folds both branches back and swaps them (0.00; Appendix F.2). We then turn to CE, whose fully mapped lineage provides ground truth for which lineages cross.

The lineages of ABpxaapap and MSxpp have distinct ancestries and become ciliated amphid neurons and body wall muscle, respectively, yet their paths cross in a two-dimensional projection of the latent state (Figure 4A), swapping order on its fate axis at the second time point (Figure 4B). scMomenta reproduces the crossing, whereas the same model reduced to first order keeps the two lineages in their initial order throughout, consistent with the structural limitation of Proposition 2 rather than a failure of optimization. Appendix F.2 extends the comparison to all 2,092 root-to-leaf lineages of CE.

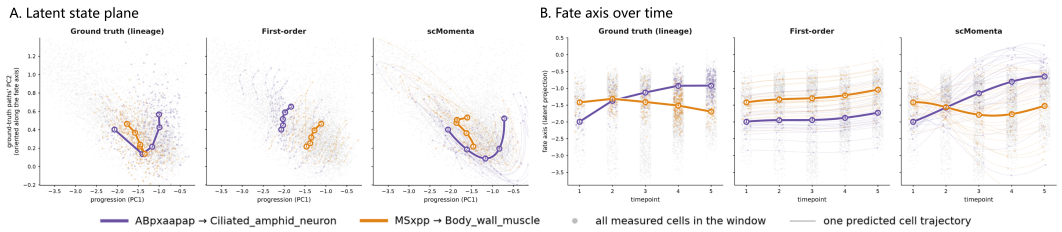

Figure 4: A crossing pair of CE lineages, in the latent state plane **(A)** and against time **(B)**. Thick purple and orange lines track each lineage at the population level, and thin lines are single cells.

4.5 Perturbations of Level and Rate

Perturbations act either on expression *level*, as exogenous mRNA does, or on its *rate*, as CRISPR interference and activation, RNA interference and knockout do; by Equation 2, the first moves a gene's position and the second its velocity. Existing in-silico perturbation protocols set the expression of a gene to a given z-score and re-simulate (Yeo et al., 2021), which represents only the first kind. The position–velocity state expresses both. A displacement Δz can be applied to the position, $z \leftarrow z + \Delta z$ (the z_0 channel, the prevailing protocol), or to the velocity, $v \leftarrow v + \gamma\,\Delta z$ (the v_0 channel, undefined without a velocity state); the factor γ makes a velocity impulse that decays under damping alone integrate to the same displacement Δz (Appendix F.3).

On VE, the fate shift measured in the z_0 channel largely reflects the displacement imposed: it grows almost linearly with $\|\Delta z\|$ (correlation 0.969; Figure 5, left). The pan-endocrine marker CHGA (Taupenot et al., 2003) illustrates the confound. It is expressed once a cell has become endocrine and is neither a transcription factor nor a determinant of fate, yet in the z_0 channel it shifts the endocrine fraction more than any other gene of the panel, including the fate determinants NEUROG3 (Gradwohl et al., 2000) and NKX6-1 (Schaffer et al., 2013). Its large expression variance gives it the largest imposed displacement, along the direction that separates endocrine from other cells (Appendix F.3). In the v_0 channel, CHGA no longer stands out, and the fate determinant NKX6-1 ranks first (Figure 5, right).

The v_0 channel imposes no displacement at the moment of intervention, so its points form a vertical band at zero displacement (Figure 5, left), yet it produces fate shifts of the same order as the z_0 channel: the added velocity integrates into position, so the eventual displacement is generated by the learned dynamics rather than imposed from outside. A first-order model has no velocity in its state, so the two channels collapse into one.

4.6 Ablation

We compare the full model, whose velocity encoder g_ψ is warm-started from OT, with two ablated variants: one **without OT initialization**, with a randomly initialized g_ψ, and one **reduced to first**

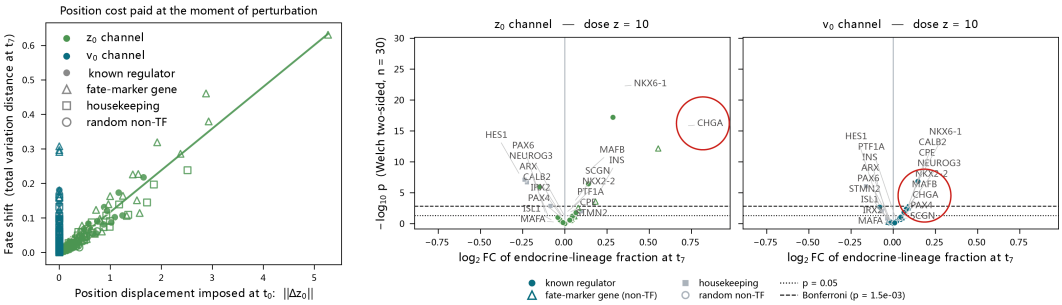

Figure 5: Perturbation through the two channels on VE. The red circle marks CHGA.

order, with no velocity in its state. No variant outperforms the one above it in any of the eight entries of Table 2. The largest gap is on SC extrapolation, the longest horizon, where reducing the model to first order raises W_2 by a third. Both components therefore contribute, and the gain from OT initialization is consistent with Proposition 4: it supplies coupling information that the marginal-matching loss alone does not determine.

Table 2: Ablation on ZB and SC. Values are mean ± standard deviation over five seeds.

Variant	ZB				SC			
	Interpolation		Extrapolation		Interpolation		Extrapolation	
	$W_2\downarrow$	Pearson↑	$W_2\downarrow$	Pearson↑	$W_2\downarrow$	Pearson↑	$W_2\downarrow$	Pearson↑
Full model	$\mathbf{23.24}_{\pm 0.14}$	$\mathbf{.9913}_{\pm .0001}$	$\mathbf{35.79}_{\pm 0.63}$	$\mathbf{.9768}_{\pm .0033}$	$\mathbf{27.48}_{\pm 0.24}$	$\mathbf{.9744}_{\pm .0006}$	$\mathbf{40.05}_{\pm 2.20}$	$\mathbf{.9718}_{\pm .0046}$
w/o OT initialization	$26.96_{\pm 0.23}$	$.9910_{\pm .0002}$	$39.27_{\pm 0.89}$	$.9631_{\pm .0038}$	$28.44_{\pm 0.20}$	$\mathbf{.9744}_{\pm .0005}$	$46.60_{\pm 3.17}$	$.9683_{\pm .0061}$
reduced to first order	$28.37_{\pm 0.28}$	$.9810_{\pm .0004}$	$42.55_{\pm 1.01}$	$.9597_{\pm .0025}$	$31.95_{\pm 0.30}$	$.9720_{\pm .0008}$	$53.32_{\pm 2.07}$	$.9660_{\pm .0027}$

5 Conclusion and Limitations

We present scMomenta, a second-order neural ODE framework for single-cell temporal prediction from unpaired snapshots. It augments the latent position with a velocity state, learns a damped acceleration field, and initializes the unobserved velocity from optimal transport between the first two snapshots. Across five datasets and six competing methods, it achieves the best prediction accuracy on most datasets in both interpolation and extrapolation, and ablations show that both the velocity state and the OT initialization contribute. Mechanistic experiments show that compressing first-order dynamics into the latent space breaks closure at first order, and one additional order largely restores it; that the velocity state represents trajectory crossings that a deterministic first-order model cannot; and that it lets perturbations act on the rate of expression as well as on its level. In downstream biological analyses, scMomenta also performs well, predicting clonal fates in mouse haematopoiesis and recovering the timing of zygotic genome activation in *Drosophila* (Appendix G). Position alone is thus not always a sufficient state for latent single-cell dynamics, and a velocity state with learned damping offers a practical way to represent temporal history.

The order analysis is restricted to linear dynamics and projections, and the OT initialization supplies a geometrically motivated rather than lineage-resolved coupling. The perturbation results of Section 4.5 remain in-silico predictions: scMomenta is trained on observational time courses only, so measured responses do not constrain its predicted effects. Large-scale single-cell CRISPR screens (Przybyla & Gilbert, 2022; Bock et al., 2022) now measure such responses, and integrating them with time courses is a natural direction for future work.

Reproducibility Statement

scMomenta is specified in Section 3: the latent space in Section 3.2, the second-order dynamics in Equations 4 and 5, the optimal-transport initialization of the velocity encoder in Equations 6 and 7, and the training objective and protocol in Section 3.4. Appendix C lists the architecture, the hyperparameters and the training schedule; one setting is used for all datasets and tasks. Each proposition states its assumptions, and complete proofs are given in Appendix B. All datasets are publicly avail-

able, and Appendix D gives their sources, accessions and preprocessing. Appendix E gives the time point splits for interpolation and extrapolation, the random seeds, and the definitions and parameters of both metrics. All methods are trained on identical splits and scored by one evaluation pipeline, and the competing methods were run from their official implementations with the recommended settings. The mechanistic and downstream analyses are specified in Appendices F.1 to G.2: the order probe of Section 4.3 through Equations 3 and 11, the synthetic crossing with known coupling, the perturbation protocol, and the fate and timing analyses. The anonymous repository in Appendix A contains the code of the model and the scripts of every experiment.

AI Use Statement

Generative AI tools were used for language editing of author-written text and for assistance with LaTeX formatting. They were not used to generate research ideas, to design experiments, to produce or analyze experimental results, or to conduct the literature search on which the related work is based. All citations were verified by the authors against the original publications. The authors have reviewed all AI-assisted text and take responsibility for the final content of this work.

References

V. Bergen, M. Lange, S. Peidli, F. A. Wolf, and F. J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12):1408–1414, 2020. doi: 10.1038/s41587-020-0591-3.

C. Bock, P. Datlinger, F. Chardon, M. A. Coelho, M. B. Dong, K. A. Lawson, T. Lu, L. Maroc, T. M. Norman, B. Song, et al. High-content CRISPR screening. *Nature Reviews Methods Primers*, 2 (1):9, 2022. doi: 10.1038/s43586-022-00098-7.

Diego Calderon, Ronnie Blecher-Gonen, Xingfan Huang, et al. The continuum of *Drosophila* embryonic development at single-cell resolution. *Science*, 377(6606):eabn5800, 2022. doi: 10.1126/science.abn5800.

J. A. Campos-Ortega and V. Hartenstein. *The Embryonic Development of Drosophila melanogaster*. Springer, Berlin, second edition, 1997. doi: 10.1007/978-3-662-22489-2.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Tianrong Chen, Guan-Horng Liu, Molei Tao, and Evangelos A. Theodorou. Deep momentum multimarginal Schrödinger bridge. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 57058–57086, 2023. doi: 10.52202/075280-2495.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2292–2300, 2013.

Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural ODEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Jeffrey A. Farrell, Yiqun Wang, Samantha J. Riesenfeld, Karthik Shekhar, Aviv Regev, and Alexander F. Schier. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, 360(6392):eaar3131, 2018. doi: 10.1126/science.aar3131.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and MMD using Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 2681–2690, 2019.

A. Filipczyk, C. Marr, S. Hastreiter, J. Feigelman, M. Schwarzfischer, P. S. Hoppe, D. Loeffler, K. D. Kokkaliaris, M. Endele, B. Schauberger, O. Hilsenbeck, S. Skylaki, J. Hasenauer, K. Anastassiadis, F. J. Theis, and T. Schroeder. Network plasticity of pluripotency transcription factors in embryonic stem cells. *Nature Cell Biology*, 17(10):1235–1246, 2015. doi: 10.1038/ncb3237.

Luz Garcia-Alonso, Valentina Lorenzi, Cecilia Icoresi Mazzeo, et al. Single-cell roadmap of human gonadal development. *Nature*, 607(7919):540–547, 2022. doi: 10.1038/s41586-022-04918-4.

Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1608–1617, 2018.

G. Gradwohl, A. Dierich, M. LeMeur, and F. Guillemot. *neurogenin3* is required for the development of the four endocrine cell lineages of the pancreas. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4):1607–1611, 2000. doi: 10.1073/pnas.97.4.1607.

M. M. Harrison, A. J. Marsh, and C. A. Rushlow. Setting the stage for development: The maternal-to-zygotic transition in *Drosophila*. *Genetics*, 225(2):iyad142, 2023. doi: 10.1093/genetics/iyad142.

P. S. Hoppe, M. Schwarzfischer, D. Loeffler, K. D. Kokkaliaris, O. Hilsenbeck, N. Moritz, M. Endele, A. Filipczyk, A. Gambardella, N. Ahmed, M. Etzrodt, D. L. Coutu, M. A. Rieger, C. Marr, M. K. Strasser, B. Schauberger, I. Burtscher, O. Ermakova, A. Bürger, H. Lickert, C. Nerlov, F. J. Theis, and T. Schroeder. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*, 535(7611):299–302, 2016. doi: 10.1038/nature18320.

Guillaume Huguet, Daniel Sumner Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. Manifold interpolating optimal-transport flows for trajectory inference. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 29705–29718, 2022. doi: 10.52202/068431-2154.

Qi Jiang and Lin Wan. A physics-informed neural SDE network for learning cellular dynamics from time-series scRNA-seq data. *Bioinformatics*, 40(Supplement_2):ii120–ii127, 2024. doi: 10.1093/bioinformatics/btae400.

Qi Jiang, Longquan Li, Lei Zhang, and Lin Wan. Learning collective multicellular dynamics with an interacting mean field neural SDE model. *PLOS Computational Biology*, 22(1):e1013916, 2026. doi: 10.1371/journal.pcbi.1013916.

Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations (ICLR)*, 2014.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. RNA velocity of single cells. *Nature*, 560(7719):494–498, 2018. doi: 10.1038/s41586-018-0414-6.

Zhuo Ma, Xiaofei Zhang, Wen Zhong, Hongyan Yi, Xiaowei Chen, Yinsuo Zhao, Yanlin Ma, Eli Song, and Tao Xu. Deciphering early human pancreas development at the single-cell level. *Nature Communications*, 14(1):5354, 2023. doi: 10.1038/s41467-023-40893-8.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Hazime Mori. Transport, collective motion, and Brownian motion. *Progress of Theoretical Physics*, 33(3):423–455, 1965. doi: 10.1143/PTP.33.423.

Alexander Norcliffe, Cristian Bodnar, Ben Day, Nikola Simidjievski, and Pietro Liò. On second order behaviour in augmented neural ODEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Adrien Osakwe, Eric H. Huang, and Yue Li. scTimeBench: A streamlined benchmarking platform for single-cell time-series analysis. *Bioinformatics*, 42(Supplement_2):btag414, 2026. doi: 10.1093/bioinformatics/btag414.

J. S. Packer, Q. Zhu, C. Huynh, P. Sivaramakrishnan, E. Preston, H. Dueck, D. Stefanik, K. Tan, C. Trapnell, J. Kim, R. H. Waterston, and J. I. Murray. A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science*, 365(6459):eaax1971, 2019. doi: 10.1126/science.aax1971.

L. Przybyla and L. A. Gilbert. A new era in functional genomics screens. *Nature Reviews Genetics*, 23(2):89–103, 2022. doi: 10.1038/s41576-021-00409-w.

T. Konstantin Rusch, Benjamin P. Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael M. Bronstein. Graph-coupled oscillator networks. In *International Conference on Machine Learning (ICML)*, 2022.

Michael E. Sander, Pierre Ablin, Mathieu Blondel, and Gabriel Peyré. Momentum residual neural networks. In *International Conference on Machine Learning (ICML)*, pp. 9276–9287, 2021.

A. E. Schaffer, B. L. Taylor, J. R. Benthuysen, J. Liu, F. Thorel, W. Yuan, Y. Jiao, K. H. Kaestner, P. L. Herrera, M. A. Magnuson, C. L. May, and M. Sander. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic beta cell identity. *PLoS Genetics*, 9 (1):e1003274, 2013. doi: 10.1371/journal.pgen.1003274.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. doi: 10.1016/j.cell.2019.01.006.

Yutong Sha, Yuchi Qiu, Peijie Zhou, and Qing Nie. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6(1):25–39, 2024. doi: 10.1038/s42256-023-00763-w.

J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology*, 100(1):64–119, 1983. doi: 10.1016/0012-1606(83)90201-4.

Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pp. 366–381. Springer, Berlin, 1981. doi: 10.1007/BFb0091924.

F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao, and M. A. Surani. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5):377–382, 2009. doi: 10.1038/nmeth.1315.

L. Taupenot, K. L. Harper, and D. T. O'Connor. The chromogranin–secretogranin family. *New England Journal of Medicine*, 348(12):1134–1149, 2003. doi: 10.1056/NEJMra021405.

Panagiotis Theodoropoulos, Augustinos Saravanos, Evangelos A. Theodorou, and Guan-Horng Liu. Momentum multi-marginal Schrödinger bridge matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 95572–95609, 2025. doi: 10.52202/085713-2877.

A. Veres, A. L. Faust, H. L. Bushnell, E. N. Engquist, J. H. Kenty, G. Harb, Y. C. Poh, E. Sintov, M. Gürtler, F. W. Pagliuca, Q. P. Peterson, and D. A. Melton. Charting cellular identity during human in vitro β-cell differentiation. *Nature*, 569(7756):368–373, 2019. doi: 10.1038/s41586-019-1168-5.

Cédric Villani. *Optimal Transport: Old and New*. Springer, Berlin, 2009.

Michael E. Vinyard, Anders W. Rasmussen, Ruitong Li, Allon M. Klein, Gad Getz, and Luca Pinello. Learning cell dynamics with neural differential equations. *Nature Machine Intelligence*, 7(12):1969–1984, 2025. doi: 10.1038/s42256-025-01150-3.

Dongyi Wang, Yuanwei Jiang, Zhenyi Zhang, Xiang Gu, Peijie Zhou, and Jian Sun. Joint velocity-growth flow matching for single-cell dynamics modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 178180–178216, 2025. doi: 10.52202/085713-5363.

C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, and A. M. Klein. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479):eaaw3381, 2020. doi: 10.1126/science.aaw3381.

Hedi Xia, Vai Suliafu, Hangjie Ji, Tan M. Nguyen, Andrea L. Bertozzi, Stanley J. Osher, and Bao Wang. Heavy ball neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Grace Hui Ting Yeo, Sachit D. Saksena, and David K. Gifford. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature Communications*, 12(1):3222, 2021. doi: 10.1038/s41467-021-23518-w.

Çağatay Yıldız, Markus Heinonen, and Harri Lähdesmäki. ODE^{2}VAE: Deep generative second order ODEs with Bayesian neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Jiaqi Zhang, Erica Larschan, Jeremy Bigness, and Ritambhara Singh. scNODE: Generative model for temporal single cell transcriptomic data prediction. *Bioinformatics*, 40(Supplement_2):ii146–ii154, 2024. doi: 10.1093/bioinformatics/btae393.

Zhenyi Zhang, Tiejun Li, and Peijie Zhou. Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. In *International Conference on Learning Representations (ICLR)*, 2025.

Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. Symplectic ODE-Net: Learning Hamiltonian dynamics with control. In *International Conference on Learning Representations (ICLR)*, 2020.

Robert Zwanzig. Nonlinear generalized Langevin equations. *Journal of Statistical Physics*, 9(3): 215–220, 1973. doi: 10.1007/BF01008729.

A Code Availability

The main code is available at
`https://anonymous.4open.science/r/scMomenta-code-anonymous`.

B Proofs

B.1 Proof of Proposition 1

With $\dot{x} = Ax$ and $z = Wx$, induction gives $z^{(j)} = WA^j x$, which is Equation 3. Closure at order m means that $z^{(m)}$ is a linear function of $z, \dots, z^{(m-1)}$ for every x. Equivalently, every row of WA^m lies in the row space of $O_m = [\, W; WA; \dots; WA^{m-1} \,]$. This is the stated criterion.

The row space $\mathrm{row}(O_m)$ is non-decreasing in m by construction. Suppose that it fails to grow at some step, so that $\mathrm{row}(WA^m) \subseteq \mathrm{row}(O_m)$. Then $\mathrm{row}(O_m)$ is A-invariant. Hence $\mathrm{row}(WA^{m+j}) \subseteq \mathrm{row}(O_m)$ for all $j \geq 0$, and the row space never grows again. Growth therefore stops at the first step at which it fails to increase, and this step is the minimal closure order m^*. By the Cayley–Hamilton theorem, A^G is a linear combination of $I, A, \dots, A^{G-1}$. Hence $\mathrm{row}(WA^G) \subseteq \mathrm{row}(O_G)$, and growth stops at or before $m = G$. Thus m^* exists and $m^* \leq G$.

For the lower bound, each additional order appends at most d rows to O_m, so $\mathrm{rank}(O_m) \leq md$. Once growth has stopped, $\mathrm{row}(O_{m^*}) = \mathrm{row}(O_G)$, which is the smallest A-invariant subspace containing $\mathrm{row}(W)$, that is, the part of the dynamics that the retained coordinates observe. Hence $r = \mathrm{rank}(O_G) = \mathrm{rank}(O_{m^*}) \leq m^* d$, and $m^* \geq \lceil r/d \rceil$.

Finally, $m^* = 1$ means $\mathrm{row}(WA) \subseteq \mathrm{row}(W)$, that is, $\mathrm{row}(W)$ is A-invariant. For fixed A with distinct eigenvalues, the d-dimensional invariant subspaces are the finitely many spans of d left eigenvectors of A, a finite and hence measure-zero subset of the Grassmannian of d-dimensional subspaces. A W chosen without reference to A, such as principal component analysis or a variational autoencoder trained on reconstruction, therefore satisfies $m^* = 1$ only exceptionally. The assumption on the eigenvalues is needed: for $A = \lambda I$, every subspace is invariant. □

B.2 Proof of Proposition 2

Let $z_a(\cdot)$ and $z_b(\cdot)$ solve $\dot{z} = f(z, t)$ with $z_a(t^*) = z_b(t^*)$. Take t^* as the initial time and the common value as the initial condition. The Picard–Lindelöf theorem gives a unique solution on a neighbourhood of t^*, so $z_a \equiv z_b$ there. The set on which the two solutions agree is therefore open. It is also closed by continuity, and it is non-empty. Since the interval is connected, this set is the whole interval, so $z_a \equiv z_b$. Injectivity of φ_t follows by applying this argument with $t^* = t$. □

B.3 Proof of Proposition 3

Write f for f_θ.

Existence and uniqueness of v^*. For fixed (z, t), the map $v \mapsto f(z, v, t)/\gamma$ has Lipschitz constant $\Lambda/\gamma < 1$. It is thus a contraction on $\mathbb{R}^L$. The Banach fixed point theorem gives a unique $v^*(z, t)$ with $\gamma v^* = f(z, v^*, t)$.

The bound. Let $w(t) = v(t) - v^*(z(t), t)$. Differentiating and using Equation 5 gives

$$\dot{w} = f(z, v, t) - f(z, v^*, t) - \gamma w - \frac{\mathrm{d}}{\mathrm{d}t} v^*\big(z(t), t\big).$$

Taking norms and using the Lipschitz property of f in v gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\|w\| \;\leq\; \Lambda\|w\| - \gamma\|w\| + \left\| \tfrac{\mathrm{d}}{\mathrm{d}t} v^* \right\| \;=\; -(\gamma - \Lambda)\|w\| + \left\| \tfrac{\mathrm{d}}{\mathrm{d}t} v^* \right\|.$$

Since $\gamma > \Lambda$, Grönwall's inequality yields

$$\|w(t)\| \;\leq\; e^{-(\gamma-\Lambda)(t-t_0)}\, \|w(t_0)\| \;+\; \frac{1}{\gamma - \Lambda} \sup_s \left\| \tfrac{\mathrm{d}}{\mathrm{d}s} v^*\big(z(s), s\big) \right\|. \tag{9}$$

□

Scaling of the limit. Taking $\gamma \to \infty$ at fixed v^* requires f to grow in proportion to γ. The learned f_θ absorbs this scale. Without it, $v^* = f/\gamma \to 0$, and the limit is a system at rest rather than first-order dynamics. Writing $f = \gamma F$ with F fixed, the Lipschitz constant scales as $\Lambda = \gamma \Lambda_F$, so the condition $\gamma > \Lambda$ becomes $\Lambda_F < 1$. The fixed point then solves $v^* = F(z, v^*, t)$ and does not depend on γ, and the residual in Equation 9 is $O\big(1/(\gamma(1 - \Lambda_F))\big)$, which vanishes as $\gamma \to \infty$.

B.4 Proof of Proposition 4

Let $\mathcal{L}$ be any loss that depends on the parameters only through the marginals, for instance $\mathcal{L}(\theta, \psi) = \sum_{k \in \mathcal{O}} S(\hat{\mu}_k^{\theta,\psi}, \mu_k)$, in which the parameters enter only the first argument of S through the marginal $\hat{\mu}_k^{\theta,\psi}$. Then $\mathcal{L}$ factors through the map $(\theta, \psi) \mapsto \{\hat{\mu}_k^{\theta,\psi}\}_{k \in \mathcal{O}}$. Any two parameter settings with the same image therefore incur the same loss, whatever coupling they induce. □

A sequence of marginals underdetermines the coupling. Unless one of two marginals is a point mass, the set of couplings between them contains more than one element, for example the independent coupling and the optimal transport coupling. Since this set is convex, it then contains infinitely many couplings, in any dimension.

C Implementation Details

Appendix Table 3 lists the hyperparameters of scMomenta. The initial velocity targets are computed on at most 500 anchor cells of the first observed time point, which bounds the size of the exact transport problem.

Table 3: Hyperparameters of scMomenta, shared by all datasets and tasks.

Module	Hyperparameter	Value
Autoencoder	latent dimension L	50
	decoder	single linear layer
	pretraining	500 iterations, learning rate 10^{-3}, batch 2048
Acceleration field f_θ	hidden layers	2 layers of width 128
	time embedding	sinusoidal, dimension 16
Damping	γ	$\gamma_{\max} \cdot \mathrm{sigmoid}(\omega)$ (Xia et al., 2021), with $\gamma_{\max} = 100$ and a single learnable scalar ω initialized at -3
Velocity encoder g_ψ	hidden width	128
	warm start	300 steps with a separate optimizer
ODE solver	method	`dopri5`, relative and absolute tolerance 10^{-5}
	gradients	backpropagation through the solver
Training	optimizer	AdamW with a cosine learning-rate schedule
	epochs	1000
	learning rate	10^{-2}
	batch size	256
	Sinkhorn divergence	blur 0.05, scaling 0.9
	velocity weight λ_v	10^{-3}
Prediction	simulated cells	2000

Because the decoder is a single linear layer with constant W_{dec}, and the bias vanishes under differentiation, latent velocity and acceleration map to gene space exactly,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v\, W_{\mathrm{dec}}^\top, \qquad \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = a\, W_{\mathrm{dec}}^\top, \qquad a = f_\theta(z, v, t) - \gamma v. \tag{10}$$

These relations are identities, not approximations. The gene-space rate in Appendix G.2 rests on them.

D Datasets and Preprocessing

Appendix Table 4 lists the five temporal prediction datasets, and Appendix Table 5 lists the three datasets used for mechanistic and downstream evaluation. All eight datasets follow a standard pipeline of quality control, library size normalization to 10^4 counts per cell and $\log(1 + x)$ transformation. We keep about $2{,}000$ genes, namely the top $2{,}000$ highly variable genes together with known marker genes of each process. DR is subsampled to a tenth of its cells, stratified by time point, and the counts in Appendix Table 4 are after subsampling. SC retains a single batch replicate, so that the time axis is not confounded with batch.

Table 4: The five time-course datasets used for temporal prediction.

Code	Organism	Process	Times	Cells	Source
ZB (Farrell et al., 2018)	Zebrafish	Embryogenesis	12	38,731	Broad SCP `SCP162`
DR (Calderon et al., 2022)	*Drosophila*	Embryogenesis	11	44,931	GEO `GSE190147`
SC (Schiebinger et al., 2019)	Mouse	Fibroblast→iPSC	19	62,840	GEO `GSE122662`
GA (Garcia-Alonso et al., 2022)	Human	Germ cell development	11	10,895	CELLxGENE
MA (Ma et al., 2023)	Human	Embryonic pancreas	8	19,224	NGDC `OMIX001616`

Table 5: The three datasets used for mechanism and downstream evaluation.

Code	Organism	Content	Times	Use in this work
CE (Packer et al., 2019)	*C. elegans*	Embryogenesis, full lineage	10	Crossing (Section 4.4)
MH (Weinreb et al., 2020)	Mouse	Haematopoiesis, LARRY clones	3	Fate (Appendix G.1)
VE (Veres et al., 2019)	Human	hPSC to pancreatic β cell	8	Perturbation (Section 4.5)

E Evaluation Protocol

E.1 Time Point Splits

Each dataset has two fixed splits, listed in Appendix Table 6, which are identical for every method and seed. The splits on ZB, DR and SC follow the benchmark configuration of Zhang et al. (2024), and those on GA and MA follow the same criterion. Entries are time point indices rather than real times, so consecutive indices do not imply equal spacing. This division into interpolation and extrapolation also matches a recent benchmarking platform for single-cell temporal analysis (Osakwe et al., 2026).

Table 6: Time point splits. Held-out time points are given by index.

Dataset	Times	Interpolation		Extrapolation	
		Held out	Observed	Held out	Observed
ZB	12	$\{4, 6, 8\}$	9	$\{10, 11\}$	10
DR	11	$\{4, 6, 8\}$	8	$\{8, 9, 10\}$	8
SC	19	$\{5, 7, 9, 11\}$	15	$\{15, 16, 17, 18\}$	15
GA	11	$\{4, 6, 8\}$	8	$\{8, 9, 10\}$	8
MA	8	$\{2, 6\}$	6	$\{7\}$	7

E.2 Seeds

Every temporal prediction experiment is repeated over the five seeds 42 to 46, and we report the mean and standard deviation. The perturbation screen and the damping sweep of Appendix F.3 use the seeds 42 to 44. A single seed fixes the autoencoder fit, the transport pairing and the stochasticity of the model.

E.3 METRICS

W_2 is the debiased Sinkhorn divergence (Feydy et al., 2019) between the predicted and real cells, computed with GeomLoss for $p = 2$, blur 0.05 and scaling 0.5. It is computed in a 50-dimensional principal component space that is fitted once on the real cells of each dataset and shared by all methods, tasks and seeds. For $p = 2$ the ground cost is $\frac{1}{2}\|x-y\|^2$, so the reported value approximates $\frac{1}{2}W_2^2$ rather than the distance W_2 itself; we keep the symbol W_2 for brevity. Pearson denotes the correlation, along the gene axis, between the per-gene mean expression of the real cells at a time point and that of the predicted cells. The two metrics measure different quantities. W_2 acts on the multivariate distribution of cells and is sensitive to the location and spread of the population and to absolute expression level. The correlation uses the full gene space of about $2{,}000$ genes but only per-gene means. As a correlation coefficient, it is insensitive to an overall rescaling of the expression profile.

F MECHANISTIC EXPERIMENTS

F.1 THE ORDER COST OF DIMENSIONALITY REDUCTION

The order probe. The probe uses the gene space of the benchmark, with at most 800 cells per time point. The snapshots are unpaired, so per-cell velocity cannot be read off directly. We solve the exact optimal transport problem between adjacent time points in gene space and convert the barycentric projection of each cell into a velocity by Equation 7. We then fit a linear field $\dot{x} = Ax$ to these velocities by ridge regression and *define* the fitted A as the ground-truth gene-space dynamics. Both sides of the regression are centred within each time point, and the penalty is chosen on held-out cells. The rows of W are the first d principal components of the data. For the retained coordinates $z = Wx$ and their derivatives $z^{(j)}$, the m-th order closure error of Section 4.3 is

$$D_m \;=\; 1 - R^2_{\mathrm{oos}}\Big(z^{(m)} \;\Big|\; z, \dot{z}, \ldots, z^{(m-1)},\ \mathrm{onehot}(t)\Big). \tag{11}$$

Each derivative is scaled by its root mean square, so D_m does not depend on the overall scale of A. The out-of-sample R^2 is computed by five-fold cross-validation stratified by time point, and the ridge penalty is chosen by generalized cross-validation within each training fold.

Detailed closure errors. Appendix Table 7 gives D_m for every retained dimension and order. The last row is the no-reduction anchor, where d equals the numerical rank of the data matrix, $1{,}999$ on ZB, $1{,}827$ on SC and $1{,}570$ on DR. D_4 is not computed there, because the regression would have more than $6{,}200$ features. In each D_1 column, the closure error decreases as d grows, with one exception at the smallest dimension on ZB, and it falls below the threshold only at $d = 500$ for SC and DR and at $d = 1{,}000$ for ZB. D_2 falls below the threshold at $d = 10$ for DR, $d = 20$ for SC and $d = 100$ for ZB; at $d = 50$, D_2 of ZB is 0.0502, just above the threshold. D_3 is below it at every dimension.

Table 7: Closure error D_m (Equation 11). **Bold** values exceed the threshold 0.05.

Dimension d	ZB				SC				DR			
	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4
2	**0.3929**	**0.1139**	0.0264	0.0053	**0.6938**	**0.1535**	0.0222	0.0028	**0.3651**	**0.0861**	0.0127	0.0018
5	**0.5700**	**0.1737**	0.0396	0.0078	**0.5794**	**0.1226**	0.0154	0.0017	**0.3340**	**0.0727**	0.0110	0.0016
10	**0.4490**	**0.1274**	0.0294	0.0061	**0.3869**	**0.0652**	0.0075	0.0008	**0.2403**	0.0478	0.0074	0.0011
20	**0.3648**	**0.0930**	0.0197	0.0039	**0.2844**	0.0402	0.0043	0.0005	**0.1718**	0.0296	0.0044	0.0006
50	**0.2378**	**0.0502**	0.0096	0.0016	**0.1661**	0.0175	0.0017	0.0002	**0.1195**	0.0165	0.0022	0.0003
100	**0.1735**	0.0318	0.0054	0.0008	**0.1084**	0.0089	0.0008	<0.0001	**0.0911**	0.0100	0.0010	<0.0001
200	**0.1274**	0.0199	0.0026	0.0003	**0.0693**	0.0041	0.0002	<0.0001	**0.0637**	0.0044	0.0002	<0.0001
500	**0.0744**	0.0055	0.0001	<0.0001	0.0263	0.0003	<0.0001	<0.0001	0.0182	<0.0001	<0.0001	<0.0001
1000	0.0201	<0.0001	<0.0001	<0.0001	0.0013	<0.0001	<0.0001	<0.0001	0.0004	<0.0001	<0.0001	<0.0001
Numerical rank	<0.0001	<0.0001	<0.0001	—	<0.0001	<0.0001	<0.0001	—	<0.0001	<0.0001	<0.0001	—

F.2 Trajectory Crossings under Second-Order Dynamics

Synthetic data with known coupling. Two groups of cells travel along intersecting straight lines in the plane and cross at the midpoint, so the terminal branch of each cell is fixed by construction. Coupling accuracy is the fraction of cells whose predicted terminal branch matches the constructed one, and random guessing gives 0.5. As shown in Appendix Figure 6, scMomenta recovers the X of the ground truth with an accuracy of 1.00. The model reduced to first order produces a ")(" shape in which both branches fold back before the crossing point, with an accuracy of 0.00. This failure reflects a systematic inversion rather than a lack of learning: by Proposition 2, an injective flow cannot let the branches pass through one another, so it can match both end marginals only by folding each branch back, which swaps the endpoints.

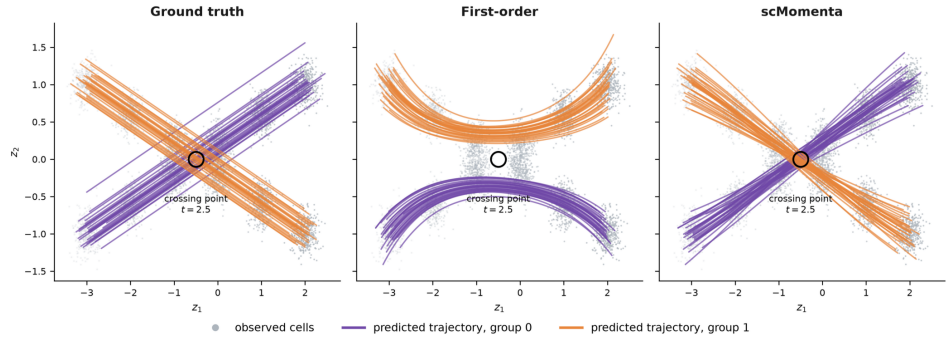

Figure 6: Trajectory crossing on synthetic data with known coupling.

The lineage forest of CE. We extend the crossing analysis from a single pair to the $2{,}092$ root-to-leaf lineage chains of CE and colour each predicted trajectory by its terminal fate, adding scNODE and PI-SDE for comparison. As shown in Appendix Figure 7, the circled region contains two lineages in the ground truth, ciliated amphid neurons in purple and ciliated non-amphid neurons in orange. scMomenta returns both. scNODE retains only a short orange segment, and PI-SDE assigns every trajectory in this region to a single fate.

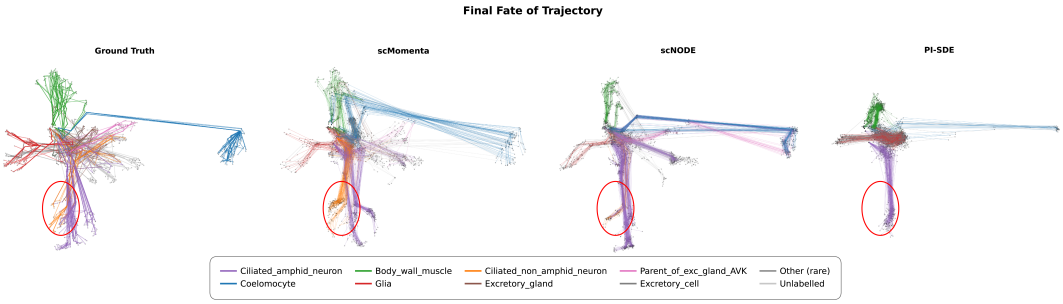

Figure 7: Predicted terminal fates over the lineage forest of CE. Panels show, from left to right, the ground truth, scMomenta, scNODE and PI-SDE. The red circle in each panel marks a region where two lineages coexist in the ground truth.

F.3 Perturbations of Level and Rate

Interventions on level and on rate. Interventions on expression level, such as delivery of exogenous mRNA, change abundance directly. Interventions on rate, such as CRISPR interference and activation, induced overexpression, RNA interference or genomic knockout, change the rate of change of abundance rather than abundance itself. Equation 2 locates the distinction. At fixed s, $\dot{s}$ can change only through the unspliced pool u or the rates β and γ_{deg}: transcriptional regulation, as in CRISPR interference and activation, acts on u through α, and RNA interference raises γ_{deg}

by accelerating the degradation of the mature transcript. A genomic knockout therefore does not set expression to zero at once. It abolishes the production of functional transcripts, and abundance decays over several mRNA half-lives.

Perturbation protocol. The panel on VE contains 34 genes, namely known regulators, fate-marker genes, housekeeping genes and random genes that are not transcription factors, each perturbed at seven doses, -1, 0, 1, 2, 3, 5 and 10. A dose is the z-score to which the expression of the target gene is set, and Δz is the change that this edit makes to the latent code of a cell. The z_0 channel follows the prevailing protocol and restarts the simulation from $z + \Delta z$. The v_0 channel keeps the position and adds $\gamma\,\Delta z$ to the velocity, where γ is the damping learned by the checkpoint. Under the damping term alone, this impulse decays as $\gamma\,\Delta z\,e^{-\gamma t}$ and integrates to the displacement Δz, which calibrates the two channels to the same total displacement. The perturbed cells are simulated to the last time point, decoded to gene space and assigned to cell types by a classifier trained on real cells, whose accuracy on held-out real cells is about 0.90. The fate shift of an intervention is the total variation distance between the cell-type compositions of the perturbed and unperturbed populations. The right panels of Figure 5 report, at the largest dose, 10, the $\log_2$ fold change of the endocrine-lineage fraction and its two-sided Welch test, with $n = 30$ values per condition: three seeds, each contributing ten resamples of 200 simulated trajectories.

The displacement of `CHGA`. `CHGA` is highly expressed in endocrine cells and near zero elsewhere, so its standard deviation across cells is large. The prevailing protocol sets expression in units of this standard deviation, so the displacement of `CHGA` is the largest on the panel. This displacement falls along the direction that the read-out measures (Section 4.5), which explains the lead of `CHGA` in the z_0 channel.

Separability and damping. The two channels differ only in the time course over which the intervention is delivered. This difference acts as a low-pass filter with timescale $\tau = 1/\gamma$, so the two channels approach each other as γ grows. Let e^{z_0} and e^{v_0} be the effect vectors of the two channels for a given gene and dose. We define separability as $1 - \cos(e^{v_0}, e^{z_0})$, which vanishes when the two channels report the same measurement. We test this prediction on VE and ZB by retraining at five fixed values of γ, with three seeds each. As shown in Appendix Figure 8, separability on VE decreases monotonically with γ, and in both gene space and latent space it spans more than two orders of magnitude over the scanned range. Every individual curve is monotone (Fisher combined $p = 3.4 \times 10^{-13}$).

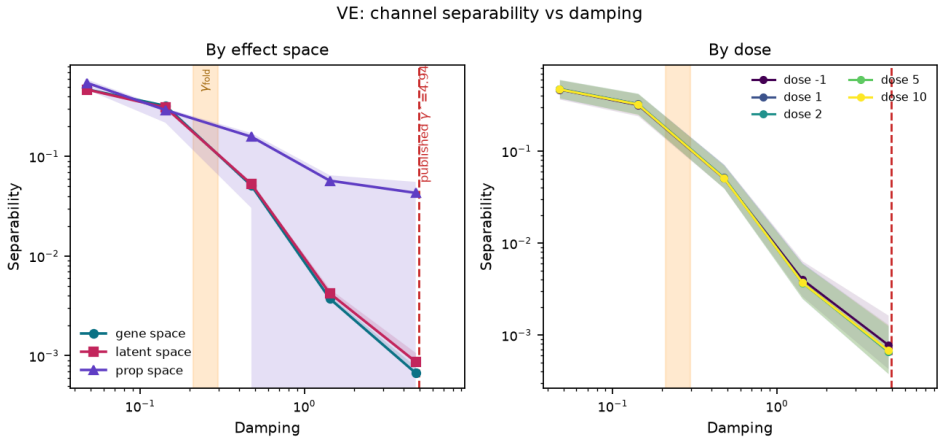

Figure 8: Separability of the two perturbation channels against the damping γ fixed during training, on VE. Curves are grouped by effect space (left) and by dose (right).

G Downstream Biological Analyses

G.1 Cell Fate Prediction

Section 1 notes that progenitors with highly similar transcriptional states can give rise to different descendants. We test whether scMomenta predicts these fates on MH (Weinreb et al., 2020). Its LARRY barcodes label clones before the first sampling, so the cell types of a clone's members on later days serve as the ground-truth fate of its day-two cells. scMomenta, scNODE, PRESCIENT and scDiffEq evolve the same day-two cells to day six. The terminal states are projected into the global 50-dimensional principal component space of the dataset, and a fate posterior is read out from the $k = 20$ nearest real cells. The read-out space, the neighbourhood size and the reference pool are the same for all methods. We report the accuracy of the dominant fate, the macro-averaged AUROC, the AUROC between the neutrophil and monocyte branches, the Pearson correlation of clonal bias and the negative cross-entropy. A random baseline and a static ridge regression serve as reference lines. The ridge regression is trained on the clonal fates themselves, with five-fold cross-validation grouped by clone, so it is an upper reference rather than a competing method.

As shown in Appendix Table 8, scMomenta is best on all five metrics among the four dynamical methods, and scNODE is second on all metrics except macro-averaged AUROC. The gap is largest on negative cross-entropy, where the other three methods have similar values far below that of scMomenta. Their fate posteriors are therefore worse in calibration as well as in ranking. Appendix Figure 9 shows the trajectories of scMomenta grouped by the observed fate of each clone, together with the barcoded cells of that fate on days two, four and six.

Table 8: Clonal fate prediction on MH. **Best** and second best are marked among the four dynamical methods.

Method	Pearson r ↑	AUROC↑ neut./mono.	Macro AUROC↑	Accuracy (%)↑	Neg. cross entropy↑
Random baseline (reference)	0.000	0.500	0.500	$28.2_{\pm 1.5}$	-1.61
Static ridge regression (upper reference)	0.557	0.826	0.894	65.9	-1.10
scDiffEq	$0.202_{\pm .023}$	$0.576_{\pm .008}$	$0.627_{\pm .020}$	$26.4_{\pm 1.1}$	-14.07
PRESCIENT	$0.290_{\pm .005}$	$0.619_{\pm .001}$	$0.572_{\pm .005}$	$28.6_{\pm 0.4}$	-14.86
scNODE	$0.319_{\pm .021}$	$0.634_{\pm .012}$	$0.617_{\pm .009}$	$42.3_{\pm 1.4}$	-13.98
scMomenta	$\mathbf{0.385_{\pm .017}}$	$\mathbf{0.660_{\pm .010}}$	$\mathbf{0.629_{\pm .014}}$	$\mathbf{44.2_{\pm 2.0}}$	$\mathbf{-10.47}$

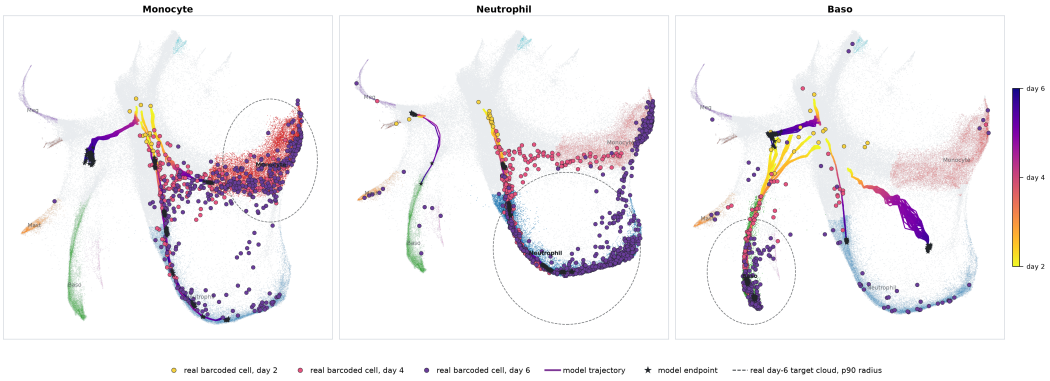

Figure 9: scMomenta trajectories on MH, grouped by the observed fate of each clone.

G.2 Timing of Zygotic Genome Activation

We ask whether the internal velocity field of the model agrees with an independently measured fact of developmental biology. The major wave of zygotic genome activation in the *Drosophila*

embryo occurs at nuclear cycle fourteen. The literature places it at 2.15 to 3.0 hours after egg laying (AEL) (Harrison et al., 2023; Campos-Ortega & Hartenstein, 1997), and the study that produced DR places it at about two hours and twenty minutes (Calderon et al., 2022). Fertilization immediately precedes egg laying in *Drosophila*, so we give all times in hours AEL, the clock of the DR collection windows. We read velocity out on real cells as the quasi-static velocity of Proposition 3, solving $\gamma v^* = f_\theta(z, v^*, t)$ by damped fixed-point iteration (at most 500 iterations, damping 0.3) and keeping cells whose residual is below 10^{-3}. Proposition 3 guarantees a unique solution when $\gamma > \Lambda$; the residual check verifies convergence cell by cell instead of assuming this condition. We take the median of $\|v^*\|$ per hour at each time point as the rate of transcriptomic change. This timing did not enter training in any form.

As shown in the right panel of Appendix Figure 10, the rate rises by a factor of 6.9 from the first time point to the second, at 1 and 2 hours, and then falls. Its peak lies at the second or third time point in all ten runs under both splits, and on the continuous time grid of the model it lies at 2.20 to 2.30 hours. The gene-space rate obtained through Equation 10 peaks at the same time point. A model-free control (left panel), the distance per hour between the centroids of adjacent time points in gene space, with each gene scaled to unit standard deviation, peaks in the same window. The model reduced to first order has no peak there, and its rate is highest at the end of the time course.

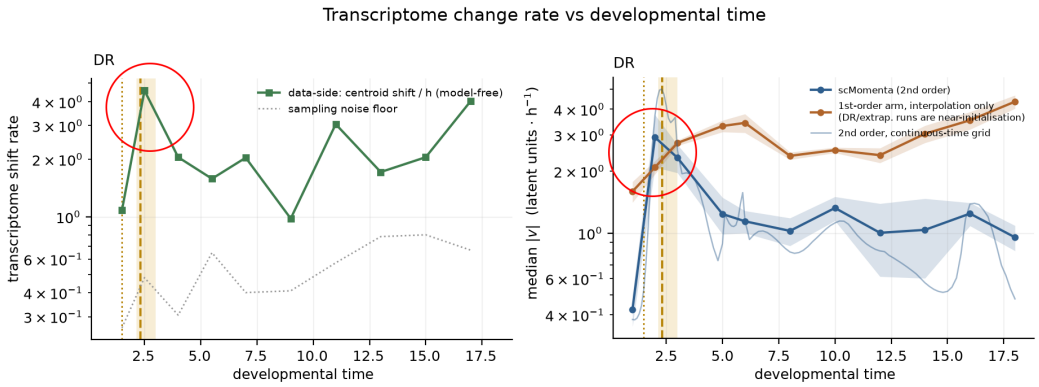

Figure 10: Rate of transcriptomic change on DR, from the data alone (left) and from the model (right). The gold band is the literature window of zygotic genome activation, and the dashed and dotted lines mark its major and minor waves. Red circles mark the peak in each panel. Time is given in hours after egg laying, at the midpoint of each collection window.

H Additional Results on Temporal Prediction

Appendix Figure 11 is the counterpart of Figure 2 on SC, which has 19 time points and is the longest time course in the benchmark. scMomenta reproduces the order of the held-out time points along the main body and the islands of the last time points at the upper right.

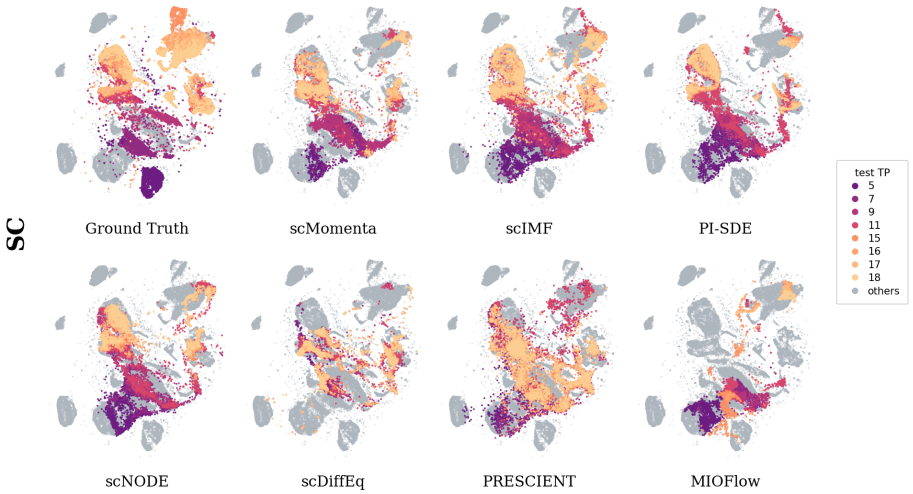

Figure 11: Held-out time points on SC. Conventions follow Figure 2.