Learning Stochastic Agent Dynamics from Pooled Population Snapshots

Wenjun Zhao[1]

[1]Department of Mathematics, Wake Forest University, Winston-Salem, NC, USA

October 9, 2026

Abstract

Many biological and ecological systems are observed through population snapshots rather than tracked trajectories, yielding a pooled mixture of snapshot distributions with no per-observation time label. For linear Gaussian stochastic differential equations (SDEs), we show that pooled moments determine the parameters locally for almost every parameter value, and for general snapshot families we give conditions under which the pooled law identifies the dynamics globally. Building on this, we propose a population-matching algorithm that bypasses time recovery entirely, fitting dynamics by minimizing a discrepancy between the observed pooled distribution and the pooled distribution simulated under the candidate parameters. On linear SDEs, we compare against a Gaussian hidden-assignment baseline that alternates between soft time-label recovery and parameter fitting via expectation–maximization (EM); direct population matching is more stable, confirming recovery in identifiable regimes and the predicted failure under temporal aliasing. We extend the direct method to interacting particle systems, where snapshot densities lack closed form, and to real single-cell gene-expression data, recovering regulatory rankings without per-cell stage labels. In both, it attains accuracy better than or at least competitive with baselines using temporal information, suggesting that time may not need to be an inferential target for learning dynamics.

1 Introduction

Agent-based models describe collective biological and ecological behavior through local rules acting on individual cells, particles, animals, or organisms [10]. These local rules determine the drift, diffusion, and interaction terms of the population flow, and estimating them is often the scientific goal. In single-cell biology, the learned flow can reveal regulatory interactions that drive transitions between cell states [32]. In animal movement and ecological systems, it can quantify migration tendencies, attraction or avoidance, and responses to environmental cues [13]. More broadly, a learned dynamical model supports prediction, perturbation analysis, and downstream decision making.

However, the observations available in these systems often do not take the form of complete trajectories. Temporal information can be missing in two distinct ways. First, individual trajectories may be unavailable: one observes each cell, particle, or animal only once, or loses identity across frames, so that observations cannot be matched into complete histories. Second, the sampling time itself may be unavailable or unreliable: even if one observes a collection of agents, one may not know where each observation lies along the relevant biological or behavioral clock. These two forms of missing temporal information often occur together. In destructive assays, such as many single-cell measurements, one observes a population of asynchronous cells distributed across different stages of an unobserved biological process, with each cell measured only once. In such data, the relevant time is each cell's intrinsic progression along the process; the observed collection time need not specify it and should be regarded

as a surrogate for that progression [34]. In ecological surveys based on indirect traces, such as footprints, scat, or other field evidence, the observation may record that an animal visited a location before the survey, but the precise visit time and the identity of the animal may be unknown. These settings motivate inference at the level of population snapshot laws pooled across time, where methods that require known time-indexed marginals or well-matched time series across frames are not directly applicable.

In this work, we focus on the case in which both individual trajectories and per-observation time labels are unavailable. We model the local stochastic rules by stochastic differential equations, so that a parameter θ determines a family of time-indexed population laws representing the distribution of agents at different stages of the process. Such systems serve as tractable continuous surrogates for agent-based models [22], and we restrict to this setting to enable theoretical analysis. The data are pooled samples from this evolving family of laws, but the time label of each sample is unobserved or unreliable. The goal is to recover the governing parameter θ from the aggregated population, rather than to assign a "pseudotime" to each observation, a standard approach adopted by modern computational biology practice [29]. Thus the hidden temporal assignment is a nuisance variable. This perspective is also motivated by gene regulatory network (GRN) inference, which seeks to determine, from gene-expression data, which genes influence the rate of change of which others. These regulatory interactions govern how expression evolves along each cell's unobserved intrinsic time. Benchmarking studies have highlighted limitations of reconstructed pseudotime in capturing this temporal progression [24], and found that methods not requiring pseudotime ordering were generally more accurate [26]. We therefore consider the central question: *when and how can one learn dynamics without first learning the per-observation time labels?*

Related work We review existing work on learning dynamical system parameters according to the type of temporal information assumed.

Trajectory methods. There is an extensive literature on methods based on trajectory data, requiring each individual to be observed multiple times at known time points. To name a few, these include neural ordinary differential equations (ODEs) [6], sparse identification via SINDy [5], and estimation of interaction kernels from particle trajectories [21], the last of which additionally requires velocity or other observables beyond position. In contrast, the pooled-snapshot setting removes both requirements: individuals are observed only once and their observation times are unknown.

Temporal-snapshot methods. A second line of work assumes temporally ordered population snapshots. APPEX [12] and nn-APPEX [11] estimate drift and diffusion from temporal marginals with identifiability guarantees. A parallel literature in computational biology, such as SINCERITIES [25] and OTVelo [40], infers gene regulatory networks from time-stamped single-cell expression profiles. These methods do not require complete individual trajectories, yet observations must still be grouped into temporally ordered populations at known times. An illustration of such a regime can be found in the top left panel of Figure 1.

Static population methods. A third line learns dynamics from long-time invariant or occupation measures [4, 36, 39], requiring statistical equilibrium. StationaryOT [38] recovers a flow field from steady-state data given known source and sink terms. Neither class addresses the transient pooled-snapshot setting where the population law is still evolving and no equilibrium structure is assumed.

Non-static and missing-time methods. A highly relevant and emerging line of work targets the challenging case where both trajectories and per-sample observational times are unknown. Zeng et al. [37] reconstruct deterministic ODEs from time-label-free data, treating the missing time parametrization as a reconstruction target. Wei and Lu [33] learn potentials of interacting particle systems from unlabeled snapshots via a trajectory-free loss based on the weak-form stochastic evolution equation. Our setting differs in that the hidden time assignment is treated as a nuisance variable rather than a reconstruction target, and we additionally provide identifiability guarantees for the pooled inference problem under suitable conditions.

Contributions This paper studies inference of dynamics from pooled population snapshots when individual observation times are hidden. Our central message: under conditions of identifiability, recovering each sample's latent time may not always be necessary, since dynamics can be learned by matching population laws directly.

(i) We formulate pooled-snapshot inference as an inverse problem and propose a population-matching estimator based on characteristic-kernel maximum mean discrepancy (MMD) [9], which recovers the dynamics whenever the pooled forward map is identifiable.

(ii) We prove that for linear Gaussian stochastic differential equations (SDEs), pooled moments of bounded degree determine the drift and diffusion locally for almost every parameter, without recovering time labels (Theorem 2.5). For general snapshot families, we give a conditional global result: mixture identifiability, nonrecurrence of the snapshot curve, and trajectory identifiability together recover the dynamics from the pooled law (Proposition 2.6).

(iii) Numerically, on linear systems we compare against a Gaussian hidden-assignment expectation–maximization (EM) baseline and find direct matching more stable, particularly as snapshots overlap or approach recurrence. On interacting particles and real single-cell gene-expression data, the method attains competitive accuracy without per-observation time labels.

2 Pooled snapshot laws and parameter identifiability

This section defines the observational model considered in this manuscript, and analyzes when a pooled law determines the underlying dynamics. Figure 1 summarizes our setup and the logical structure of this section. We begin by introducing an observational model under which time labels are hiddenin Section 2.1. We show that for linear Gaussian SDEs, the pooled law gives local identifiability in Section 2.2, and give a conditional global result for general families in Section 2.3. Finally, Section 2.4 discusses the practical implications for parameter inference given finite samples.

2.1 Observation model

Let $X_i \in \mathbb{R}^d$ denote the observed state of one agent, recorded at a sampling time $T_i \in [0, t_{\max}]$, and let ρ_t^θ denote the distribution of the state at time t under model parameters θ, a probability distribution on $\mathbb{R}^d$. We work under the following structural assumptions.

Assumption 2.1 (Observation model)**.** *The following quantities are known.*

(a) Initial law. *The initial population law $\rho_0 = \rho_0^\theta$ is known and independent of the unknown parameter θ.*

(b) Time marginal. *The sampling times T_i have a known marginal distribution ν on $[0, t_{\max}]$. In the continuous idealization, we focus on a uniform distribution over the time interval $[0, t_{\max}]$:*

$$T \sim U(0, t_{\max}).$$

The realized time label of each individual observation is not observed.

(c) Empirical approximation. *The empirical measure*

$$\widehat{\rho}_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}$$

approximates the data-generating pooled law $P_* = G(\theta_0) = \int \rho_t^{\theta_0} \, \nu(dt)$ in the sense that $\widehat{\rho}_N \xrightarrow{w} P_*$ as $N \to \infty$, where ν is the known sampling-time marginal, and θ_0 is the unknown true parameter, for example when the observations are independent draws from P_*.

Here δ_x denotes a unit point mass at x, and weak convergence means convergence of integrals against bounded continuous test functions. Under this model, conditional on its latent sampling time $T_i = t$, an observation is drawn from the model-implied population law ρ_t^θ. Thus the continuous pooled law is

$$G(\theta) = \int_0^{t_{\max}} \rho_t^\theta \nu(dt), \quad \text{which is } \frac{1}{t_{\max}} \int_0^{t_{\max}} \rho_t^\theta \, dt \text{ when } T \sim U(0, t_{\max}). \tag{2.1}$$

The observed law is a mixture of the model-implied snapshot laws, with no time label attached to any sample. The data consist of N pooled observations $X_1, \ldots, X_N$ from the population at unknown times. The latent times are nuisance variables; the inferential target is θ_0, not the assignment of observations to times.

The question is whether:

$$G(\theta) = G(\theta') \implies \theta = \theta'. \tag{2.2}$$

The parameters are *globally* identifiable on a parameter set if equation (2.2) holds for all pairs in the set of possible parameter values, and *locally* identifiable at θ if it holds for all θ' in some neighborhood of θ.

Remark 2.2 (Known initial population distribution)**.** *The initial population distribution is treated as known in the identifiability results as stated in Assumption 2.1(a). In the global argument of Section 2.3, it orients the curve of snapshot distributions (Remark 2.8). In practice, the initial distribution can often be characterized without per-observation time labels: in single-cell settings, progenitor cell types at the start of a differentiation process can be identified by known marker genes, while in ecological movement studies, the initial spatial distribution of animals can be estimated from habitat surveys prior to the observation period. Such estimates are subject to sampling error and uncertainty in population selection. We examine sensitivity to the supplied initial distribution through numerical experiments in Appendix B.*

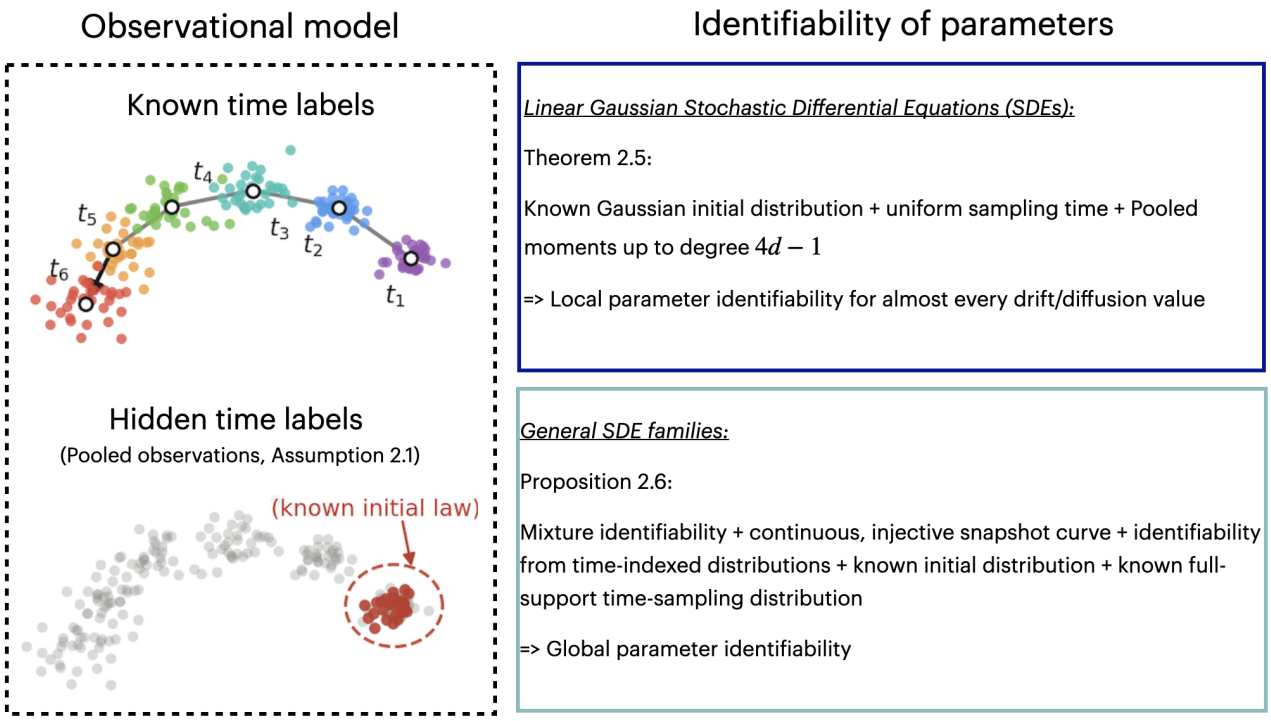


Figure 1: *Summary of the observational model considered in this work (left) and the theoretical contribution to the parameter identifiability (right).*

Remark 2.3 (Sampling-time distribution)**.** *The assumption of uniform time marginal, as stated in Assumption 2.1(b), can also be understood as a choice of temporal coordinate. If the sampling-time distribution has a continuous, strictly positive density, the transformed time $u = F_\nu(t)$ is uniformly distributed on $[0,1]$, where F_ν is its cumulative distribution function. Expressing the dynamics in this coordinate leaves the pooled distribution unchanged. Thus, at the level of general SDEs, a nonuniform sampling distribution can be absorbed into the temporal parameterization of the dynamics. The transformed drift and diffusion may, however, depend explicitly on time, so the uniform-sampling assumption remains a restriction for the time-homogeneous linear family considered in our main theorem.*

2.2 Local identifiability for linear Gaussian dynamics

For an analytically tractable case study, we now specialize to the linear Gaussian SDE of d dimensions ($d \geq 1$). Under

$$dX_t = (AX_t + b)\,dt + \Sigma\,dW_t, \qquad D = \Sigma\Sigma^\top, \tag{2.3}$$

with Gaussian initial condition $X_0 \sim \mathcal{N}(m_0, C_0)$, $C_0 \succ 0$, each snapshot law is Gaussian, $\rho_t^\theta = \mathcal{N}(m_\theta(t), C_\theta(t))$, with moments satisfying

$$\dot{m}_t = Am_t + b, \qquad \dot{C}_t = AC_t + C_tA^\top + D. \tag{2.4}$$

The parameter of interest is $\theta = (A, b, D)$. Here W_t is standard Brownian motion, and D is the constant diffusion covariance matrix. The notation $C_0 \succ 0$ means that C_0 is symmetric positive definite. Different matrices Σ can give the same $D = \Sigma\Sigma^\top$, so we identify D rather than Σ.

Our argument uses the principle behind the method of moments: different parameter values can be distinguished through the moments of the distribution they generate. Here, we use moments of the pooled observations, which do not require individual time labels. Although each snapshot distribution is Gaussian, their pooled mixture generally is not, so moments beyond the mean and covariance can carry additional information about the dynamics. For example, in one dimension, write $m_\theta(t)$ and $v_\theta(t)$ for the mean and variance at time t, generated by parameter θ. The first three moments of the pooled observation are:

$$\begin{aligned}
\mathbb{E}_\theta[X_T] &= \int m_\theta(t)\,\nu(dt), \\
\mathbb{E}_\theta[X_T^2] &= \int \big(m_\theta(t)^2 + v_\theta(t)\big)\,\nu(dt), \\
\mathbb{E}_\theta[X_T^3] &= \int \big(m_\theta(t)^3 + 3m_\theta(t)v_\theta(t)\big)\,\nu(dt).
\end{aligned} \tag{2.5}$$

All three are properties of the pooled law and require no individual time assignments. In higher dimensions, mixed moments similarly contain information about relationship between coordinates. Their total degree is the sum of the exponents; for example, $\mathbb{E}[X_1^2 X_2]$ is a moment of total degree three, with subscripts here denoting state coordinates. The following regularity fact ensures that the moments used below are well defined.

Lemma 2.4 (Regularity of the moment curve)**.** *Under Assumption 2.1(a) and model (2.3), $t \mapsto (m_\theta(t), C_\theta(t))$ is real-analytic on $[0,\infty)$, and on $[0, t_{\max}]$:*

(a) $\sup_t \|m_\theta(t)\| < \infty$ *and* $\sup_t \|(\dot{m}_\theta(t), \dot{C}_\theta(t))\| < \infty$*;*

(b) $C_\theta(t) \succ 0$ *for every* $t \geq 0$*, and there exist* $0 < s_1 \leq s_2 < \infty$ *with* $s_1 I \preceq C_\theta(t) \preceq s_2 I$ *for all* $t \in [0, t_{\max}]$*. Here* $B \preceq C$ *means that* $C - B$ *is positive semidefinite, and* I *is the identity matrix.*

Every pooled polynomial moment is finite and depends real analytically on $\theta = (A, b, D)$. The proof is in Appendix A.1.

Theorem 2.5 (Local identifiability for linear Gaussian SDEs)**.** *Consider model (2.3) with* $X_0 \sim \mathcal{N}(m_0, C_0)$, $C_0 \succ 0$, *and let* $T \sim U(0, t_{max})$. *On the parameter space:*

$$\Theta_d = \mathbb{R}^{d\times d} \times \mathbb{R}^d \times \mathbb{S}^d_{++}, \qquad \mathbb{S}^d_{++} := \{D \in \mathbb{R}^{d\times d} : D = D^\top,\ D \succ 0\},$$

the parameters $\theta = (A, b, D)$ *are locally identifiable from the pooled law for Lebesgue-almost every* θ. *Specifically, outside a set of Lebesgue measure zero, the pooled moments of total degree at most* $4d - 1$ *uniquely determine each* θ *within a neighborhood* V_θ*:*

$$\theta', \theta'' \in V_\theta, \quad G(\theta') = G(\theta'') \implies \theta' = \theta''.$$

Idea of the proof It suffices to show that the pooled moments of degree at most $4d - 1$ distinguish nearby parameters. By the inverse function theorem, this holds wherever no nonzero perturbation $(\delta A, \delta b, \delta D)$ leaves all of these moments unchanged to first order, that is, wherever their Jacobian has full column rank. Because the moments depend analytically on the parameters, full rank at a single parameter implies full rank outside a set of measure zero [23]. The proof therefore reduces to checking full rank at one convenient parameter (Appendix A.2).

In one dimension, the calculation follows directly from the first three moments in (2.5). We give an explicit point at which their Jacobian determinant is nonzero, starting from a noise-free model with $C_0 = 0$ and $D = 0$. Continuity and an invertible change of variables then give a full-rank point with positive diffusion and any prescribed Gaussian initial law with $C_0 \succ 0$, as required in our main setup. In higher dimensions, computing the Jacobian explicitly becomes infeasible. Instead, we use test polynomials to show that a parameter perturbation leaving all the specified pooled moments unchanged to first order must itself vanish, again at a noise-free model. The same continuity and change-of-variables argument then extends the construction to the setting of the theorem (Appendices A.3–A.5).

2.3 A general occupation-measure argument

The local identifiability result uses the explicit Gaussian moment equations to identify the dynamical parameters directly. For general dynamical systems, such a direct argument may not be available. We therefore give a general sufficient condition based on mixture identifiability: if the pooled distribution determines the distribution of snapshot distributions contributing to it, we can recover their temporal evolution under the assumptions below.

We describe each snapshot distribution by a parameter z that specifies its full probability distribution, denoted by ρ_z. Let $\mathcal{M}$ be the space of possible values of z. In the one-dimensional Gaussian example, a snapshot is specified by its mean and variance:

$$z = (m, v), \qquad \mathcal{M} = \mathbb{R} \times (0, \infty), \qquad \rho_z = \mathcal{N}(m, v).$$

Thus $\mathcal{M}$ consists of all possible mean–variance pairs. In d dimensions, $z = (m, C)$, $\mathcal{M} = \mathbb{R}^d \times \mathbb{S}^d_{++}$, and $\rho_z = \mathcal{N}(m, C)$. For other distribution families, z must likewise specify the full distribution; a mean and covariance alone need not suffice.

As the population evolves, its distribution parameters trace a curve $z = \Phi_\theta(t)$, so that

$$\Phi_\theta : [0, t_{\max}] \longrightarrow \mathcal{M}, \qquad \rho^\theta_t = \rho_{\Phi_\theta(t)}.$$

For example, $\Phi_\theta(t) = (m_\theta(t), v_\theta(t))$ in one dimension. Here θ specifies the dynamics, whereas $\Phi_\theta(t)$ specifies the population distribution at time t. We equip $\mathcal{M}$ with a metric, which makes continuity of this curve meaningful,

and assume that Φ_θ is continuous. We also assume that $z \mapsto \rho_z(B)$ is measurable for every Borel set $B \subset \mathbb{R}^d$. This measurability requirement holds, for example, when $z \mapsto \rho_z$ is weakly continuous.

We denote the probability distribution of $\Phi_\theta(T)$ by Γ_θ, called the occupation measure:

$$\Gamma_\theta = (\Phi_\theta)_{\#}\nu, \qquad \Gamma_\theta(B) = \nu\{t : \Phi_\theta(t) \in B\}, \tag{2.6}$$

where ν is the sampling-time distribution, $\#$ denotes the pushforward, and $B \subset \mathcal{M}$ is any Borel set.

Proposition 2.6 (Conditional global identifiability)**.** *Let ν be a known probability distribution of time with full support on the interval $[0, t_{max}]$. Let*

$$\Theta_{z_0} = \{\theta \in \Theta : \Phi_\theta(0) = z_0\}, \quad \rho_{z_0} = \rho_0$$

be the admissible family with a common known initial point. Suppose:

(i) (Mixture identifiability) *For every $\theta, \theta' \in \Theta_{z_0}$, $G(\theta) = G(\theta')$ implies $\Gamma_\theta = \Gamma_{\theta'}$.*

(ii) (Nonrecurrence) *For every $\theta \in \Theta_{z_0}$, the parameter map Φ_θ is continuous and injective on $[0, t_{\max}]$.*

(iii) (Trajectory identifiability) *For every $\theta, \theta' \in \Theta_{z_0}$, $\Phi_\theta(t) = \Phi_{\theta'}(t)$ for all $t \in [0, t_{\max}]$, implies $\theta = \theta'$.*

Then $G(\theta) = G(\theta')$ implies $\theta = \theta'$, for any $\theta, \theta' \in \Theta_{z_0}$.

Idea of the proof Condition (i) recovers the occupation measure Γ_θ, and hence its support, the curve traced by Φ_θ. By condition (ii) this curve does not revisit any point, so it is a simple arc, and the shared starting point z_0 orients it. Each point on the arc can then be assigned its visiting time from the mass that Γ_θ places between z_0 and that point: for uniform sampling, this mass is $t/t_{\max}$ at $\Phi_\theta(t)$; for general ν, it is the cumulative probability $\nu([0, t])$. This recovers the time parametrization $t \mapsto \Phi_\theta(t)$ without assigning a time to any observation, and condition (iii) then identifies the dynamical parameters. The full proof is in Appendix A.6.

The following remarks explain the role of each assumption.

Remark 2.7 (On assumption (i): mixture identifiability)**.** *Assumption (i) asks that the mixing measure Γ_θ on $\mathcal{M}$ be recoverable from the pooled mixture $G(\theta)$. For general nonparametric infinite mixtures this is notoriously hard [1, 2] and remains largely open. Recent work by Le et al. [20] gives negative results that point to a fundamental difficulty: in broad classes, including Gaussian location–scale mixtures, distinct mixing distributions can produce the same mixture. We therefore retain mixture identifiability as an assumption, to be established for the model at hand.*

Remark 2.8 (On assumption (ii): nonrecurrence)**.** *Assumption (ii) asks that the parameter map $\Phi_\theta \colon [0, t_{\max}] \to \mathcal{M}$ be injective: no two distinct times visit the same point in parameter space. This fails whenever the dynamics produces a periodic or stationary trajectory in $\mathcal{M}$. For the linear Gaussian SDE, important failure modes include rotational recurrence, where the mean and covariance together repeat, and equilibrium degeneracy, where the initial distribution is a fixed-point of the SDE, making Φ_θ constant. Note that nonrecurrence is a condition on the occupied moment curve as a whole, and is logically independent of the auto-rotational invariance condition of [12], which concerns identifiability of parameters from ordered marginals.*

Remark 2.9 (On assumption (iii): trajectory identifiability)**.** *Assumption (iii) is the usual identifiability problem with temporally resolved observations. Assumptions (i)–(ii) recover the anchored trajectory $t \mapsto \Phi_\theta(t)$ from the pooled distribution; Assumption (iii) asks whether that trajectory determines θ. For linear Gaussian SDEs this is the separated-marginal problem addressed by [12], while for gradient-flow SDEs the analogous result is given by [11].*

2.4 From identifiability to estimation in practice

We have discussed identifiability of parameters under a continuous sample-time distribution as in 2.1. In applications, however, data are finite samples, often collected at finitely many times, and the snapshot distributions need not be Gaussian. We now discuss what the results imply in general settings, and how they motivate the estimation procedure used in Section 3.

Remark 2.10 (Finitely many sampling times)**.** *On a finite grid, if the Gaussian components in* $G_K(\theta) = \frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(m_\theta(t_k), C_\theta(t_k))$ *are distinct and nondegenerate, finite Gaussian-mixture identifiability [30, 35] recovers the unordered component set. If the components admit a unique ordering compatible with the known initial point* z_0 *and the assumed dynamics, the ordered trajectory on the grid follows. However, in general, the continuous identifiability result does not transfer automatically: recovering* (A, b, D) *from finitely many snapshot moments needs the grid to resolve the dynamics without aliasing, and [12] show that discrete sampling of a rotating system produces Nyquist aliasing, which is why their identifiability results are stated for continuously observed marginals. Theorem 2.5 assumes uniform continuous sampling and supplies no finite-grid guarantee.*

The general identifiability argument may appear to suggest a pseudotime learning approach: recover the components, order them by the known initial condition, and fit the dynamics. We clarify that this is not necessary. The identifiable object is the occupation measure Γ_θ, which determines the one-dimensional curve structure of the snapshot parameter map without assigning a time coordinate to any individual observation. The situation is analogous to an ordinary Gaussian mixture: identifiability recovers the component distributions, but the label of each individual sample in overlapping regions need not be recovered and may remain ambiguous. Here the latent times T_i are nuisance variables, not inferential targets.

Assuming that the pooled forward map $\theta \mapsto G(\theta)$ is identifiable for the model class at hand, we fit the dynamical parameters by matching the full observed and simulated pooled distributions, rather than a finite set of moments, whose use in Theorem 2.5 is restricted to linear SDEs with Gaussian snapshots. Concretely, we minimize a population-level discrepancy

$$\mathcal{D}\big(\widehat{\rho}_N,\, G(\theta)\big)$$

between the empirical pooled distribution and the model-implied pooled distribution, where $\mathcal{D}(P, Q) \geq 0$ with equality only when $P = Q$. Since $G(\theta)$ is generally not available in closed form, it is itself approximated by the empirical distribution of simulated particles (Section 3.2), and the objective computed in practice compares two empirical distributions. At the population level, global identifiability makes true θ_0 the unique zero of $\theta \mapsto \mathcal{D}(P_*, G(\theta))$; local identifiability gives uniqueness of the zero only in a neighborhood. These statements do not alone establish convergence of a finite-sample optimizer. They hold for any such $\mathcal{D}$, for example Wasserstein distances (for distributions with finite moments of the relevant order), the energy distance, or the maximum mean discrepancy below.

The discrepancy $\mathcal{D}$ used in Section 3 is the kernel maximum mean discrepancy (MMD) [9]. For a bounded, continuous, positive-definite kernel k with reproducing kernel Hilbert space $\mathcal{H}_k$, it is defined as

$$\mathrm{MMD}_k(P, Q) = \sup_{\|f\|_{\mathcal{H}_k} \leq 1} \big|\mathbb{E}_P f - \mathbb{E}_Q f\big|,$$

the maximum difference in expectations over the unit ball of $\mathcal{H}_k$. It can be computed as

$$\mathrm{MMD}_k^2(P, Q) = \mathbb{E}[k(X, X')] - 2\,\mathbb{E}[k(X, Y)] + \mathbb{E}[k(Y, Y')], \tag{2.7}$$

where X, X' have distribution P, Y, Y' have distribution Q, and all four are independent. The kernel is called characteristic if $\mathrm{MMD}_k(P, Q) = 0$ only when $P = Q$, so that MMD is a discrepancy of the above type; the Gaussian kernel is characteristic [9]. We use MMD because the expectations in (2.7) are straightforward to estimate and differentiate on mini-batches.

Section 3 discusses two approaches. One follows the pseudotime reading literally, assigning each X_i a soft responsibility for each stage and alternating between label recovery and parameter update in the manner of EM. The other bypasses label recovery entirely, comparing the observed pooled distribution with the model-implied one directly. The first is effective when snapshot distributions are well separated but can settle into incorrect assignments when they overlap, the system is near equilibrium, or the moment curve is periodic. The second targets the identifiable object of Theorem 2.5 and Proposition 2.6 directly and avoids this instability. We describe both next.

3 Algorithms

3.1 Gaussian hidden-assignment EM

For the linear Gaussian model, the latent times T_i can be handled by a generalized EM procedure, analogous to what has been applied to discrete linear systems [15]. The E-step computes posterior responsibilities for snapshot membership. The M-step forms soft snapshot moments and fits the drift and diffusion parameters to the ordered moment sequence.

Algorithm 1 Gaussian hidden-assignment dynamics inference

Require: Pooled observations $\{X_i\}_{i=1}^N$, stage times $t_1 < \cdots < t_K$

Require: Initial parameters $\theta^{(0)} = (A^{(0)}, b^{(0)}, D^{(0)})$

1: **for** $r = 0, 1, 2, \ldots$ until convergence **do**

2: Compute model snapshot moments $m_k^{(r)} = m_\theta(t_k; \theta^{(r)})$ and $C_k^{(r)} = C_\theta(t_k; \theta^{(r)})$.

3: Compute responsibilities

$$\gamma_{ik}^{(r)} = \frac{\mathcal{N}(X_i; m_k^{(r)}, C_k^{(r)})}{\sum_{\ell=1}^K \mathcal{N}(X_i; m_\ell^{(r)}, C_\ell^{(r)})}.$$

4: Compute soft snapshot moments

$$\hat{m}_k^{(r)} = \frac{\sum_i \gamma_{ik}^{(r)} X_i}{\sum_i \gamma_{ik}^{(r)}}, \quad \hat{C}_k^{(r)} = \frac{\sum_i \gamma_{ik}^{(r)} (X_i - \hat{m}_k^{(r)})(X_i - \hat{m}_k^{(r)})^\top}{\sum_i \gamma_{ik}^{(r)}}.$$

5: Initialize ordering $\pi \leftarrow \mathrm{id}_{\{1,\ldots,K\}}$.

6: **repeat**

7: **for** $j = 1, \ldots, K-1$ **do**

8: Let π' swap positions $j, j{+}1$ of π; refit $A_\pi, A_{\pi'}$ by least squares to the midpoint mean-difference equation under each ordering, and let π' replace π if the summed mean- and covariance-Lyapunov-equation residual over consecutive pairs is smaller under π' than under π.

9: **end for**

10: **until** no swap in the sweep reduces the residual

11: Fit $\theta^{(r+1)}$ to the ordered soft moments $(\hat{m}_{\pi(1)}^{(r)}, \hat{C}_{\pi(1)}^{(r)}), \ldots, (\hat{m}_{\pi(K)}^{(r)}, \hat{C}_{\pi(K)}^{(r)})$ using (2.4).

12: **end for**

13: **return** $\hat{\theta}$ and responsibilities $\hat{\gamma}$.

The method is fast when the Gaussian model is correct. The local-search ordering procedure (Lines 6–10) is the practical locus of the failure modes in Remark 2.8: when components overlap, sit near equilibrium, or lie on a recurrent curve, the ordering is ambiguous and the M-step can lock onto a self-consistent but incorrect assignment. It also degrades when the snapshot laws are strongly non-Gaussian.

Remark 3.1. *The M-step of Algorithm 1 fits dynamics parameters from an ordered sequence of soft-assigned moments. This step can be replaced by any method that takes separated temporal marginals as input, such as*

APPEX [12] or other temporal-snapshot estimators. The EM wrapper adds hidden-assignment handling on top of whatever separated-marginal method is used in the inner loop.

3.2 Population-MMD simulator fitting

Explicitly recovering the one-dimensional curve structure from finite samples is a difficult statistical problem in its own right [1, 2], even when the snapshot law family is known. The population-MMD algorithm sidesteps this by avoiding likelihoods and matching population laws directly [7]. Let $\hat{\rho}^{\theta}_{t_k}$ be the empirical law of simulated particles at time t_k. Assuming equal weights for each time point, the simulated pooled law is

$$\hat{\rho}^{\theta}_{\mathrm{sim}} = \frac{1}{K}\sum_{k=1}^{K} \hat{\rho}^{\theta}_{t_k}, \tag{3.1}$$

the simulated analogue of the pooled forward map $G_K(\theta)$ of Remark 2.10. We minimize the squared MMD between $\hat{\rho}_N$ and $\hat{\rho}^{\theta}_{\mathrm{sim}}$.

Algorithm 2 Population-MMD dynamics inference

Require: Pooled observations $\{X_i\}_{i=1}^N$, stage times $t_1 < \cdots < t_K$
Require: Differentiable simulator with parameters θ
Require: Kernel k, initial parameters $\theta^{(0)}$
1: **for** $r = 0, 1, 2, \ldots$ until convergence **do**
2: Sample an observed mini-batch $\mathcal{B} \subset \{X_i\}_{i=1}^N$.
3: Sample initial simulated agents and Brownian noise variables.
4: Simulate the agent system forward and collect particles at $t_1, \ldots, t_K$.
5: Form the simulated mixture $\frac{1}{K}\sum_k \hat{\rho}^{\theta}_{t_k}$.
6: Compute the mini-batch loss

$$\mathcal{L}^{(r)} = \mathrm{MMD}_k^2\Big(\hat{\rho}_{\mathcal{B}},\ \frac{1}{K}\sum_{k=1}^{K}\hat{\rho}^{\theta}_{t_k}\Big).$$

7: Update θ by stochastic gradient descent.
8: **end for**
9: **return** $\hat{\theta}$.

Given observed data $\{X_i\}_{i=1}^N$ and simulated particles $\{Y_j\}_{j=1}^M$ pooled over $t_1, \ldots, t_K$, the expectations in (2.7) are approximated by sample averages,

$$\widehat{\mathrm{MMD}}_k^2 = \frac{1}{N^2}\sum_{i,i'=1}^{N} k(X_i, X_{i'}) - \frac{2}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} k(X_i, Y_j) + \frac{1}{M^2}\sum_{j,j'=1}^{M} k(Y_j, Y_{j'}).$$

In practice, we compute gradients by automatic differentiation and optimize the mini-batch objective with stochastic first-order methods [3, 28]. In all experiments, for MMD, we use a Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2/2h^2)$ with bandwidth h set by the median pairwise-distance heuristic on the pooled observations, and random Fourier features are used to approximate MMD on mini-batches [27].

4 Numerical experiments

We evaluate the pooled-law inference framework in three settings. First, we use two-dimensional linear Gaussian SDEs, where the snapshot laws are available in closed form and the ground-truth drift matrix is known. This setting lets us compare direct population matching (Algorithm 2) with a hidden-assignment EM baseline (Algorithm 1) under controlled identifiable and non-identifiable regimes. Second, we study a position-only

Table 1: *Linear-Gaussian test regimes. All cases use $K = 10$ uniformly spaced sampling times in the stated window, with equal spacing from $t_0 = 0$, and known initial law $X_0 \sim \mathcal{N}(m_0, \nu_0 I)$.*

Regime	Drift matrix A	m_0	ν_0	$[t_1, t_K]$
Damped rotation	$\begin{pmatrix} -0.05 & -3.93 \\ 3.93 & -0.05 \end{pmatrix}$	$(0.5, 0)$	0.004	$[0.135, 1.35]$
Stable linear drift	$\begin{pmatrix} -1.10 & 0.50 \\ 0.10 & -1.30 \end{pmatrix}$	$(2.2, -2.0)$	0.005	$[0.11, 1.10]$
Growing spiral	$\begin{pmatrix} 0.30 & -2.90 \\ 2.90 & 0.30 \end{pmatrix}$	$(0.55, 0)$	0.002	$[0.308, 3.08]$
Long rotation (aliased)	$\begin{pmatrix} -0.03 & -3.93 \\ 3.93 & -0.03 \end{pmatrix}$	$(1.0, 0)$	0.004	$[0.34, 3.40]$

interacting-particle system, where snapshot densities are unavailable and inference must be carried out through a differentiable simulator. This experiment demonstrates that the same population-MMD principle applies beyond Gaussian closure and linear dynamics. Finally, we apply the method to single-cell gene-expression data, one of the motivating biological settings for pooled-snapshot inference.

Across the simulations, the population-MMD method (Algorithm 2) receives only pooled observations and matches the empirical pooled distribution to a model-implied mixture of snapshot laws, without assigning a stage label to any observation. The sampling grid and initial population law are assumed known; the hidden variables are the realized stage labels of individual observations. The hidden-EM baseline, used in the linear Gaussian experiments, receives the same pooled observations but explicitly introduces latent stage assignments before fitting the dynamics. In the interacting particle experiments, where Gaussian mixture responsibilities and closed-form moment updates are unavailable, we compare pooled population-MMD with simulator-based diagnostic baselines. The single-cell experiment is used as a real-data illustration: per-cell stage labels are withheld from the pooled estimator, and the resulting gene-regulatory ranking is compared with published gene regulatory network benchmarks.

4.1 Linear stochastic agents

We first consider two-dimensional linear additive-noise SDEs

$$dX_t = AX_t\,dt + \sigma\,dW_t,$$

with known initial law. This setting provides a controlled testbed in which the snapshot laws are Gaussian and the ground-truth drift matrix is known, while still allowing the direct population-matching method to be implemented through the same simulator-based procedure used in the nonlinear experiments.

We use four regimes: damped rotation, stable linear drift, growing spiral, and a long rotation designed to create temporal aliasing after pooling. Table 1 lists the parameters. The damped rotation is observed for less than one period, while the long rotation is observed over two periods and therefore makes temporal traversal ambiguous from the pooled law alone.

For each regime, observations are generated at the $K = 10$ sampling times and then pooled. The pooled methods do not observe the sampling-time labels. We vary

$$\sigma \in \{0.02,\ 0.04,\ 0.08,\ 0.12,\ 0.16\},$$

use 1000 observations per stage, and average over 10 random seeds. Population MMD uses mini-batches of size

128, 96 simulated particles, 192 random Fourier features and 400 gradient steps. The known-time oracle is fit for 400 steps.

We now compare the two algorithms proposed in Section 3. Population MMD (Algorithm 2) simulates particles forward using Euler–Maruyama with step size 0.01, pools the simulated snapshots at $t_1, ..., t_K$ with equal weights, and matches the result to the observed pooled population by MMD. The time grid is known, but per-observation time labels are hidden; no pointwise matches or stage assignments are used. Hidden EM (Algorithm 1) instead introduces soft latent stage assignments and alternates between estimating these assignments and refitting the linear Gaussian dynamics from the resulting soft moments. Both pooled methods are initialized from the same label-free ordered-moment estimate. The known-time oracle receives the true stage groups, but not individual trajectories, and fits the Gaussian moment equations from empirical per-stage moments. To examine sensitivity to the assumed sampling-time distribution, we also fit population MMD assuming $T \sim \mathrm{Unif}[0, t_K]$, while the data remain generated on the ten-point grid. All other settings, including the initialization, match the grid-based fit.

Figure 2 summarizes the results. Each row corresponds to one regime. Columns 1 and 3 show observed snapshots together with fitted mean trajectories at the smallest and largest diffusion levels, $\sigma = 0.02$ and $\sigma = 0.16$, while columns 2 and 4 show the corresponding uncolored pooled inputs available to the pooled estimators. The rightmost column reports the Frobenius drift error $\|\widehat{A} - A\|_F$ as a function of σ, averaged over 10 random seeds. Stage colors are used only for visualization and are not observed by the pooled estimators.

The known-time oracle provides a best-case reference because it is given the true temporal grouping. Since the pooled methods observe only the samples, without their sampling times, their drift errors are expected to exceed the oracle's at a given sample size. The central comparison is therefore between the pooled methods: population MMD and hidden EM. In the identifiable regimes, population MMD is consistently more stable than hidden EM as diffusion increases. Hidden EM matches population MMD only in the linear case at low to moderate noise, where the snapshots are sequential and disjoint, so stage assignment is unambiguous. In the rotation and the growing spiral, population MMD is more accurate even at the lowest noise level, and as noise grows, hidden EM degrades sharply in all three cases. Population MMD degrades more gradually in all three cases, consistent with its direct optimization of the pooled law without an intermediate label-recovery step. Appendix B, Figure 7 shows the time labels recovered by hidden EM: they are accurate when snapshots are well separated but degrade as snapshot distributions overlap, and in the growing spiral they are assigned to the wrong times even at low noise. Assuming a continuous uniform time law instead of the true grid increases the drift error of population MMD in most identifiable cases, but the fit generally remains more stable than hidden EM as noise increases.

The long-rotation case tells a different story. Both pooled methods fit visually plausible orbits at all noise levels, yet neither approaches the oracle: drift errors remain near 10^0–10^1 regardless of σ. This confirms the intended aliasing regime as anticipated by Remark 2.10.

Finally, we also examine sensitivity to errors in the supplied initial law in Appendix B, Figures 8 and 9. In the linear regime, the drift estimate is nearly unaffected. In the rotation case, increasing the first coordinate of the supplied initial mean by one initial standard deviation roughly doubles the drift error, while the MMD between the fitted and observed distributions changes little. A plausible explanation is that the stable linear system attenuate errors, whereas in the rotation example, a shift in the starting position on the orbit may be partly compensated by changes in the fitted drift.

4.2 Position-only interacting particles

We next test the pooled estimator in a nonlinear simulator where closed-form snapshot densities are unavailable. Models of this type are commonly used to simulate collective-animal-behavior such as swarming and schooling [13,

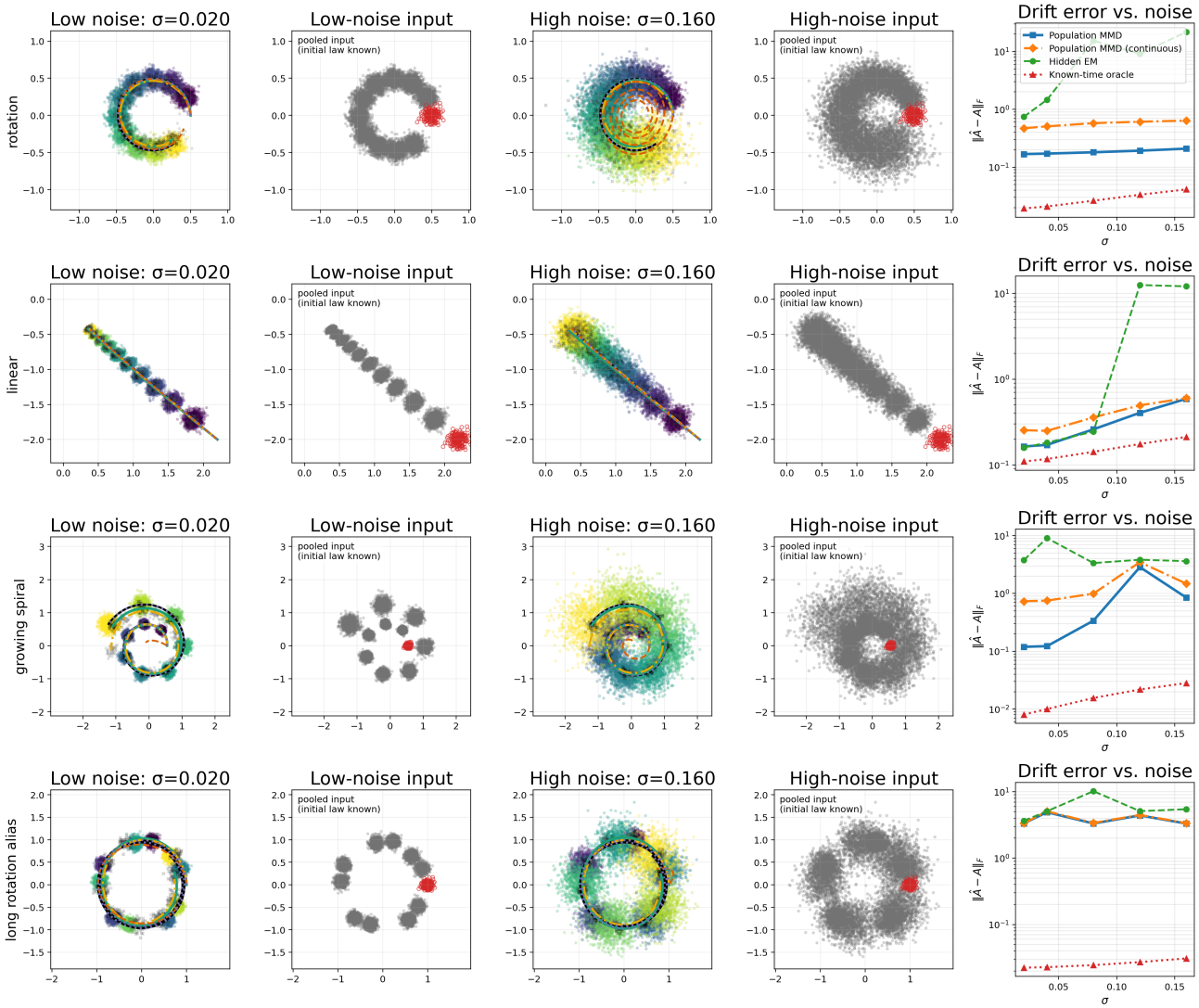


Figure 2: *Linear-Gaussian experiments, one row per regime. Columns 1 and 3: observed snapshots with fitted mean trajectories at* $\sigma = 0.02$ *and* $\sigma = 0.16$*; columns 2 and 4: pooled inputs (stage labels hidden); column 5:* $\|\widehat{A} - A\|_F$ *vs.* σ *(log scale). Trajectory colors: black = true mean, green = population-MMD with discrete temporal marginal, yellow = population-MMD with continuous temporal marginal, red dashed = hidden-EM, purple dotted = known-time oracle. In identifiable regimes, population MMD is more stable than hidden EM as snapshots overlap; in the long-rotation regime, a good pooled fit does not imply correct drift recovery (temporal aliasing).*

Table 2: *Interacting-particle regimes. During fitting, κ and ℓ are fixed at their true values, while λ and σ are estimated. All regimes use $K = 5$ snapshot times.*

Regime	κ	λ	ℓ	σ	snapshot times $\{t_k\}$
Identifiable attraction	0.25	2.20	0.75	0.045	$\{0.20, 0.45, 0.70, 0.95, 1.20\}$
Weak broad attraction	0.34	0.45	2.75	0.045	$\{0.20, 0.45, 0.70, 0.95, 1.20\}$
Short-range clustering	0.06	3.20	0.33	0.035	$\{0.16, 0.32, 0.52, 0.76, 1.00\}$

21], where each agent represents the location of an animal. Agent i evolves according to

$$dX_i = \left[-\kappa X_i + \frac{\lambda}{N} \sum_{j=1}^{N} \exp\left(-\frac{\|X_j - X_i\|^2}{2\ell^2} \right) (X_j - X_i) \right] dt + \sigma \, dW_i.$$

The first term is global contraction, while the second is finite-range aggregation. The observations are position-only: velocities are not observed, particle identities are not tracked across time, and the pooled estimator does not observe the stage label of any particle. The known initial population is used to initialize the simulator.

We consider three regimes, listed in Table 2. In all regimes, κ and ℓ are fixed at their true values during fitting, and we estimate only the interaction strength λ and diffusion scale σ. This isolates whether the aggregation strength is recoverable from position snapshots. Each stage contains 500 observed particles. Results are averaged over ten random seeds. The simulator is integrated by Euler–Maruyama with step size 0.04, and the MMD objective is approximated using 192 random Fourier features and optimized for 300 gradient steps with mini-batches of size 250 and 128 simulated particles.

Let X_i denote the pooled observations and let $Z_i \in \{1, \dots, K\}$ denote the unobserved stage label. The pooled estimator minimizes

$$\mathcal{L}_{\text{pooled}}(\theta) = \text{MMD}^2 \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}, \frac{1}{K} \sum_{k=1}^{K} \widehat{\rho}^{\theta}_{t_k} \right),$$

where $\widehat{\rho}^{\theta}_{t_k}$ is the simulated particle distribution at time t_k. This objective uses only the pooled position cloud.

We compare against two diagnostics. The first is a time-label oracle, which is given the stage membership Z_i of each observation and minimizes

$$\mathcal{L}_{\text{oracle}}(\theta) = \frac{1}{K} \sum_{k=1}^{K} \text{MMD}^2 \left(\frac{1}{N_k} \sum_{i:Z_i=k} \delta_{X_i}, \widehat{\rho}^{\theta}_{t_k} \right).$$

This oracle is not given particle identities, velocities, or trajectories; it only receives the otherwise hidden time grouping. The second diagnostic is a $\lambda = 0$ control, which removes the aggregation force and tests whether interaction is needed to explain the observed positional evolution.

Figure 3 summarizes the experiment. The first column shows the stage-colored snapshots for evaluation. The second column shows the actual input to the pooled estimator: one uncolored pooled point cloud, together with the known initial population. The third column shows the fitted simulator from pooled MMD, and the final column reports the estimated interaction strength and final pooled MMD for the pooled fit, the time-label oracle, and the $\lambda = 0$ control.

In the identifiable-attraction and weak broad-attraction regimes, pooled MMD captures the main stage-wise contraction and spread pattern and recovers λ close to its true value, with estimates near the time-label oracle (Figure 3, top panel of the last column). In both regimes, the $\lambda = 0$ control has substantially larger pooled MMD, showing that the attraction term leaves a detectable signature in the position snapshots.

The short-range-clustering regime shows the complementary failure mode. Both the pooled estimator and the time-label oracle severely underestimate λ, and the $\lambda = 0$ control attains comparable pooled MMD to the

interacting fits. This is due to the near-equilibrium degeneracy of the initial condition, a known failure mode noted in Remark 2.8: at this parameter setting, strongly clustered snapshots can be explained almost as well by clustered initial data evolving with little or no interaction, and the interaction strength is therefore not identifiable from position-only snapshots alone, even when time stamps are provided.

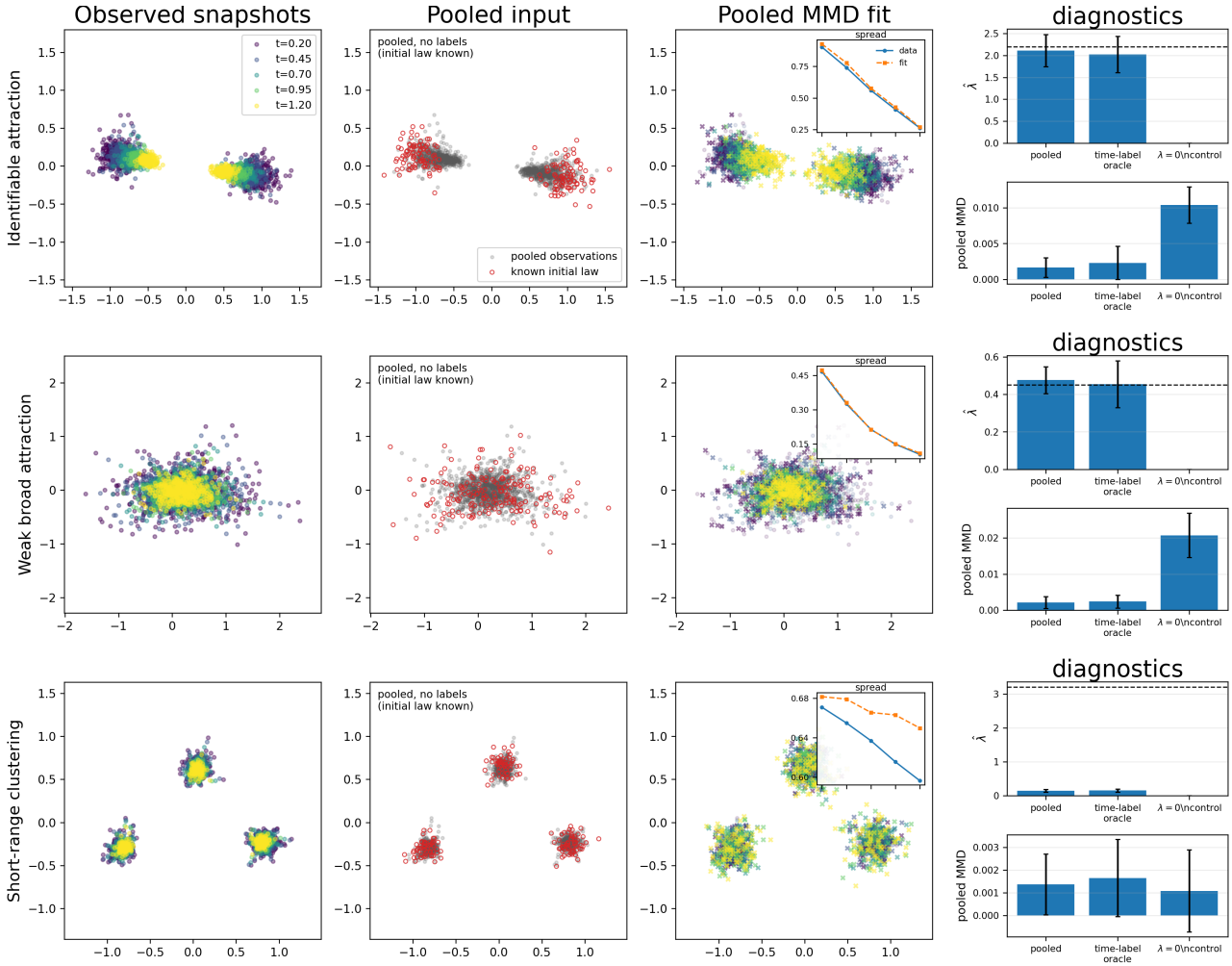


Figure 3: *Position-only interacting-particle experiment, one row per regime. Column 1: stage-wise snapshots (colors for visualization only); column 2: pooled input to the estimator (stage labels hidden, initial population highlighted); column 3: simulated snapshots from the pooled-MMD fit; column 4: estimated interaction strength and final pooled MMD for the pooled estimator, time-label oracle, and $\lambda = 0$ control.*

4.3 Gene regulatory dynamics from real single-cell snapshots

As a final illustration, we apply population-MMD to a single-cell gene expression dataset: qRT-PCR expression profiles of 45 transcription factors (TFs) in 960 THP-1 human myeloid leukemia cells collected after PMA-induced differentiation [19]. The data contain $K = 8$ collection times,

$$t = 0, 1, 6, 12, 24, 48, 72, 96 \text{ hours},$$

with 120 cells per time point and each cell observed once. We choose this dataset because its recorded collection times allow a direct comparison of inference with and without per-cell time labels: we withhold the per-cell stage label $Z_i \in \{1, \dots, 8\}$ from the pooled estimator, and compare against methods that use these labels. The

estimator receives the pooled expression vectors and a discrete sampling-time grid with equal weights. The initial population is estimated from the $t = 0$ cells, using Gaussian moment estimates for the linear model and empirical resampling for the neural models. We additionally assess sensitivity to the sampling-time marginal by replacing the discrete distribution with a continuous uniform distribution over the same normalized observation window. The performance changes modestly under this replacement (Appendix Figure 10). This is one of our motivating biological settings: individual cells are destructively observed, and their position along an underlying biological process may be latent or unreliable. Each cell is a vector in $\mathbb{R}^{45}$, one coordinate per gene. The goal of gene regulatory network inference is to determine which coordinates drive the increase or decrease of which others: an edge $j \to i$ means that the expression level of gene j affects the rate of change of gene i, with a positive (activating) or negative (repressing) sign. In our setting, these dependencies are read off from the fitted drift and compared with an experimentally derived network.

To visualize the 45-dimensional expression data, we show three pairwise projections of the leading three principal components in Figure 4. Cells from different collection times overlap in these projections, qualitatively resembling the overlapping snapshot distributions in the noisy linear-SDE examples. Such overlap illustrates why collection time need not uniquely specify expression state.

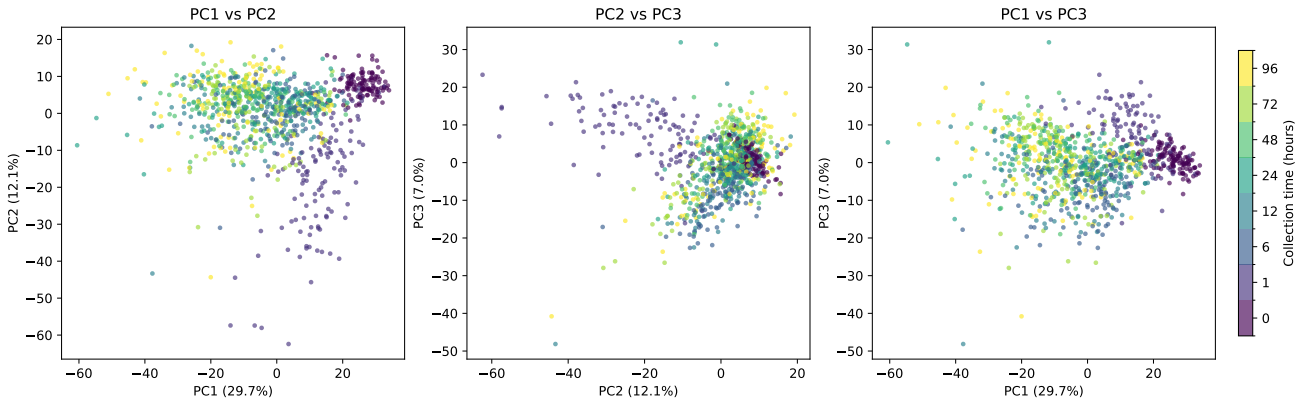


Figure 4: *Observed THP-1 expression profiles projected onto PC1–PC2, PC2–PC3, and PC1–PC3. Each point represents one cell, colored by collection time. All three projections use the same PCA fitted to the observed data, including the initial population.*

Drift models and GRN score We fit three drift/diffusion parameterizations by direct pooled-law matching (Algorithm 2), using uniform mixture weights:

(i) a linear SDE,

$$dX = (AX + b)\,dt + \sigma\,dW,$$

with a constant isotropic diffusion scale $\sigma > 0$;

(ii) a two-layer fully connected drift model,

$$dX = g_\theta(X)\,dt + \sigma\,dW,$$

again with a constant isotropic diffusion scale, where g_θ is a two-layer fully connected network with tanh activation;

(iii) the same two-layer fully connected drift model, but with state-dependent diagonal diffusion,

$$dX = g_\theta(X)\,dt + \mathrm{diag}(\sigma_\phi(X))\,dW,$$

where $\sigma_\phi : \mathbb{R}^d \to \mathbb{R}^d_{>0}$ is a second two-layer fully connected network with tanh activation and a softplus output map to guarantee positivity.

Thus the first two models use a single constant diffusion scale, while the third allows noise levels to vary with expression state and gene.

For the linear model, the directed GRN score for $\text{gene}_j \to \text{gene}_i$ is $|A_{ij}|$. For the neural drift models, the score is the cell-averaged absolute Jacobian of the drift,

$$\mathbb{E}_c \left| \frac{\partial g_{\theta,i}(X_c)}{\partial x_j} \right|,$$

representing the average effect of gene j on the rate of change in gene i across all observed cells c. The diffusion network in the third model is used only to fit the stochastic dynamics and does not directly enter the GRN score. In all cases, diagonal self-edges are excluded.

These scores rank candidate regulatory interactions according to the sensitivity of the fitted drift. A larger score for $\text{gene}_j \to \text{gene}_i$ indicates a stronger local dependence of the modeled rate of change of gene i on gene j.

Evaluation set We compare the resulting edge rankings against the regulatory network inferred from Matrix RNAi by [31], restricted to the 20-gene BCL6 subnetwork used in the SINCERITIES benchmark [25]. After excluding self-edges, the evaluation set contains $20 \times 19 = 380$ directed gene pairs, of which 68 (17.9%) are annotated as regulatory edges. AUROC and AUPRC are computed over these 380 directed pairs using edge existence only. These metrics summarize complementary aspects of the rankings. AUROC measures how consistently reference edges rank above non-edges across the full evaluation set, whereas AUPRC emphasizes recovering reference edges with high precision and is particularly sensitive to false positives near the top of the ranking [8]. The main evaluation concerns unsigned edge rankings. We report signed evaluation in Appendix Figure 11.

Results Figure 5 summarizes the fitted dynamics, and Table 3 reports the resulting GRN performance. In the PCA projection, the observed point cloud broadens along the differentiation process, indicating increasing heterogeneity at later stages. This changing spread is difficult for the constant-diffusion models to reproduce. The model with state-dependent diagonal diffusion better matches the breadth of the observed cloud that expands over time (purple to yellow), and this improved visual agreement is accompanied by the highest mean AUROC.

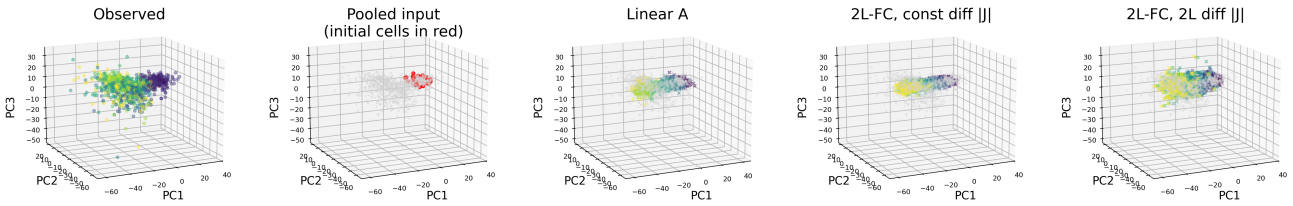


Figure 5: *Single-cell THP-1 differentiation expression profiles projected onto the first three principal components, with simulated trajectories from the pooled models. Stage colors are held out during fitting and shown only for visualization. Linear A = linear SDE; 2L-FC = two-layer fully connected drift; const diff = constant diffusion; 2L diff = state-dependent diagonal diffusion network.*

Table 3 reports the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPRC) for the pooled fits, with mean and standard deviation over 10 random seeds. Both metrics lie between 0 and 1, with larger values indicating better edge-ranking accuracy. We compare against the published unsigned GENIE3 and SINCERITIES results from [25] on the same 20-gene subnetwork. GENIE3 [16] is a time-free static baseline, while SINCERITIES is designed for this task and uses the per-cell temporal labels. All three population-MMD fits withhold those labels in the fitting objective.

The pooled linear SDE reaches AUROC/AUPRC $0.63 \pm 0.04/0.31 \pm 0.06$, improving over GENIE3 (0.46/0.23) but remaining below SINCERITIES (0.70/0.33). The nonlinear pooled models perform better: the two-layer

Table 3: *Unsigned AUROC/AUPRC on the 20-gene BCL6 subnetwork (68 positive edges among 380 directed non-self gene pairs). Our methods are fit by pooled population-MMD without using per-cell stage labels over 10 random seeds. GENIE3 and SINCERITIES numbers are the published unsigned results from [25], Table 2, on the same subnetwork and evaluation protocol.*

Method	AUROC	AUPRC
GENIE3 (time-free) [16]	0.46	0.23
Linear SDE, pooled (ours)	0.63 ± 0.04	0.31 ± 0.06
Two-layer drift, pooled (ours)	0.71 ± 0.05	0.35 ± 0.05
Two-layer drift + diffusion net, pooled (ours)	$\mathbf{0.75 \pm 0.04}$	$\mathbf{0.37 \pm 0.04}$
SINCERITIES (uses per-cell time labels) [25]	0.70	0.33

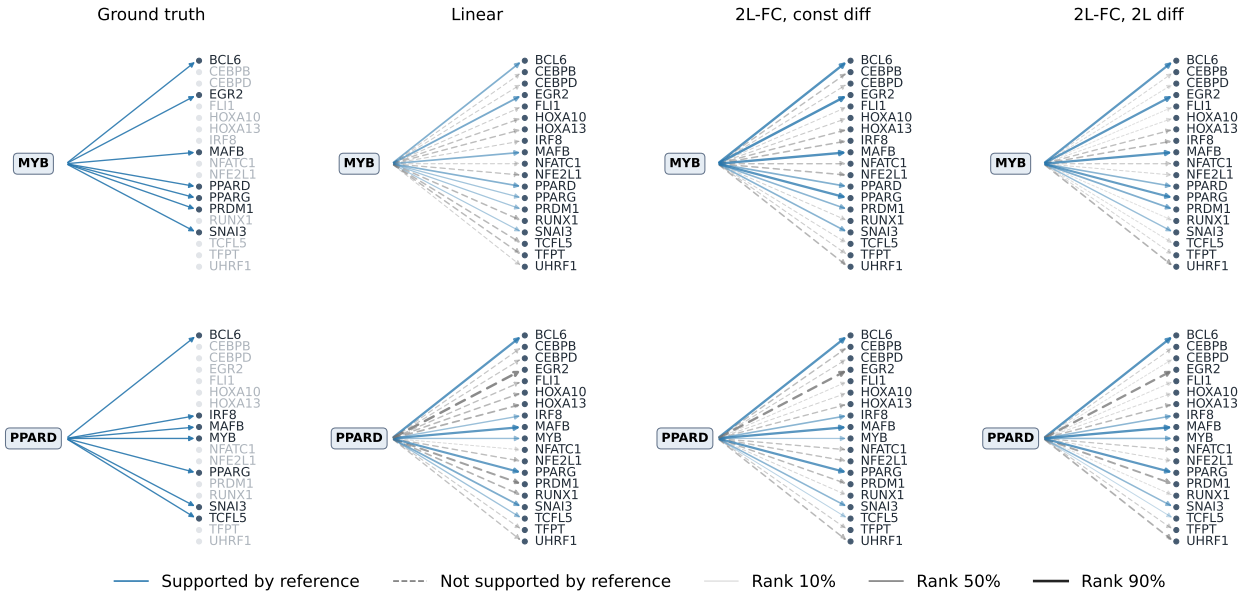


Figure 6: *Top edges outgoing from genes MYB and PPARD from reference (first column) and our algorithms with different models (second to fourth columns). Edge widths and opacity represent mean percentile ranks across all 380 evaluated gene pairs, using a common scale across panels.*

drift reaches $0.71 \pm 0.05/0.35 \pm 0.05$, and the two-layer drift with state-dependent diagonal diffusion reaches $0.75 \pm 0.04/0.37 \pm 0.04$. Both nonlinear models exceed SINCERITIES on both metrics, despite the latter using temporal information and a Granger causality-based methodology tailored for this task. Thus, on this benchmark, direct pooled-law matching with a flexible drift model gives regulatory rankings competitive with methods that use per-cell stage labels.

In addition to the aggregated metrics, we visualize the outgoing regulatory rankings of *MYB* and *PPARD*, the two genes with the largest numbers of outgoing edges in the reference. Edge width and opacity in Figure 6 indicate the mean percentile rank across 10 random seeds. In the linear model, edge weights are more evenly distributed across reference-supported and unsupported edges. The nonlinear models, particularly the model with state-dependent diffusion, tend to assign higher ranks to reference-supported edges (blue), shown as darker edges, and lower ranks to reference-unsupported edges (gray), shown as fainter edges. This pattern is consistent with their improved edge-ranking performance. For example, the reference-supported edges *MYB*→*MAFB* and *MYB*→*PPARG* receive higher ranks in both nonlinear models than in the linear model. Both nonlinear models also assign lower ranks to the reference-unsupported edges *MYB*→*RUNX1* and *PPARD*→*RUNX1*, with the largest reductions in the model with state-dependent diffusion.

The comparison shows that pooled matching alone is not enough: model flexibility matters. The linear pooled

model improves over a time-free baseline but does not match the specialized label-using method, whereas the nonlinear pooled models do. We therefore view this experiment as an illustration of the pooled-snapshot principle rather than a definitive GRN benchmark. The point is that biologically meaningful regulatory rankings can be recovered from pooled single-cell snapshots, without using per-cell stage labels in the fitting objective.

5 Discussion

This paper studies a trajectory-free inverse problem for stochastic agent systems: recovering dynamics from pooled population snapshots whose time labels are hidden. The central message is that the hidden observation time is often not the object one needs to learn. For linear Gaussian SDEs, pooled moments determine the parameters locally for almost every parameter value. For general snapshot families, the occupation-measure argument recovers the time-parametrized snapshot curve whenever the pooled law identifies its occupation measure and the curve is nonrecurrent, with periodic and stationary dynamics as the transparent exceptions. Empirically, direct pooled-law matching recovers the drift in identifiable regimes, while a hidden-assignment EM baseline that reconstructs time labels is less stable as snapshot distributions overlap, showing that the intermediate time-recovery step can be a liability.

The direct result for linear Gaussian SDEs is local and holds for almost every parameter in the admissible set. Global identifiability rests on mixture identifiability of the occupation measure, which remains open even for Gaussian snapshots [20]; the nonparametric mixture theory of Aragam et al. [1, 2] is a natural route. The result is also population-level: we do not quantify a statistical rate, nor give precise conditions for discrete observation and finitely many samples per marginal. Sharpening these, together with a characterization of nonrecurrence for nonlinear flows, is the main theoretical work left open.

The algorithm also has its own limits. The simulator-based estimator matches pooled laws through a kernel discrepancy whose cost grows with the number of particles, snapshot times, and state dimension; scaling to large single-cell panels will need more efficient discrepancies and trajectory sampling procedures, and MMD could be replaced by entropic optimal transport or other high-dimensional objectives. Two practical limitations deserve note. First, different mechanisms can produce nearly identical pooled laws: as in the last examples for linear and interacting particle systems, a good distributional fit does not imply correct mechanistic recovery, and resolving this requires external knowledge of the plausible parameter regime. Second, when the model class is misspecified, pooled matching fits the wrong dynamics to the right marginals with no internal signal that the model form is inadequate; model flexibility must therefore be chosen with the underlying biology or physics in mind, as what we observed for the GRN example.

Despite these limits, the results suggest that the pooled observation is a workable setting. Across synthetic linear systems, an interacting-particle model, and real single-cell differentiation data, dynamics learned directly from pooled population laws were competitive with estimators given the true observation times, without reconstructing per-observation labels. As destructive assays and high-throughput population measurements become the norm in biology and ecology, this perspective offers both a theoretical grounding and a practical path forward for data-driven agent-based modeling.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author used Claude (Anthropic) in order to assist with manuscript language polishing and checking consistency between the code and the manuscript description. After using this

tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

[1] Bryon Aragam, Chen Dan, Eric P. Xing, and Pradeep Ravikumar. Identifiability of nonparametric mixture models and Bayes optimal clustering. *The Annals of Statistics*, 48(4), August 2020.

[2] Bryon Aragam and Ruiyi Yang. Uniform consistency in nonparametric mixture models. *Ann. Stat.*, 51(1), February 2023.

[3] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, January 2018.

[4] Jonah Botvinick-Greenhouse, Robert Martin, and Yunan Yang. Learning dynamics on invariant measures using PDE-constrained optimization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(6), 2023.

[5] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[6] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations, 2018.

[7] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, May 2020.

[8] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning - ICML '06*, ICML '06, page 233–240. ACM Press, 2006.

[9] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

[10] Volker Grimm, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf M. Mooij, Steven F. Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, and Donald L. DeAngelis. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310(5750):987–991, November 2005.

[11] Vincent Guan, Joseph Janssen, Nicolas Lanzetti, Antonio Terpin, Geoffrey Schiebinger, and Elina Robeva. Gradient-flow SDEs have unique transient population dynamics, 2025.

[12] Vincent Guan, Joseph Janssen, Hossein Rahmani, Andrew Warren, Stephen Zhang, Elina Robeva, and Geoffrey Schiebinger. Identifying drift, diffusion, and causal structure from temporal snapshots, 2024.

[13] Mevin B. Hooten, Devin S. Johnson, Brett T. McClintock, and Juan M. Morales. *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, March 2017.

[14] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, October 2012.

[15] Tzu-Kuo Huang and Jeff Schneider. Learning linear dynamical systems without sequence information. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 425–432. ACM, 2009.

[16] Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE*, 5(9):e12776, 2010.

[17] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134, November 1918.

[18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[19] Tsukasa Kouno, Michiel de Hoon, Jessica C Mar, Yasuhiro Tomaru, Mitsuoki Kawano, Piero Carninci, Harukazu Suzuki, Yoshihide Hayashizaki, and Jay W Shin. Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome Biology*, 14(10), December 2013.

[20] Dung Le, Nicola Bariletto, Alessandro Rinaldo, and Nhat Ho. Partial differential equation barriers to identifiability in infinite mixture models, 2026.

[21] Fei Lu, Mauro Maggioni, and Sui Tang. Learning interaction kernels in heterogeneous systems of agents from multiple trajectories. *J. Mach. Learn. Res.*, 22(1), January 2021.

[22] W. Duncan Martinson, Alexandria Volkening, Markus Schmidtchen, Chandrasekhar Venkataraman, and José A. Carrillo. Linking discrete and continuous models of cell birth and migration. *Royal Society Open Science*, 11(7), 2024.

[23] B. S. Mityagin. The zero set of a real analytic function. *Mathematical Notes*, 107(3-4):529–530, March 2020.

[24] Jalil Nourisa, Antoine Passemiers, Jeremie Kalfon, Marco Stock, Berit Zeller-Plumhoff, Robrecht Cannoodt, Christian Arnold, Alexander Tong, Jason Hartford, Mihai G. Netea, Antonio Scialdone, Laura Cantini, Yves Moreau, Daniele Raimondi, Yang Li, and Malte D. Luecken. geneRNIB: a living benchmark for gene regulatory network inference. March 2025.

[25] Nan Papili Gao, S M Minhaz Ud-Dean, Olivier Gandrillon, and Rudiyanto Gunawan. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics*, 34(2):258–266, 2017.

[26] Aditya Pratapa, Amogh P. Jalihal, Jeffrey N. Law, Aditya Bharadwaj, and T. M. Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17(2):147–154, January 2020.

[27] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

[28] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[29] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A comparison of single-cell trajectory inference methods. *Nature Biotechnology*, 37(5):547–554, April 2019.

[30] Henry Teicher. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4):1265–1269, December 1963.

[31] Yasuhiro Tomaru, Christophe Simon, Alistair RR Forrest, Hisashi Miura, Atsutaka Kubosaki, Yoshihide Hayashizaki, and Masanori Suzuki. Regulatory interdependence of myeloid transcription factors revealed by Matrix RNAi analysis. *Genome Biology*, 10(11), November 2009.

[32] Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32(4):381–386, March 2014.

[33] Viska Wei and Fei Lu. Learning interacting particle systems from unlabeled data, 2026.

[34] Caleb Weinreb, Samuel Wolock, Betsabeh K. Tusi, Merav Socolovsky, and Allon M. Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10), February 2018.

[35] Sidney J. Yakowitz and John D. Spragins. On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39(1):209–214, February 1968.

[36] Yunan Yang, Levon Nurbekyan, Elisa Negrini, Robert Martin, and Mirjeta Pasha. Optimal transport for parameter identification of chaotic dynamics via invariant measures. *SIAM Journal on Applied Dynamical Systems*, 22(1):269–310, March 2023.

[37] Zhijun Zeng, Chenglong Bao, Pipi Hu, Yi Zhu, and Zuoqiang Shi. Reconstruction of dynamical systems without time label. *Communications in Computational Physics*, 39(3):941–968, January 2026.

[38] Stephen Zhang, Anton Afanassiev, Laura Greenstreet, Tetsuya Matsumoto, and Geoffrey Schiebinger. Optimal transport analysis reveals trajectories in steady-state systems. *PLOS Computational Biology*, 17(12):e1009466, December 2021.

[39] Wenjun Zhao, Elana J. Fertig, and Genevieve Stein-O'Brien. CycleGRN: Inferring gene regulatory networks from cyclic flow dynamics in single-cell rna-seq. November 2025.

[40] Wenjun Zhao, Erica Larschan, Björn Sandstede, and Ritambhara Singh. Optimal transport reveals dynamic gene regulatory networks via gene velocity estimation. *PLOS Computational Biology*, 21(5):e1012476, May 2025.

A Proofs for Section 2

A.1 Regularity and analytic dependence of moments

Proof of Lemma 2.4. The moment ODEs (2.4) are linear with constant coefficients, with closed forms

$$m_\theta(t) = e^{At}m_0 + \int_0^t e^{A(t-s)}b\,ds, \qquad C_\theta(t) = e^{At}C_0e^{A^\top t} + \int_0^t e^{A(t-s)}De^{A^\top(t-s)}\,ds, \tag{A.1}$$

verified by direct differentiation. Both are entire functions of t, giving (a) by continuity on the compact interval $[0, t_{\max}]$. For (b): e^{At} is invertible for every t, so the first term of $C_\theta(t)$ is C_0 conjugated by an invertible matrix and hence $\succ 0$; the diffusion $D = \Sigma\Sigma^\top \succeq 0$, so the second term is $\succeq 0$; a strictly positive-definite matrix plus a positive-semidefinite one is strictly positive-definite. By Weyl's inequality [14], $t \mapsto \lambda_{\min}(C_\theta(t))$ and $t \mapsto \lambda_{\max}(C_\theta(t))$ are continuous, hence attain a strictly positive minimum s_1 and a finite maximum s_2 on $[0, t_{\max}]$.

For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ with nonnegative integer entries, write $|\alpha| = \sum_i \alpha_i$ and $x^\alpha = \prod_i x_i^{\alpha_i}$. Every Gaussian moment $\mathbb{E}[X_t^\alpha]$ is a polynomial in the entries of $m_\theta(t)$ and $C_\theta(t)$. Formula (A.1) makes these entries analytic in the model parameters; the corresponding locally convergent analytic expansions are uniform on the compact time interval. Integrating against a fixed probability law ν preserves analyticity. All moments are finite

by the same compact-interval bounds. The moment formulas also extend analytically to arbitrary symmetric C_0, D, even when these matrices do not define probability laws. We use that extension only to construct a rank certificate at a degenerate limiting point.

□

A.2 Reduction to a rank condition at one point

Proof of Theorem 2.5. Fix m_0, C_0 and $t_{\max}$. Equal pooled laws have equal moments, so it suffices to show that the pooled moments of degree at most $4d-1$,

$$F(\theta) = \Big(\int x^\alpha \, G(\theta)(dx) \Big)_{1 \le |\alpha| \le 4d-1}, \tag{A.2}$$

distinguish nearby parameters. By the inverse function theorem, this holds near θ whenever the following is true at θ:

> if a perturbation $(\delta A, \delta b, \delta D)$ leaves every entry of F unchanged to first order, then $\delta A = 0$, $\delta b = 0$ and $\delta D = 0$;

that is, the Jacobian $DF(\theta)$ has full column rank $p_d = 3d(d+1)/2$, the number of parameters.

It is enough to verify this at a single parameter θ^*. Full rank at θ^* means that some p_d entries of F have a square Jacobian with nonzero determinant $\Delta(\theta^*)$. Since F depends analytically on θ (Lemma 2.4), so does Δ, and a real-analytic function on a connected open set that is nonzero at one point vanishes only on a closed set of measure zero [23]. Full rank, and hence local identifiability, therefore holds on an open dense set U of full measure. In particular, the set of locally nonidentifiable parameters has Lebesgue measure zero.

We verify the condition at one convenient point: by direct computation in one dimension (Appendix A.3), and with test polynomials in general dimension (Appendix A.4). Appendix A.5 moves the point to the prescribed initial law and horizon.

A.3 The scalar case: three moments determine three parameters

We first illustrate the argument in one dimension, where every step can be written out. Consider

$$dX_t = (aX_t + b)\, dt + \sqrt{D}\, dW_t, \qquad T \sim \mathrm{Unif}[0,1].$$

We compute at the noise-free point

$$X_0 = 0, \qquad a = 0, \qquad b = 1, \qquad D = 0,$$

where the snapshot mean and variance are $m(t) = t$ and $v(t) = 0$. This point is outside the positive-definite setting of the theorem. We use the analytic extension of the moments from Appendix A.1 to compute their derivatives there; continuity and rescaling in Appendix A.5 then give a full-rank point with the prescribed positive initial variance and positive diffusion.

The pooled moments By (2.5), the first three pooled moments are

$$\mu_1 = \int_0^1 m\, dt, \qquad \mu_2 = \int_0^1 (m^2 + v)\, dt, \qquad \mu_3 = \int_0^1 (m^3 + 3mv)\, dt.$$

How a perturbation moves them For a parameter direction $(\delta a, \delta b, \delta D)$, differentiate the mean and variance equations $\dot{m} = am + b$ and $\dot{v} = 2av + D$ at this point. With the initial law held fixed, the first-order changes satisfy

$$u'(t) = \delta a\, t + \delta b, \quad u(0) = 0, \qquad V'(t) = \delta D, \quad V(0) = 0.$$

Thus

$$u(t) := \delta m(t) = \tfrac{1}{2}\delta a\, t^2 + \delta b\, t, \qquad V(t) := \delta v(t) = \delta D\, t.$$

The term $2\delta a\, v$ vanishes because the baseline variance v is zero. Substitution into the three pooled moments gives

$$\begin{aligned} \delta\mu_1 &= \int_0^1 u\, dt && = \tfrac{1}{6}\delta a + \tfrac{1}{2}\delta b, \\ \delta\mu_2 &= \int_0^1 (2tu + V)\, dt && = \tfrac{1}{4}\delta a + \tfrac{2}{3}\delta b + \tfrac{1}{2}\delta D, \\ \delta\mu_3 &= \int_0^1 (3t^2 u + 3tV)\, dt && = \tfrac{3}{10}\delta a + \tfrac{3}{4}\delta b + \delta D. \end{aligned} \tag{A.3}$$

Each moment adds information The pooled mean does not depend on diffusion. The second moment brings in the variance change, and the third weights that change by the mean, separating diffusion from drift. With columns ordered as (a, b, D), the Jacobian is

$$J = \begin{pmatrix} 1/6 & 1/2 & 0 \\ 1/4 & 2/3 & 1/2 \\ 3/10 & 3/4 & 1 \end{pmatrix}, \qquad \det J = -\frac{1}{720} \neq 0.$$

Consequently, no nonzero parameter perturbation leaves all three moments unchanged to first order. This is the full-rank condition of Appendix A.2, using moments of degree at most $3 = 4d - 1$. The higher-dimensional construction below starts from the same noise-free setting; Appendix A.5 supplies the common extension to positive diffusion and any prescribed nondegenerate Gaussian initial law.

A.4 General dimension

For $d \geq 2$ the Jacobian is too large to compute explicitly as in Appendix A.3. We therefore verify the condition of Appendix A.2 using test polynomials, at a noise-free point where every snapshot lies on a known curve. As in the scalar case, this noise-free point is only a rank certificate: Appendix A.5 gives the common extension to positive diffusion, a prescribed nondegenerate Gaussian initial law, and the prescribed time horizon.

The baseline model Take $t_{\max} = 1$ and

$$m_0 = 0, \quad C_0 = 0, \quad A = N, \quad b = e_1, \quad D = 0,$$

where $N_{i,i-1} = 1$ for $i = 2, \ldots, d$ and all other entries of N are zero. This is the d-dimensional analogue of the scalar choice $a = 0$, $b = 1$ in Appendix A.3, to which it reduces when $d = 1$. Since $(Nx)_i = x_{i-1}$, the model reads

$$\dot{x}_1 = 1, \qquad \dot{x}_i = x_{i-1} \quad (i = 2, \ldots, d),$$

so each coordinate integrates the previous one. Starting from the origin and integrating this chain one coordinate at a time gives $x_1 = t$, $x_2 = t^2/2$, and so on. Because the initial state is fixed and there is no noise, every agent follows this same path,

$$\gamma(t) = \Big(t,\ \frac{t^2}{2!},\ \ldots,\ \frac{t^d}{d!}\Big), \tag{A.4}$$

Here d is the state dimension; the ith coordinate has degree i, so the curve has degree d. This curve is also the snapshot mean $m_\theta(t)$ at this baseline. Each snapshot is therefore a point mass at $\gamma(t)$: all agents observed at time t sit at the same point. Pooling over $T \sim \mathrm{Unif}[0,1]$ spreads the observations along the curve; for $d = 2$ this is the parabola arc $x_2 = x_1^2/2$, $0 \le x_1 \le 1$. The pooled moment of a polynomial φ is the average along the curve, $\int_0^1 \varphi(\gamma(t))\,dt$.

Two features of this choice are used below. Because N is nilpotent, e^{Nt} is a polynomial in t, so every quantity in the argument is a polynomial. And the coordinates of γ are distinct powers of t.

How a perturbation changes the snapshots Fix a perturbation direction $(\delta A, \delta b, \delta D)$ and consider the one-parameter family

$$\theta_\varepsilon = \big(N + \varepsilon\,\delta A,\ e_1 + \varepsilon\,\delta b,\ \varepsilon\,\delta D\big),$$

with snapshot mean $m_\varepsilon(t)$ and covariance $C_\varepsilon(t)$; at $\varepsilon = 0$ these are $\gamma(t)$ and 0. Write

$$u(t) := \frac{\partial m_\varepsilon(t)}{\partial \varepsilon}\Big|_{\varepsilon=0}, \qquad V(t) := \frac{\partial C_\varepsilon(t)}{\partial \varepsilon}\Big|_{\varepsilon=0}$$

for the first-order mean displacement and covariance variation, respectively. We use the analytic extension of polynomial moments from Appendix A.1 when the perturbed diffusion is not positive semidefinite. Differentiating the moment equations (2.4), $\dot m = Am + b$ and $\dot C = AC + CA^\top + D$, with respect to ε at $\varepsilon = 0$, where $m = \gamma$ and $C = 0$, gives

$$u' = Nu + \delta A\,\gamma + \delta b, \quad u(0) = 0, \qquad V' = NV + VN^\top + \delta D, \quad V(0) = 0, \tag{A.5}$$

Here the product rule gives $\delta(Am) = \delta A\,\gamma + Nu$. Solving these linear equations with zero initial data gives

$$u(t) = \int_0^t e^{N(t-s)}\big(\delta A\,\gamma(s) + \delta b\big)\,ds, \qquad V(t) = \int_0^t e^{N(t-s)}\,\delta D\,e^{N^\top(t-s)}\,ds.$$

The terms $\delta A\,C + C\,\delta A^\top$ are absent from the covariance equation because the baseline covariance C is zero. Since $(e^{Ns})_{ij} = s^{i-j}/(i-j)!$ for $i \ge j$ and zero otherwise, all entries are polynomials in t, with

$$\deg V_{ij} \le i + j - 1, \qquad \deg u_i \le d + i. \tag{A.6}$$

Indeed, a summand in V_{ij} contains powers $(t-s)^{i-k+j-l}$ with $k, l \ge 1$; integration adds one to give degree at most $i + j - 1$. A summand in u_i contains $(t-s)^{i-k}$ times a polynomial of degree at most d in s, so integration gives degree at most $d + i$.

How the snapshot changes move the pooled moments For a polynomial φ, let $\psi(m, C) := \mathbb{E}_{\mathcal{N}(m,C)}[\varphi(X)]$ be its expectation under one Gaussian snapshot. By (2.1), the pooled moment is the time average of snapshot moments:

$$M_\varphi(\varepsilon) := \int \varphi\,dG(\theta_\varepsilon) = \int_0^1 \psi\big(m_\varepsilon(t), C_\varepsilon(t)\big)\,dt.$$

Its first-order change is the derivative $\delta\int \varphi\,dG := M'_\varphi(0)$, which is the Jacobian DF applied to the perturbation direction. We compute it in three steps.

(i) Differentiate under the integral. The integrand is a polynomial in (m, C) (see (ii)) and $m_\varepsilon, C_\varepsilon$ depend smoothly on ε and t, so $M'_\varphi(0) = \int_0^1 \frac{d}{d\varepsilon}\psi(m_\varepsilon(t), C_\varepsilon(t))\big|_{\varepsilon=0}\,dt$. By the chain rule,

$$\frac{d}{d\varepsilon}\psi(m_\varepsilon, C_\varepsilon)\Big|_{\varepsilon=0} = \nabla_m\psi(\gamma, 0)\cdot u + D_C\psi(\gamma, 0)[V].$$

Thus u and V are the inner derivatives in this chain rule; $D_C\psi[V]$ denotes the directional derivative in the symmetric covariance direction V.

(ii) The partial derivatives of ψ. For a polynomial φ, expand $\varphi(m+Z)$ in its finite Taylor series, where Z is centered Gaussian with covariance C. The linear and all other odd-order terms have zero expectation, and the quadratic term has expectation $\frac{1}{2}\sum_{k,l} C_{kl}\partial_{kl}\varphi(m)$. By Isserlis' (Wick's) theorem [17], terms of order at least four contribute products of at least two covariance entries. Hence

$$\psi(m,C) = \varphi(m) + \tfrac{1}{2}\textstyle\sum_{k,l} C_{kl}\,\partial_{kl}\varphi(m) + \big(\text{terms of degree} \geq 2 \text{ in } C\big).$$

This formula also holds at $C=0$, where the snapshot is a point mass and $\psi = \varphi(m)$. Differentiating it at $C=0$ gives $\nabla_m\psi(m,0) = \nabla\varphi(m)$ and $D_C\psi(m,0)[V] = \frac{1}{2}\sum_{k,l} V_{kl}\partial_{kl}\varphi(m)$. The sum is over all ordered pairs, so it accounts for both symmetric off-diagonal entries.

(iii) Combine. Substituting (ii) into (i) at $m=\gamma(t)$ and integrating over t,

$$\delta\int \varphi\, dG = \int_0^1 \Big[\underbrace{\nabla\varphi(\gamma(t))\cdot u(t)}_{\text{displacement}} + \underbrace{\tfrac{1}{2}\textstyle\sum_{k,l} V_{kl}(t)\,\partial_{kl}\varphi(\gamma(t))}_{\text{spread}}\Big]dt. \tag{A.7}$$

For example, $\varphi(x) = x_1^2$ gives $\psi = m_1^2 + C_{11}$, so $\delta\int\varphi\,dG = \int_0^1 \big(2\gamma_1 u_1 + V_{11}\big)\,dt$, as in the second row of (A.3). The strategy is to choose test polynomials that switch off one of the two terms, so that the diffusion and drift perturbations can be recovered separately.

Test polynomials Any polynomial $\varphi(x)$, $x\in\mathbb{R}^d$, of degree at most $4d-1$ is a combination of monomials, so its pooled moment is a combination of the entries of F, and (A.7) applies to it. Suppose now that (A.7) vanishes for every such φ. We show that $\delta D = 0$, $\delta A = 0$ and $\delta b = 0$, which is injectivity of the Jacobian. Each step below makes one specific choice of φ and reads off what (A.7) then says. Both steps use the same mechanism: suitable tests give polynomial orthogonality identities, and testing a polynomial against itself forces it to vanish. We first eliminate diffusion so that the covariance term cannot contribute to the subsequent drift tests. This is the multidimensional counterpart of separating the drift and diffusion contributions in Appendix A.3.

The test polynomials are built from the following functions. For $i = 2,\dots,d$ let

$$g_i(x) = x_i - \frac{x_1^i}{i!}, \qquad n_i(t) := \nabla g_i(\gamma(t)) = e_i - \frac{t^{i-1}}{(i-1)!}\,e_1.$$

Since $\gamma_i(t) = t^i/i! = \gamma_1(t)^i/i!$, each g_i vanishes on the curve: $g_i(\gamma(t)) = 0$ for all t. Thus these functions detect departure from the curve. Two such factors will remove the mean-displacement term in Step 1; one factor will detect displacement transverse to the curve in Step 2(a). Also $\gamma_1(t) = t$, so a factor $p(x_1)$ evaluated on the curve becomes $p(t)$. We write δD_{ij} for the entries of δD, and use:

Lemma A.1. *If Q is a polynomial and $\int_0^1 p(t)Q(t)\,dt = 0$ for every polynomial p with $\deg p \leq \deg Q$, then $Q\equiv 0$.*

Proof. Take $p = Q$: then $\int_0^1 Q^2\,dt = 0$. □

Step 1: $\delta D = 0$ Fix $i,j\geq 2$ and a polynomial p in one variable, and take

$$\varphi(x) = p(x_1)\,g_i(x)\,g_j(x).$$

(a) The displacement term vanishes for this test. By the product rule, every term of $\nabla\varphi$ contains a factor g_i or g_j, so $\nabla\varphi(\gamma(t)) = 0$ and this term of (A.7) is zero. *(b) The covariance term gives polynomial constraints.* In $\partial_{kl}\varphi$, the only terms without a factor g_i or g_j are those in which one derivative falls on g_i and the other on g_j. Hence $\partial_{kl}\varphi(\gamma(t)) = p(t)\big[(n_i)_k(n_j)_l + (n_j)_k(n_i)_l\big]$, and, since V is symmetric, $\frac{1}{2}\sum_{k,l} V_{kl}\,\partial_{kl}\varphi(\gamma(t)) = p(t)\,Q_{ij}(t)$ with

$$Q_{ij}(t) := n_i(t)^\top V(t)\,n_j(t), \qquad \deg Q_{ij} \leq i+j-1$$

by (A.6). So (A.7) reads $\int_0^1 p(t)\,Q_{ij}(t)\,dt = 0$. Taking $\deg p \le \deg Q_{ij}$ keeps $\deg\varphi \le 2(i+j)-1 \le 4d-1$, so Lemma A.1 gives $Q_{ij} \equiv 0$ for all $i, j \ge 2$. The largest degree required here occurs at $i = j = d$, explaining the bound $4d-1$.

It remains to see that these identities force $\delta D = 0$. From (A.5), $V(t) = \delta D\, t + O(t^2)$, and $n_i(t) = e_i + O(t)$. Therefore:

- The coefficient of t in Q_{ij} is δD_{ij}, so $\delta D_{ij} = 0$ for all $i, j \ge 2$.
- Integrating the first two coordinates of (A.5) gives

$$V_{11} = \delta D_{11} t, \quad V_{12} = \delta D_{12} t + \tfrac{1}{2}\delta D_{11} t^2, \quad V_{22} = \delta D_{22} t + \delta D_{12} t^2 + \tfrac{1}{3}\delta D_{11} t^3.$$

  Since $n_2 = e_2 - te_1$, substitution gives

$$Q_{22} = V_{22} - 2tV_{12} + t^2 V_{11} = \delta D_{22} t - \delta D_{12} t^2 + \tfrac{1}{3}\delta D_{11} t^3.$$

  The coefficients must vanish, so $\delta D_{12} = \delta D_{11} = 0$.
- The only remaining entries are δD_{1j} with $j \ge 3$. With the entries identified above zero, $V_{11} = V_{12} = 0$ and

$$V_{1j} = \delta D_{1j} t + O(t^2), \qquad V_{2j} = \tfrac{1}{2}\delta D_{1j} t^2 + O(t^3).$$

  For the second formula, use $V_{2j}' = V_{1j} + V_{2,j-1}$ and $V_{2,j-1} = O(t^2)$, since $\delta D_{2,j-1} = 0$, and integrate once more. Thus

$$Q_{2j} = V_{2j} - tV_{1j} = -\tfrac{1}{2}\delta D_{1j} t^2 + O(t^3).$$

  The first-order coefficient had only detected δD_{2j}; the remaining entry δD_{1j} appears at second order through the shift dynamics and the factor t in n_2. Hence $\delta D_{1j} = 0$.

Thus $\delta D = 0$, and then $V \equiv 0$ by (A.5).

Step 2: $\delta A = 0$ and $\delta b = 0$ Now $V \equiv 0$, so the spread term of (A.7) is zero for every φ, and (A.7) involves only the mean displacement u. We show $u \equiv 0$ in two parts, and then read off δA and δb.

(a) The displacement lies along the curve. Fix $i \ge 2$ and a polynomial p, and take

$$\varphi(x) = p(x_1)\, g_i(x).$$

On the curve $g_i = 0$, so $\nabla\varphi(\gamma(t)) = p(t)\, n_i(t)$, and (A.7) reads

$$\int_0^1 p(t)\, H_i(t)\, dt = 0, \qquad H_i(t) := n_i(t)\cdot u(t) = u_i(t) - \frac{t^{i-1}}{(i-1)!}\, u_1(t), \qquad \deg H_i \le d + i.$$

Taking $\deg p \le \deg H_i$ keeps $\deg\varphi \le d + 2i \le 3d \le 4d-1$, so Lemma A.1, now applied to H_i just as it was to Q_{ij} in Step 1, gives $H_i \equiv 0$, that is, $u_i = \frac{t^{i-1}}{(i-1)!} u_1 = \gamma_i'(t)\, u_1$ for every $i \ge 2$. Since also $\gamma_1' = 1$, this says $u(t) = u_1(t)\,\gamma'(t)$: the displacement points along the curve.

(b) The displacement vanishes. Part (a) still allows the snapshots to slide along the curve. To rule out this remaining tangential displacement, we choose a test whose variation is the integral of u_1^2. Let P be a polynomial with $P' = u_1$, so $\deg P \le d+2 \le 4d-1$, and take $\varphi(x) = P(x_1)$. Then $\nabla\varphi(\gamma(t)) = u_1(t)\, e_1$, so the displacement term is $u_1(t)^2$, and (A.7) reads $\int_0^1 u_1(t)^2\, dt = 0$. Hence $u_1 \equiv 0$, and $u \equiv 0$ by part (a).

(c) Conclusion for the drift. Since $u \equiv 0$, also $u' \equiv 0$, and the mean equation in (A.5) reduces to $\delta A\,\gamma(t) + \delta b = 0$ for all t. The coordinates of γ are the distinct powers $t, t^2/2!, \ldots, t^d/d!$ noted at the baseline construction, which together with 1 are linearly independent, so $\delta A = 0$ and $\delta b = 0$.

Conclusion A perturbation that leaves every pooled moment of degree at most $4d-1$ unchanged to first order is therefore zero, so DF has full column rank p_d at this point. The choices of p and P above depend on the perturbation; this is allowed because (A.7) was assumed to vanish for every polynomial within the degree bound.

A.5 Moving the full-rank point to the prescribed initial law

Appendix A.3 for $d = 1$ and Appendix A.4 for $d \geq 2$ give full rank at the noise-free point

$$m_0 = 0, \quad C_0 = 0, \quad A = N, \quad b = e_1, \quad D = 0, \quad t_{\max} = 1,$$

with $N = 0$ and $e_1 = 1$ when $d = 1$. The following two steps work in every dimension and give an admissible full-rank point for the initial law $\mathcal{N}(m_0, C_0)$ and horizon $t_{\max}$ prescribed in Theorem 2.5.

(i) Small changes keep full rank. The Jacobian depends continuously on C_0 and θ, so a nonzero determinant stays nonzero nearby. Full rank therefore also holds at $m_0 = 0$, $C_0 = D = \epsilon I$, $A = N$, $b = e_1$, $t_{\max} = 1$, for some small $\epsilon > 0$. Here we vary the initial covariance only to construct a full-rank point; the Jacobian is still taken with respect to the dynamical parameters.

(ii) Changing variables keeps full rank. With $LL^\top = C_0$, the change of variables $Y_s = L^{-1}(X_{t_{\max} s} - m_0)$, $0 \leq s \leq 1$, turns the model with initial law $\mathcal{N}(m_0, C_0)$ and horizon $t_{\max}$ into one with initial law $\mathcal{N}(0, I)$ and horizon 1, with parameters

$$\bar{A} = t_{\max} L^{-1} A L, \quad \bar{b} = t_{\max} L^{-1}(A m_0 + b), \quad \bar{D} = t_{\max} L^{-1} D L^{-\top}. \tag{A.8}$$

This map of parameters is invertible. An invertible affine state transformation also gives an invertible affine map of the moments of degrees 1 through $4d - 1$ (the zeroth moment is the fixed constant 1), so the chain rule multiplies the moment Jacobian on the left and right by invertible matrices. Its rank is therefore unchanged.

Applying (ii) to the point in (i), with $L = \sqrt{\epsilon}\, I$, gives a full-rank point for $\mathcal{N}(0, I)$ and horizon 1; applying (ii) in reverse gives one for $\mathcal{N}(m_0, C_0)$ and $t_{\max}$ in every dimension. The argument of Appendix A.2 completes the proof. □

A.6 Recovery of the anchored curve from its occupation measure

Proof of Proposition 2.6. Recall that $\Phi_\theta : [0, t_{\max}] \to \mathcal{M}$ records the snapshot descriptor at each time, and $\Gamma_\theta = (\Phi_\theta)_\# \nu$ is its occupation measure, the distribution of $\Phi_\theta(T)$ for $T \sim \nu$. The argument proceeds in three stages: condition (i) recovers this measure; condition (ii), the common initial descriptor z_0, and full support of ν recover the time parametrization; condition (iii) then identifies the dynamical parameters.

Let $\theta, \theta' \in \Theta_{z_0}$ and suppose $G(\theta) = G(\theta')$. Condition (i) gives $\Gamma_\theta = \Gamma_{\theta'}$. By (ii), each Φ_θ is a continuous injection from a compact interval into the metric space $\mathcal{M}$, hence a homeomorphism onto its image. Full support of ν implies that the support of its pushforward is exactly that image. Thus the two curves have the same image.

Define $h = \Phi_\theta^{-1} \circ \Phi_{\theta'}$: $h(t)$ is the time at which the θ-curve visits the snapshot reached by the θ'-curve at time t. This is a homeomorphism of $[0, t_{\max}]$. The common initial descriptor gives $\Phi_\theta(0) = \Phi_{\theta'}(0) = z_0$, so injectivity yields $h(0) = 0$. Therefore h is strictly increasing. On the common image, equality of the occupation measures gives

$$h_\# \nu = (\Phi_\theta^{-1})_\# \Gamma_{\theta'} = (\Phi_\theta^{-1})_\# \Gamma_\theta = \nu.$$

Consequently, with $F_\nu(t) = \nu([0, t])$,

$$F_\nu(h(t)) = \nu([0, h(t)]) = \nu\big(h^{-1}([0, h(t)])\big) = \nu([0, t]) = F_\nu(t).$$

Here monotonicity gives $h^{-1}([0, h(t)]) = [0, t]$. Full support is used a second time to make F_ν strictly increasing: for $s < t$, the nonempty interval (s, t) has positive mass. It follows that $h(t) = t$ for every t. Thus the snapshot parametrizations agree, reducing identification to the known-time case, and condition (iii) gives $\theta = \theta'$.

□

For uniform time, the equivalent subarc formula is $t = t_{\max}\Gamma(\Phi_\theta([0, t]))$. Allowing atoms in the proposition does not permit arbitrary finite grids: a finite grid has no full support on an interval.

B Additional results for linear stochastic agents

B.1 Time labels recovered by hidden EM

Hidden EM assigns each pooled observation a posterior probability r_{ik} of belonging to sampling time t_k, computed from the fitted Gaussian snapshot laws with equal prior weights. We summarize these assignments by the posterior-mean time $\hat{t}_i = \sum_k r_{ik} t_k$ and compare it with the true sampling time (Figure 7), for one dataset per regime at the lowest and highest noise levels.

When the recovered times are accurate, hidden EM effectively fits the dynamics with known stage labels. In the stable linear case at low noise, times are recovered nearly exactly, and hidden EM is as accurate as population MMD (Figure 2). As noise increases and neighboring snapshot distributions overlap, the recovered times become less reliable. In the damped rotation, the time error increases from 0.078 at $\sigma = 0.02$ to 0.337 at $\sigma = 0.16$, together with a rapid increase in the drift error of hidden EM. In the growing spiral, the snapshot clusters are assigned to the wrong times even at low noise, consistent with the failure of hidden EM at every noise level. In the aliased long rotation, times are not recovered at either noise level.

B.2 Sensitivity to the supplied initial law

The theoretical results assume a known initial law, whereas in practice it must be estimated. To assess the effect of errors in this input, we generate data from the true initial law $\mathcal{N}(m_0, \nu_0 I)$ and fit population MMD with a perturbed supplied law, keeping the data and all other settings fixed. The supplied mean is shifted along the first coordinate by ± 0.5 and ± 1 initial standard deviations ($\sqrt{\nu_0} \approx 0.06$–$0.07$), or the supplied covariance is scaled by 0.5, 0.75, 1.5, or 2 with the mean held at its true value. The label-free initialization is recomputed with the supplied law. We consider the linear and rotation regimes at $\sigma = 0.08$ over 10 seeds.

We report the errors in A, in σ, and in the rotation frequency $\omega = (A_{21} - A_{12})/2$. We also report two squared MMDs of the fitted pooled law, computed with a Gaussian kernel whose bandwidth is set from the median pairwise distance of the observations: one against a fixed subset of the observed data (empirical), and one against the true data-generating mixture (population).

In the linear regime (Figure 8), the drift and frequency errors are nearly unchanged under both perturbations, while the diffusion error increases with the size of the perturbation. In the rotation (Figure 9), the drift error increases with the mean shift, from 0.18 to 0.38 at a shift of $+1$ standard deviation and to 0.24 at -1, whereas covariance scaling has little effect. In both regimes, the pooled MMDs change little across perturbations.

C Additional details and results for GRN inference

We first provide some implementation details. Expression values are centered gene-wise without log transformation or variance scaling. Models are fitted in all 45 gene dimensions; PCA is used only for visualization.

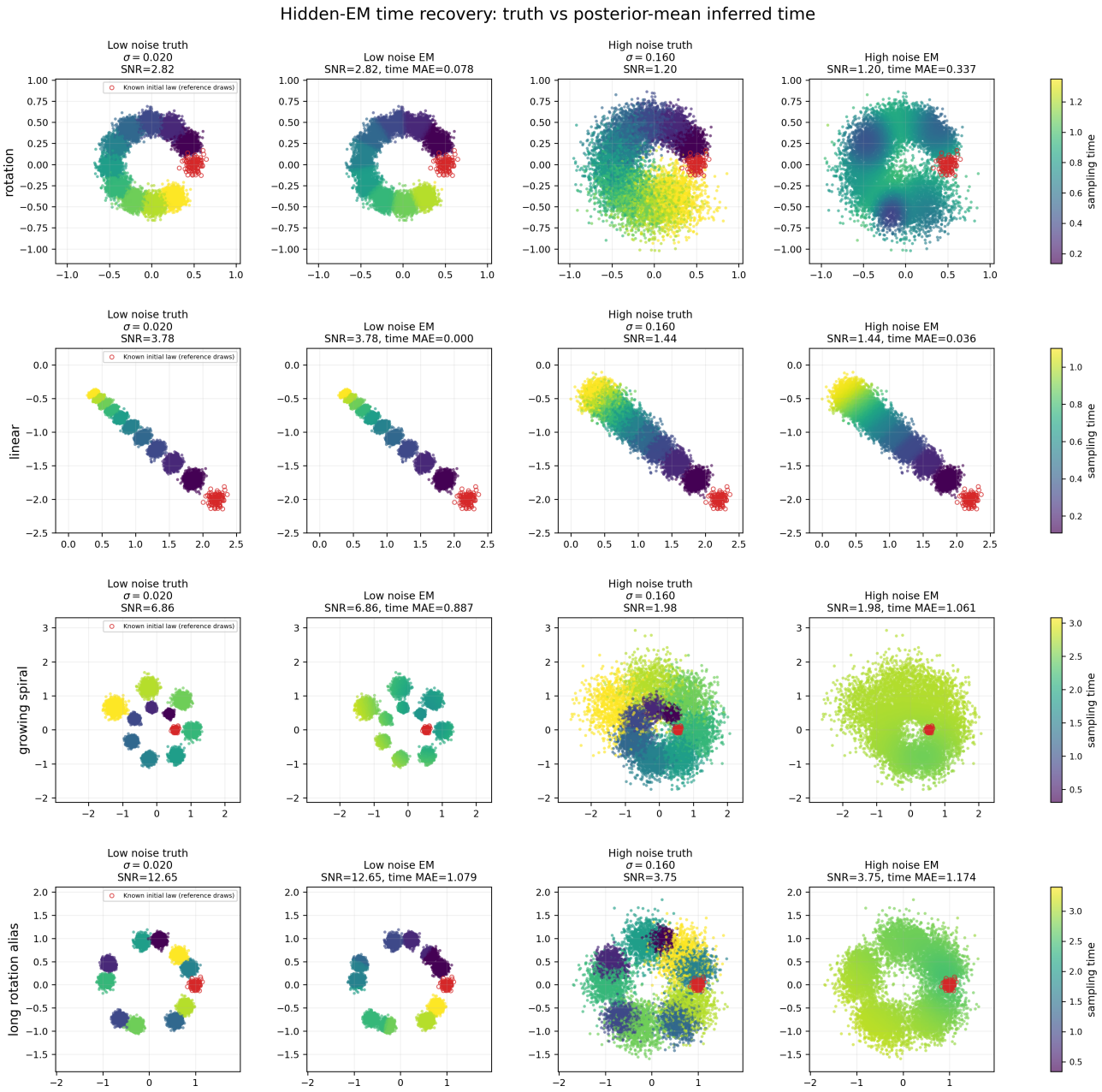


Figure 7: *True sampling times (first and third columns) and posterior-mean times recovered by hidden EM (second and fourth columns), at* $\sigma = 0.02$ *and* $\sigma = 0.16$*, for one dataset per regime. Colors within each row share the same time scale. Red circles: draws from the known initial law. Panel titles report the signal-to-noise ratio and the mean absolute time error.*

Neural drift and diffusion networks have architecture $45 \to 32 \to 45$ with tanh hidden activations and positive softplus diffusion outputs. We use Adam [18] with learning rates 0.02 for the linear model and 0.01 for the neural models, mini-batches of 128 cells, and at most 500 iterations with early stopping and gradient clipping at norm 20. MMD is approximated using 256 random Fourier features across Gaussian bandwidths $h/4, h/2, h, 2h$, where h is estimated by the median-distance heuristic. Linear moment equations are integrated by fourth-order Runge–Kutta; neural models use Euler–Maruyama with maximum step size 0.05 on the normalized time interval. Neural GRN scores are evaluated by averaging all 960 cells, and results are reported over 10 random seeds.

Next, we report two supplementary evaluations of the THP-1 application in Section 4.3: sensitivity to the assumed

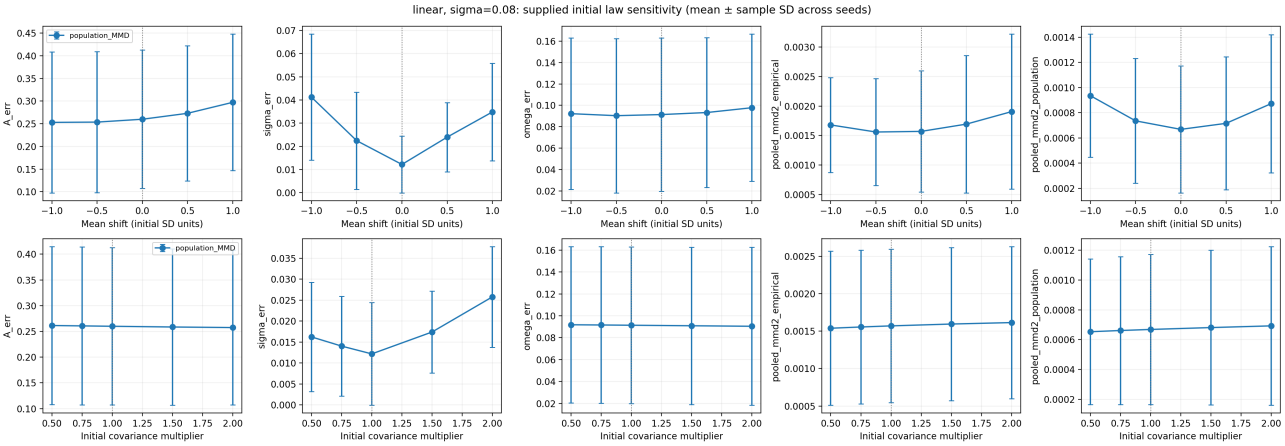


Figure 8: *Sensitivity to the supplied initial law, stable linear regime, $\sigma = 0.08$. Top row: supplied mean shifted in units of the initial standard deviation. Bottom row: supplied covariance scaled. Columns: errors in A, σ, and rotation frequency; empirical and population pooled squared MMD. Dotted vertical line: correct initial law. Mean $\pm$ SD over seeds.*

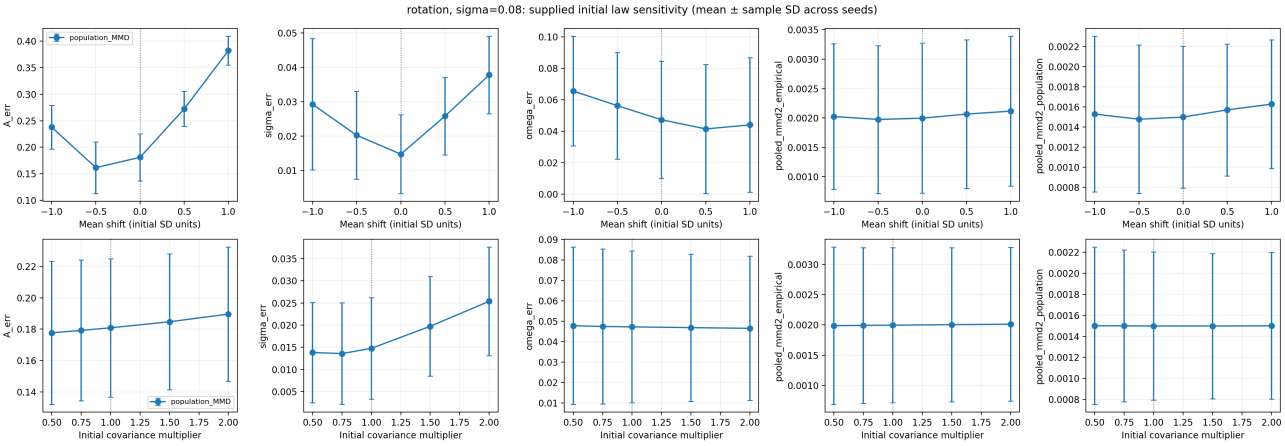


Figure 9: *As in Figure 8, for the damped rotation regime.*

sampling-time marginal (Figure 10) and signed edge evaluation (Figure 11). The modeling and validation setup are those of the main text, and results are mean $\pm$ standard deviation over the same 10 random seeds.

Collection hours are rescaled to $[0, 1]$. Each figure compares three fits per model class. *Pooled: grid* is the main-text fit, whose sampling law places equal weight on the eight rescaled collection times. *Pooled: uniform* instead assumes $T \sim \mathrm{Unif}[0, 1]$, which places more mass at late times than the grid does. For the neural models, times are drawn uniformly at each iteration, particles are integrated by Euler–Maruyama with steps ending exactly at the drawn times, and each particle is read off at one randomly assigned drawn time. For the linear model, particles are sampled from Gaussian snapshot laws whose means and covariances are computed numerically from the moment equations at the drawn times. Only the assumed time marginal differs between the two pooled fits. *Known-time oracle* fits the same model class by minimizing the average squared MMD between simulated and observed distributions at each collection time. It uses collection-time labels but requires no tracking of individual cells. It indicates what the time labels add and is not affected by the choice of time marginal.

For signed evaluation, each directed gene pair carries two candidate labels, activation and repression; the label matching the reference sign is positive and all others are negative. A predicted edge assigns its unsigned score to its predicted sign and zero to the opposite sign. A correct edge with the wrong sign therefore yields a

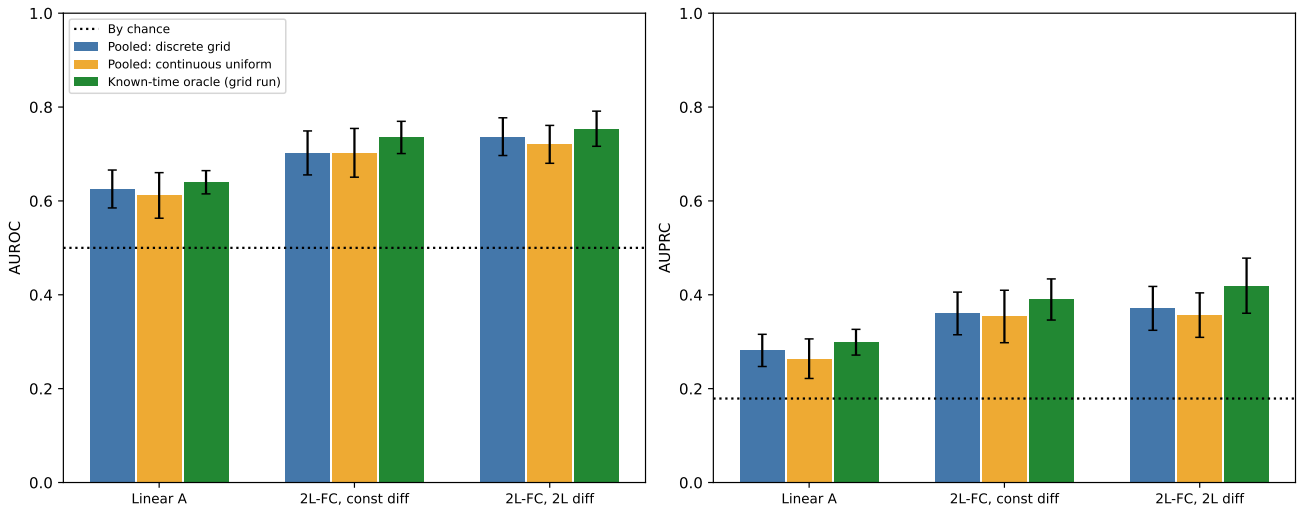


Figure 10: *Sensitivity of unsigned AUROC (left) and AUPRC (right) to the assumed sampling-time marginal on the 20-gene BCL6 subnetwork. Pooled fits assume either the discrete collection grid or a continuous uniform law on the same rescaled window; the oracle uses per-cell collection times. Mean ± SD over 10 seeds. Dotted line: chance (0.5 for AUROC; fraction of reference edges,* 68/380*, for AUPRC).*

missed positive and a scored negative. Two sources of predicted signs are compared. *Drift Jacobian signs* are $\mathrm{sign}(A_{ij})$ for the linear model and the sign of the cell-averaged signed Jacobian $\mathbb{E}_c[\partial g_{\theta,i}(X_c)/\partial x_j]$ for the neural models. *Partial-correlation signs* follow the convention of SINCERITIES [25]: signs are taken from partial correlations of the pooled expression data, here computed as ridge-regularized Pearson rather than Spearman partial correlations. These signs do not depend on the fitted model, so only the edge ranking differs across fits.

Inferring regulatory signs is considerably harder than detecting edges. Among the twelve methods benchmarked in BEELINE [26], only five output signed networks, and recovery of activating and inhibitory edges among top-ranked predictions was generally poor. BEELINE also found that a large fraction of false-positive edges corresponded to indirect, length-two paths in the reference network. The fitted drift Jacobian describes a local, instantaneous sensitivity, whereas an experimental knockdown response may include indirect effects through other genes. Their signs therefore need not coincide. This distinction may contribute to disagreement with the reference, but does not establish that the fitted signs are correct. We therefore evaluate signed recovery separately from unsigned edge ranking.

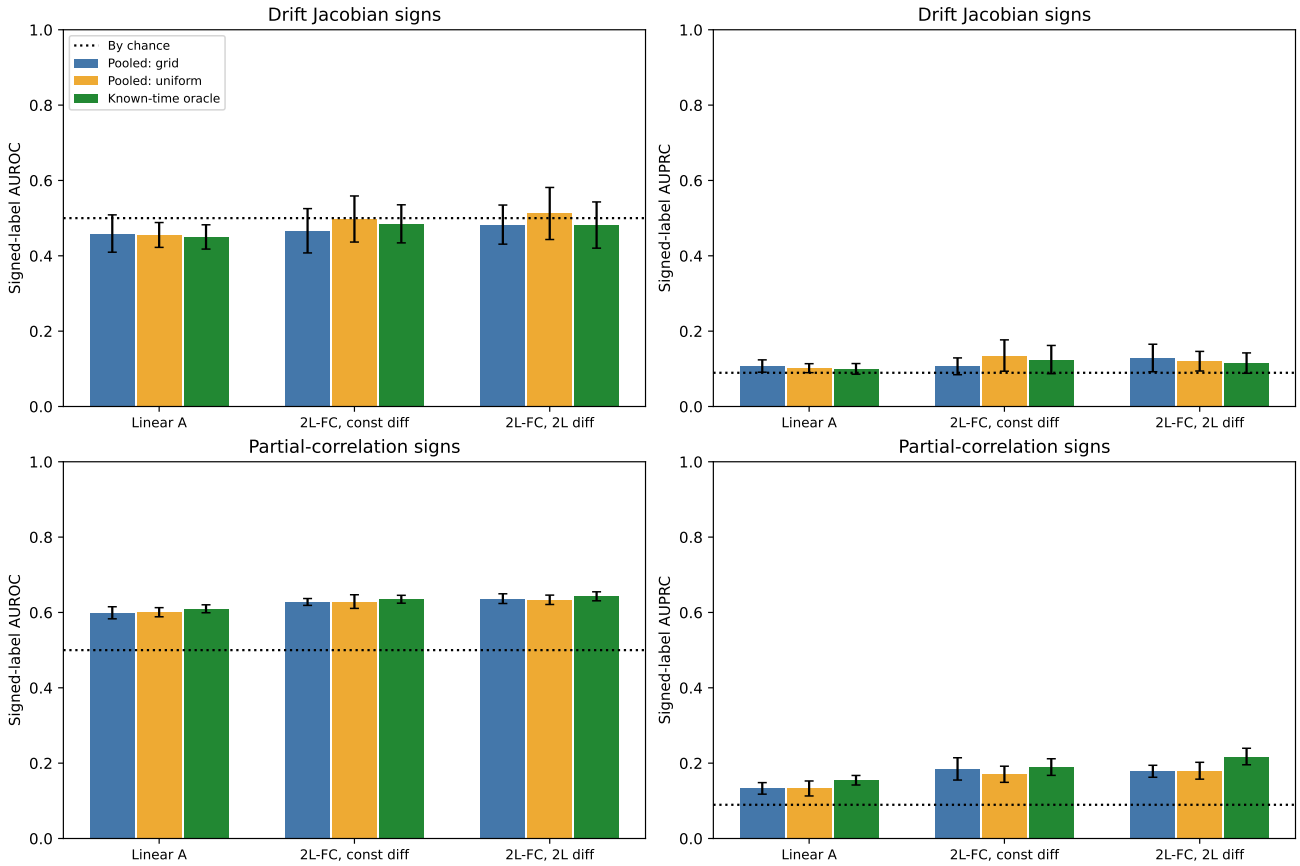


Figure 11: *Signed evaluation on the 20-gene BCL6 subnetwork. An edge is credited only if its sign matches the reference. Top: signs from the fitted drift. Bottom: signs from partial correlations of the pooled data. AUPRC is average precision. Mean ± SD over 10 seeds. Dotted line: chance (0.5 for AUROC; positive-label prevalence,* 68/760*, for AUPRC).*