GENERATING CELLULAR SPATIAL ORGANIZATION WITH DENOISING PAPANGELOU SCORE MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Biological tissues are organized not only by which cell types are present, but also by how they form cellular neighborhoods, compartments, and interfaces. Generative models of spatial tissues should therefore capture not only cell counts and composition, but also spatial organization. We introduce Denoising Papangelou Score Matching (DPSM), a framework for generating cellular spatial configurations consisting of cell locations and types. DPSM uses time dependent Papangelou fields to model where cells of different types are likely to be inserted given the current configuration. Learned along a corruption path to a Poisson reference, these fields determine the reverse generative processes. We show that noisy pseudolikelihood performs population field matching, derive a denoising identity for the corrupted fields, and relate field estimation error to reverse generation error. Experiments on controlled interacting point processes, synthetic cellular patterns, and real world spatial omics data show that DPSM recovers local cell type associations and spatial organization while preserving cell counts and composition.

1 INTRODUCTION

Biological tissues are organized not only by the cell types they contain, but also by how those cells form cellular neighborhoods, compartments, and interfaces. Spatial studies have linked cellular organization to disease state and progression in colorectal and breast cancer (Schürch et al., 2020; Risom et al., 2022). Even tissues with similar cell counts and type compositions can exhibit substantially different spatial arrangements. These observations motivate generative models that capture how cells are jointly arranged within tissues. Given a collection of observed tissue regions, our goal is to learn a distribution over complete cellular spatial configurations, including cell locations and types, from which new tissue arrangements can be sampled.

Existing work approaches spatial tissue modeling from several directions. Spatial omics simulators often focus on generating molecular measurements over observed or constructed spatial coordinates (Zhu et al., 2023; Song et al., 2024; Li et al., 2025). Classical spatial point processes model variation in point locations, counts, and spatial dependence (Møller et al., 1998; Baddeley & Turner, 2000). Neural set models learn distributions over unordered collections (Yang et al., 2020; Kim et al., 2021). More recent point-process diffusion and jump-based models allow configurations to change cardinality through point insertion and removal (Lüdke et al., 2023; Campbell et al., 2023; Lüdke et al., 2025; 2026). These developments provide useful foundations for cellular spatial generation, but a gap remains in how spatial dependencies between cell types and their surrounding environments can be represented and integrated into the generative process.

To incorporate such dependencies into generation, each local update should adapt to the current cellular configuration. A cell type that is common in one neighborhood may be rare in another, and each insertion or removal changes the context for subsequent updates. This motivates a configuration dependent field that characterizes the propensity to insert a cell of a given type at a particular location, conditional on the cells already present.

The Papangelou conditional intensity (Baddeley & Turner, 2000) provides a natural representation of this insertion propensity: given the current configuration, it describes how favorable it is to insert a cell of a particular type at a particular location. We consider a thinning–superposition path that gradually transforms the data distribution toward a Poisson reference. Each marginal along this

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

path has a time-dependent Papangelou field, reducing generation to learning these noisy fields from corrupted configurations.

Here we introduce Denoising Papangelou Score Matching (DPSM) to learn these time dependent Papangelou score fields through noisy negative pseudolikelihood. We show that this objective performs population field matching and derive a denoising identity linking the noisy fields to the clean insertion field. A single learned time dependent field determines both reverse insertion and removal rates, thereby specifying the reverse generative dynamics. We also show that field estimation error controls reverse generation error. Experiments on controlled point processes, synthetic cellular patterns, and real world heldout spatial omics data evaluate interaction recovery, configuration level organization, and generative fidelity. Our contributions are:

- **Configuration space generation.** We formulate generation of finite marked configurations through a time dependent Papangelou field that determines both reverse insertion and removal rates along a thinning–superposition path.
- **Learning Papangelou fields from corrupted data.** We show that noisy negative pseudolikelihood matches the Papangelou fields of corrupted distributions and derive a denoising identity relating these fields to the original insertion field.
- **Empirical recovery of spatial organization.** We show that DPSM recovers known statistical interactions on controlled point processes, coherent configuration level organization on synthetic cellular patterns, and multiple spatial statistics on real world held-out multi-conditional spatial omics data.

2 Related work

Spatial dependence and point-process models. Spatial point processes provide a statistical framework for modeling variation in point locations, counts, and dependence. Log Gaussian Cox processes represent spatial dependence through random intensity fields (Møller et al., 1998), while Papangelou conditional intensities describe insertion propensity conditional on the current configuration and can be estimated through pseudolikelihood (Baddeley & Turner, 2000). Neural models have also incorporated these ideas. SI-VAE learns a Papangelou-based spatial interaction model and uses the learned point-process model for conditional simulation (Sturm et al., 2026). CytoSpatio learns multirange, multitype point-process models of cell-type relationships and generates synthetic tissues by iteratively updating cell types on Poisson-sampled locations (Chen et al., 2025). DPSM builds on this statistical perspective but learns time-dependent Papangelou fields for corrupted data distributions, so that conditional spatial dependence directly parameterizes reverse-time generation.

Generative models for variable-cardinality configurations. Neural set models learn distributions over unordered collections using latent or energy-based representations, including EBP (Yang et al., 2020) and SetVAE (Kim et al., 2021). Other structured generators transport coordinates or empirical distributions, as in EquiFM and WFM (Song et al., 2023; Haviv et al., 2025). Cardinality-changing models instead allow points to appear or disappear during generation. ADD-THIN introduced thinning and Poisson superposition for temporal point processes (Lüdke et al., 2023), and PSD extended this construction to unordered point sets on general metric spaces (Lüdke et al., 2025). PSD approximates its reverse posterior by predicting which current points correspond to retained data points together with the number and locations of thinned data points. In contrast, DPSM parameterizes continuous-time reverse insertion and removal rates through a single time-dependent Papangelou field conditioned on the current configuration. Trans-dimensional and edit-based approaches provide alternative mechanisms for changing cardinality during generation (Campbell et al., 2023; Lüdke et al., 2026).

Score and generator learning. Score-based methods learn local information about a distribution that can be used to construct generative dynamics. For point-process data, Score-DKMPP applies score matching to gradients of the point-process log density with respect to event coordinates (Zhang et al., 2023), while SMASH learns scores for event times and locations together with mark probabilities (Li et al., 2024). In discrete diffusion, SEDD instead learns probability ratios that determine reverse transition rates (Lou et al., 2024). Under the Poisson-reference density representation used in DPSM, the normalized Papangelou field is an add-one density ratio on finite marked configurations.

Generator Matching shows that marginal generators can be expressed as posterior averages of conditional generators and learned using Bregman objectives (Holderrieth et al., 2025), with energy-based extensions to unnormalized targets (Woo et al., 2025). DPSM specializes these ideas to configuration space, where the GNZ identity turns Papangelou field matching into a noisy pseudolikelihood objective and the learned field directly specifies the reverse insertion–removal generator.

3 PRELIMINARIES

Let $W \subset \mathbb{R}^d$ be a bounded Borel window with positive Lebesgue measure and let $\mathcal{M} = \{1, \dots, K\}$ be a finite mark space. Write $E = W \times \mathcal{M}$ and $\mu(\mathrm{d}u) = \mathrm{d}x\,\#(\mathrm{d}m)$ for $u = (x, m)$, where $\#$ denotes counting measure. A configuration is a finite simple unordered set $X \subset E$, with cardinality $|X|$ and counting map $N_X(B) = |X \cap B|$ for Borel $B \subseteq E$. The configuration space $\mathcal{N}_f(E)$ carries the sigma-algebra generated by these counting maps. We denote the data law by P_0 and assume $\mathbb{E}_{P_0}|X| < \infty$.

3.1 FIRST-ORDER INTENSITY AND POISSON REFERENCE

When the mean counting measure of a law P is absolutely continuous with respect to μ, its first-order intensity ρ_P is defined by

$$\mathbb{E}_{X\sim P}[N_X(B)] = \int_B \rho_P(u)\,\mu(\mathrm{d}u), \qquad B \subseteq E \text{ Borel.} \tag{1}$$

It describes abundance averaged over realizations (Cronie et al., 2024).

For a nonnegative integrable function $\eta : E \to [0, \infty)$, let $\Pi_\eta = \mathrm{PPP}(\eta\mu)$ denote the Poisson point process law with intensity measure $\eta\mu$. Under $Z \sim \Pi_\eta$, counts in disjoint Borel sets are independent and

$$N_Z(B) \sim \mathrm{Poisson}\left(\int_B \eta(u)\,\mu(\mathrm{d}u)\right). \tag{2}$$

Because μ is diffuse and $\int_E \eta\,\mathrm{d}\mu < \infty$, Z is simple and finite almost surely (Decreusefond & Vasseur, 2018). We fix a measurable reference intensity $\nu : E \to (0, \infty)$ satisfying $0 < \bar{\nu} = \int_E \nu\,\mathrm{d}\mu < \infty$. The law Π_ν is both the reference for configuration densities and the Poisson noise reference used below.

3.2 PAPANGELOU CONDITIONAL INTENSITY

The Papangelou conditional intensity $\lambda_P(u; X)$ describes insertion propensity given the surrounding configuration; we call it an insertion field. Formally, a law P admits a Papangelou conditional intensity relative to μ if there exists a nonnegative measurable function $\lambda_P : E \times \mathcal{N}_f(E) \to [0, \infty)$ satisfying the Georgii–Nguyen–Zessin (GNZ) identity:

$$\mathbb{E}_P\left[\sum_{u\in X} h(u, X \setminus \{u\})\right] = \mathbb{E}_P\left[\int_E h(u, X)\lambda_P(u; X)\,\mu(\mathrm{d}u)\right] \tag{3}$$

for every nonnegative measurable h.

Insertion ratio representation. Suppose additionally that $P \ll \Pi_\nu$ and that its density $p_P = \mathrm{d}P/\mathrm{d}\Pi_\nu$ can be chosen so that removing points preserves positive density: $p_P(X) > 0$ implies $p_P(Y) > 0$ for every $Y \subseteq X$. Then the Papangelou intensity satisfies

$$r_P(u; X) := \frac{\lambda_P(u; X)}{\nu(u)} = \frac{p_P(X \cup \{u\})}{p_P(X)}, \qquad u \notin X, \quad p_P(X) > 0. \tag{4}$$

On configurations with $p_P(X) = 0$, we set the intensity to zero. The reference Π_ν yields the factor $\nu(u)$ (Ba & Coeurjolly, 2023). For the Poisson law, $\lambda_{\Pi_\nu}(u; X) = \nu(u)$ and $r_{\Pi_\nu}(u; X) = 1$. We write $\lambda_t = \lambda_{P_t}$ for the Papangelou field of a corrupted marginal. When the original data distribution P_0 also admits a Papangelou intensity, we denote it by $f = \lambda_{P_0}$. Here, "original" means before the imposed corruption.

3.3 THINNING AND SUPERPOSITION

For $a \in [0,1]$, the independent thinning of $X = \{u_1, \dots, u_N\}$ retains each point with probability a:

$$\mathrm{Thin}_a(X) = \{u_i \in X : B_i = 1\}, \qquad B_i \mid X \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(a). \tag{5}$$

Retained points keep their locations and marks (Lüdke et al., 2025).

Superposition $X \oplus Y$ adds the associated counting measures and agrees with set union for disjoint configurations. Poisson laws are closed under independent thinning and superposition:

$$\begin{aligned} Z \sim \Pi_\eta &\implies \mathrm{Thin}_a(Z) \sim \Pi_{a\eta}, \\ Z_1 \sim \Pi_{\eta_1},\ Z_2 \sim \Pi_{\eta_2},\ Z_1 \perp Z_2 &\implies Z_1 \oplus Z_2 \sim \Pi_{\eta_1+\eta_2}. \end{aligned} \tag{6}$$

Here the intensity functions are nonnegative and integrable (Decreusefond & Vasseur, 2018). These operations define the corruption process below.

Denoising score perspective. For $Z_0 \in \mathbb{R}^q$ and a corruption kernel with density $k_t(z \mid Z_0)$, let $q_t(z) = \mathbb{E}[k_t(z \mid Z_0)]$ be the corrupted marginal density. Assume that the conditional densities are positive and differentiable and that differentiation can be interchanged with expectation. Then

$$\nabla_z \log q_t(z) = \mathbb{E}[\nabla_z \log k_t(z \mid Z_0) \mid Z_t = z] \tag{7}$$

for q_t-almost every z. Denoising score matching learns the unavailable marginal score by regression onto the conditional score; squared-error regression recovers this conditional mean when the target has finite second moment and the model is realizable (Song et al., 2020). For marked configurations, DPSM uses the normalized insertion field $r_t(u; X) = \lambda_t(u; X)/\nu(u)$ as its score. The next section describes how to learn this score from corrupted configurations and use it for reverse generation.

4 DENOISING PAPANGELOU SCORE MATCHING

DPSM generates marked point configurations using a time-dependent Papangelou conditional intensity. We first corrupt data by independently removing points and inserting Poisson points. The conditional intensity of each corrupted distribution can be learned from corrupted samples using negative pseudolikelihood. A single learned intensity determines both insertion and removal rates in the reverse process. We develop the unconditional construction here; Appendix C gives the conditional formulation. Generation starts from the Poisson reference and follows the learned reverse rates; Algorithm 1 summarizes the practical workflow.

4.1 POISSON CORRUPTION AND CONDITIONAL INTENSITIES

Using the notation and Poisson reference of Section 3, assume

$$P_0 \ll \Pi_\nu, \qquad M_0 := \mathbb{E}_{P_0}|X_0| < \infty. \tag{8}$$

Consider the insertion and removal process with insertion intensity $\nu\,\mu$ and unit removal rate for each present point. Its transition at time $t \geq 0$ is

$$X_t = \mathrm{Thin}_{a_t}(X_0) \oplus Z_t, \qquad a_t = e^{-t}, \qquad Z_t \sim \Pi_{(1-a_t)\nu}, \tag{9}$$

with independent thinning and inserts. Write P_t for the resulting marginal law. The reference Π_ν is invariant under this transition.

For every $t > 0$, corruption gives mutual absolute continuity $P_t \sim \Pi_\nu$ and a Papangelou intensity λ_t satisfying

$$\lambda_t(u; X) \geq (1 - a_t)\nu(u) > 0 \qquad \text{for } \mu \otimes P_t\text{-almost every } (u, X). \tag{10}$$

Appendix A.1 proves these properties using the invariant Poisson law and the contribution from inserted points to the GNZ identity. They ensure that the conditional intensity is well defined and positive at every positive noise level under equation 8. Given a candidate point $u = (x, m)$, a corrupted configuration X_t, and a noise level t, the model $\lambda_\theta(u; X_t, t)$ predicts the conditional

intensity of inserting u. We normalize the true and predicted intensities by the reference intensity to obtain dimensionless scores:

$$r_t(u; X) = \frac{\lambda_t(u; X)}{\nu(u)}, \qquad r_\theta(u; X, t) = \frac{\lambda_\theta(u; X, t)}{\nu(u)}. \tag{11}$$

Thus λ_θ and r_θ express the same prediction on two scales. Energy differences and finite Poisson mixtures give structured examples in Appendix B.

4.2 Learning from corrupted configurations

Fix $0 < \varepsilon < T$ and a probability density $\pi(t) > 0$ almost everywhere on $[\varepsilon, T]$. For $a \geq 0$ and $b > 0$, define the Poisson divergence

$$D_{\mathrm{P}}(a\|b) = a \log(a/b) - a + b, \qquad 0 \log 0 = 0. \tag{12}$$

At each noise level, the population field error is

$$\mathcal{E}_t(\theta) = \mathbb{E}_{P_t} \int_E D_{\mathrm{P}}(\lambda_t(u; X)\|\lambda_\theta(u; X, t))\ \mu(\mathrm{d}u). \tag{13}$$

The objective $\mathcal{E}_\pi(\theta) = \int_\varepsilon^T \pi(t)\mathcal{E}_t(\theta)\,\mathrm{d}t$ matches insertion intensities across the corruption path.

GNZ representation. For $X_0 \sim P_0$, sample $t \sim \pi$ independently and form X_t by equation 9. The noisy negative pseudolikelihood (NPL) is

$$\begin{aligned}\mathcal{L}_{\mathrm{NPL}}(\theta) = \mathbb{E}_{t,X_t}\Big[&\int_E \lambda_\theta(u; X_t, t)\,\mu(\mathrm{d}u) \\ &- \sum_{z \in X_t} \log \lambda_\theta(z; X_t \setminus \{z\}, t)\Big].\end{aligned} \tag{14}$$

The integrated predicted intensity penalizes assigning high intensity everywhere. The pointwise term encourages high intensity at observed points. This is the pseudolikelihood construction applied to the noisy marginals (Ba & Coeurjolly, 2023).

Theorem 1 (Population field matching)**.** *Under equation 8, assume that both terms of the NPL objective have finite expectations. Specifically, for both the model intensity $g_t(u; X) = \lambda_\theta(u; X, t)$ and the true intensity $g_t(u; X) = \lambda_t(u; X)$, require*

$$\mathbb{E}_{t,X_t}\left[\int_E g_t(u; X_t)\,\mu(\mathrm{d}u)\right] < \infty,$$

$$\mathbb{E}_{t,X_t}\left[\sum_{z \in X_t} |\log g_t(z; X_t \setminus \{z\})|\right] < \infty,$$

where the expectations are over $t \sim \pi$ and $X_t \sim P_t$. Then

$$\mathcal{L}_{\mathrm{NPL}}(\theta) - \mathcal{L}_{\mathrm{NPL}}(\lambda) = \mathcal{E}_\pi(\theta). \tag{15}$$

Here, $\mathcal{L}_{\mathrm{NPL}}(\lambda)$ is the NPL objective evaluated at the true intensities λ_t. The unrestricted minimum is attained at λ_t. Within any fixed model family, minimizing NPL is equivalent to minimizing $\mathcal{E}_\pi$.

Appendix A.3 gives the details and the corresponding construction for general Bregman divergences.

Denoising representation. The NPL objective above does not require access to the Papangelou intensity of the data distribution. To relate this objective to denoising regression, suppose additionally that P_0 admits a Papangelou intensity f satisfying the GNZ identity for P_0. Here, f describes the original data distribution before corruption. For a data configuration X_0 and noise level t, define the denoising target

$$T_t(u; X_0) = (1 - a_t) + a_t \frac{f(u; X_0)}{\nu(u)}. \tag{16}$$

Let $U \sim \nu\,\mu/\bar{\nu}$ be independent of the data configuration X_0, the sampled time, and the corruption randomness. The associated denoising risk is

$$\mathcal{L}_{\mathrm{DPSM}}(\theta) = \mathbb{E}[D_{\mathrm{P}}(T_t(U; X_0)\|r_\theta(U; X_t, t))]\,. \tag{17}$$

Theorem 2 (Denoising identity and risk equivalence)**.** *Under the preceding assumptions on P_0 and f, for each $t > 0$,*

$$\lambda_t(u; X_t) = (1 - a_t)\nu(u) + a_t\, \mathbb{E}[f(u; X_0) \mid X_t] \tag{18}$$

holds $\mu \otimes P_t$-almost everywhere. If the NPL risks in Theorem 1 are finite and the target satisfies $\mathbb{E}[T_t \log_+ T_t] < \infty$ under the query and corruption law, then

$$\mathcal{L}_{\mathrm{DPSM}}(\theta) = \frac{1}{\bar{\nu}}\mathcal{L}_{\mathrm{NPL}}(\theta) + C, \tag{19}$$

where $\log_+ x = \max\{\log x, 0\}$ and C is finite and independent of the candidate field.

Thus the two objectives have the same minimizers in any fixed model family. When f can be evaluated, the denoising representation provides an explicit regression target; NPL can be evaluated without computing f. Appendices A.2 and A.4 give the proofs and analyze the effect of replacing f by an estimate $\widehat{f}$.

4.3 Reverse generation

Let τ denote elapsed reverse time and set $t = T - \tau$. Reversing the probability fluxes of forward insertions and removals gives

$$\begin{aligned} b_t(u; X) &= \lambda_t(u; X), \\ d_t(z; X) &= \frac{\nu(z)}{\lambda_t(z; X \setminus \{z\})}, \qquad z \in X. \end{aligned} \tag{20}$$

The exact reverse process started from P_T has marginal P_t at elapsed time $T - t$. Replacing λ_t by λ_θ gives the learned insertion intensity $b_{\theta,t}(u; X) = \nu(u) r_\theta(u; X, t)$ and removal rate $d_{\theta,t}(z; X) = 1/r_\theta(z; X \setminus \{z\}, t)$. A larger score therefore favors inserting a point and discourages removing it from the same surrounding configuration. For a test function H on configuration space, the learned process uses the generator

$$\begin{aligned} (\mathcal{A}^{\leftarrow}_{\theta,t} H)(X) = &\int_E \lambda_\theta(u; X, t)[H(X \cup \{u\}) - H(X)]\, \mu(\mathrm{d}u) \\ &+ \sum_{z \in X} \frac{\nu(z)}{\lambda_\theta(z; X \setminus \{z\}, t)}[H(X \setminus \{z\}) - H(X)]. \end{aligned} \tag{21}$$

Thus one positive field specifies both inserts and removals. Appendix A.5 derives the reverse rates.

Connection to the learning objective. For generation, we initialize the learned reverse process from Π_ν and run from noise level T down to ε. Under bounded positive insertion intensities, Theorem 5 bounds the total variation error relative to P_0 by three contributions: Poisson initialization, residual corruption, and a term controlled by the time-integrated intensity estimation error. The full statement and proof are given in Appendix A.6.

5 Results

We evaluate DPSM through four complementary tests. We first test whether DPSM recovers known interactions and the spatial statistics of the resulting configurations. Synthetic cellular patterns isolate configuration-level mark organization from marginal geometry and mark frequencies. Patient-held-out spatial omics datasets test whether generated windows reproduce spatial statistics beyond the training patients. And comparisons with direct MCMC under the same fitted energy examine the role of the learned noise path. Model architectures and baseline settings are collected in Appendix E.

5.1 Recovering interactions and spatial statistics

We use unmarked and three-type Strauss processes with repulsion, and a four-type Gibbs process with attraction and repulsion. These processes provide ground truth pair interactions across all mark pairs. Each method generates 512 configurations per dataset and seed (Appendix F).

Interaction recovery. On three-type Strauss, the pair terms in the fitted $t = 0$ energy distinguish all six mark pairs and approach zero near distance 0.1 (mean per-pair RMSE 0.060) (Figure 1), recovering both the strength and range of type-specific repulsion. Appendix F.3 reports recovery and heldout insertion intensity probes for all three processes.

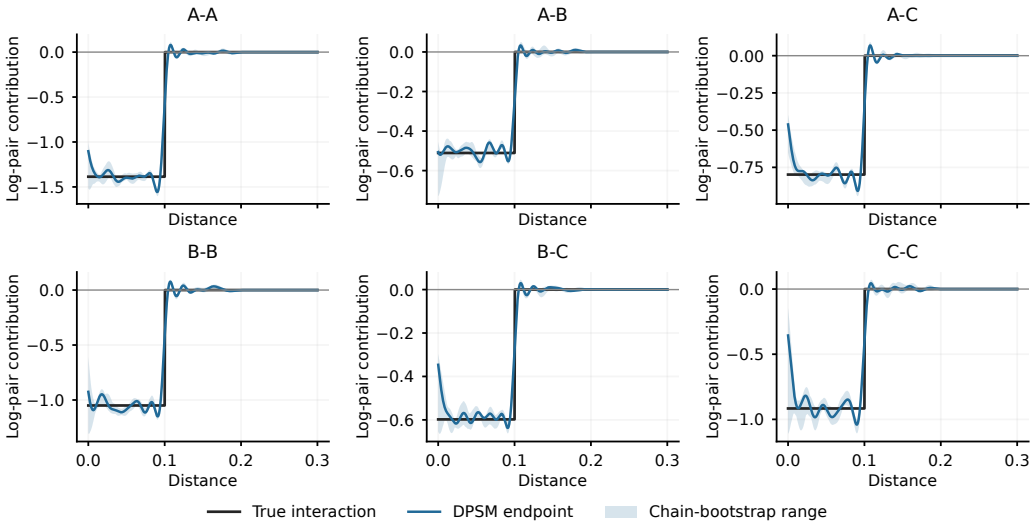

Figure 1: Three-type Strauss pair log density contributions: truth (black) and fitted energy (blue). Shading gives exploratory pointwise ranges from 12 chain bootstrap refits with model selection fixed.

Spatial generation. DPSM achieves the lowest mean errors for the reported spatial metrics on all three datasets (Table 1).

Table 1: Spatial generation errors (lower is better): mean ± sample SD over three seeds, sharing the clean endpoint for DPSM. Bold: lowest mean. Each power of ten applies to the mean and SD. Full results and definitions: Appendix F.4 and Appendix F.2.

Method	Scaled NN W_1	Neighbor mark RMSE	Spectral error	Moran RMSE
Unmarked Strauss Dataset				
DPSM	$\mathbf{6.38_{\pm 1.71} \times 10^{-3}}$	—	$\mathbf{1.066_{\pm 0.284} \times 10^{-2}}$	$\mathbf{3.834_{\pm 1.842} \times 10^{-3}}$
PSD	$9.617_{\pm 0.298} \times 10^{-2}$	—	$1.4844_{\pm 0.0355} \times 10^{-1}$	$6.673_{\pm 0.088} \times 10^{-2}$
EBP	$1.2356_{\pm 0.2229} \times 10^{-1}$	—	$2.2492_{\pm 0.5529} \times 10^{-1}$	$1.246_{\pm 0.519} \times 10^{-1}$
SI-VAE	$2.945_{\pm 0.247} \times 10^{-2}$	—	$4.123_{\pm 1.080} \times 10^{-2}$	$1.097_{\pm 0.473} \times 10^{-2}$
SetVAE	$1.259_{\pm 0.015} \times 10^{-1}$	—	$1.6522_{\pm 0.0177} \times 10^{-1}$	$7.678_{\pm 0.094} \times 10^{-2}$
Three-type Strauss Dataset				
DPSM	$\mathbf{3.82_{\pm 2.53} \times 10^{-3}}$	$\mathbf{1.889_{\pm 0.537} \times 10^{-2}}$	$\mathbf{1.128_{\pm 0.161} \times 10^{-2}}$	$\mathbf{4.040_{\pm 1.286} \times 10^{-3}}$
PSD	$7.716_{\pm 0.211} \times 10^{-2}$	$6.188_{\pm 0.196} \times 10^{-2}$	$1.4043_{\pm 0.0314} \times 10^{-1}$	$3.006_{\pm 0.202} \times 10^{-2}$
EBP	$1.312_{\pm 0.665} \times 10^{-1}$	$2.0588_{\pm 1.1000} \times 10^{-1}$	$3.6429_{\pm 2.8878} \times 10^{-1}$	$7.033_{\pm 5.456} \times 10^{-2}$
SI-VAE	$2.908_{\pm 0.112} \times 10^{-2}$	$2.201_{\pm 0.200} \times 10^{-2}$	$5.427_{\pm 0.210} \times 10^{-2}$	$1.049_{\pm 0.134} \times 10^{-2}$
SetVAE	$1.4210_{\pm 0.0610} \times 10^{-1}$	$1.1586_{\pm 0.1228} \times 10^{-1}$	$2.0191_{\pm 0.1100} \times 10^{-1}$	$4.197_{\pm 0.704} \times 10^{-2}$
Four-type Gibbs Dataset				
DPSM	$\mathbf{3.26_{\pm 0.16} \times 10^{-3}}$	$\mathbf{1.539_{\pm 0.484} \times 10^{-2}}$	$\mathbf{1.386_{\pm 0.316} \times 10^{-2}}$	$\mathbf{3.881_{\pm 1.859} \times 10^{-3}}$
PSD	$4.573_{\pm 0.236} \times 10^{-2}$	$3.185_{\pm 0.323} \times 10^{-2}$	$8.404_{\pm 0.152} \times 10^{-2}$	$1.760_{\pm 0.116} \times 10^{-2}$
EBP	$1.051_{\pm 0.224} \times 10^{-1}$	$1.379_{\pm 0.957} \times 10^{-1}$	$3.9133_{\pm 1.7751} \times 10^{-1}$	$1.991_{\pm 1.093} \times 10^{-2}$
SI-VAE	$1.297_{\pm 0.212} \times 10^{-2}$	$2.101_{\pm 1.350} \times 10^{-2}$	$3.694_{\pm 1.174} \times 10^{-2}$	$8.634_{\pm 1.676} \times 10^{-3}$
SetVAE	$1.2335_{\pm 0.1522} \times 10^{-1}$	$1.2602_{\pm 0.7842} \times 10^{-1}$	$2.1846_{\pm 0.1434} \times 10^{-1}$	$9.894_{\pm 0.928} \times 10^{-3}$

5.2 GENERATING ORGANIZATION BEYOND MARGINAL INTENSITY

We train each conditional model jointly on six datasets: three with position dependent geometry and marks, and three with hidden random partitions assignments. The latter have equal foreground mark probabilities at each location but coherent organization within configurations. Models receive only the dataset label (Appendix G).

Geometry and mark organization. On annular domains, DPSM and PSD have the same reported mean geometric precision (0.972), but organization scores of 0.896 and 0.152, respectively (Table 10). The lower row of Figure 2 illustrates this distinction: both reproduce the annular outline, while DPSM forms coherent type domains and PSD mixes the foreground marks. Table 2 reports geometry and mark alignment averaged over the three position dependent datasets, and organization separately for each relational dataset.

Table 2: Geometry and mark organization: mean ± sample SD over three independent fits, each generating 32 samples per dataset. Precision, coverage, and alignment average over three position-dependent datasets; organization scores are reported separately for the three relational datasets. Higher is better for all metrics; bold marks the best model mean.

	Geometry and marks			Mark organization ↑		
Method	Precision ↑	Coverage ↑	Alignment ↑	Annular	Elliptic	Paired
Reference	0.949	0.950	0.894	0.948	0.950	0.955
DPSM	$\mathbf{0.940}_{\pm 0.007}$	$\mathbf{0.952}_{\pm 0.002}$	$\mathbf{0.896}_{\pm 0.004}$	$\mathbf{0.896}_{\pm 0.024}$	$\mathbf{0.915}_{\pm 0.012}$	$\mathbf{0.856}_{\pm 0.129}$
PSD	$0.817_{\pm 0.022}$	$0.929_{\pm 0.004}$	$0.764_{\pm 0.012}$	$0.152_{\pm 0.161}$	$0.163_{\pm 0.159}$	$0.079_{\pm 0.015}$
EBP	$0.323_{\pm 0.022}$	$0.488_{\pm 0.023}$	$0.510_{\pm 0.008}$	$0.488_{\pm 0.312}$	$0.412_{\pm 0.265}$	$0.297_{\pm 0.064}$
SI-VAE	$0.436_{\pm 0.022}$	$0.631_{\pm 0.040}$	$0.530_{\pm 0.018}$	$0.106_{\pm 0.014}$	$0.100_{\pm 0.006}$	$0.290_{\pm 0.039}$
SetVAE	$0.927_{\pm 0.018}$	$0.787_{\pm 0.030}$	$0.892_{\pm 0.016}$	$0.377_{\pm 0.057}$	$0.386_{\pm 0.084}$	$0.112_{\pm 0.019}$

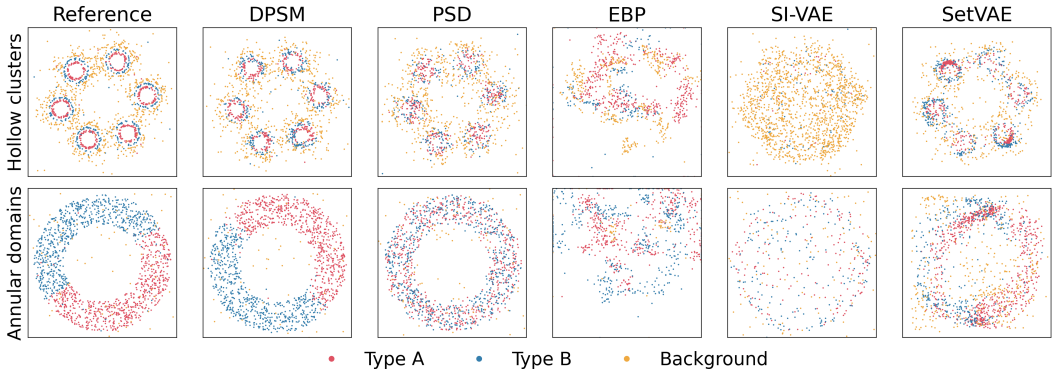

Figure 2: Hollow cluster geometry (top) and annular mark organization (bottom). Colors distinguish two foreground marks and background. Prespecified rows show the first saved sample of the first fit; partition orientations vary between configurations.

Interaction ablation. Disabling configuration interactions in the fitted generator, with spatial conditions, initial noise, reverse seed, and sampling budget fixed, reduces the annular organization score from 0.896 to 0.084, with similar reductions on the other relational datasets (Appendix G.4). This ablation supports the contribution of interactions to coherent mark organization in the generator.

5.3 PRESERVING SPATIAL ORGANIZATION IN HELD-OUT BIOLOGICAL DATA

We evaluate three multiplexed imaging datasets: breast tissue MIBI-TOF (DCIS; Risom et al., 2022), brain tumor IMC (Karimi et al., 2023), and colorectal cancer CODEX (CRC; Schürch et al., 2020). Each window is a square tissue region represented as a marked point configuration of cell centroids and aggregated cell types. We split each dataset into training, validation, and test sets by patient, keeping all windows from the same patient in a single split (Appendix H.1).

Spatial generation. We compare DPSM with PSD, SI-VAE, and CytoSpatio over three seeds, generating 512 windows per condition and seed. Table 3 reports the metrics for cell spacing, neighboring types, and density fluctuations. DPSM has the lowest condition averaged mean error on all five metrics in DCIS and CRC and on four of five metrics in IMC, where SI-VAE has the lowest mean enrichment RMSE. Complete condition level results appear in Appendix H.2. Taken together, these results support the recovery of several spatial statistics beyond the training patients, including cell spacing, type associations, and density fluctuations.

Table 3: Real world data generation (lower is better). Values report mean with sample SD over three seeds. Each seed generates 512 windows per condition; dataset scores average conditions equally. Bold marks the lowest mean.

Method	Scaled NN W_1	Neighbor mark RMSE	Spectral error	Enrichment RMSE	Moran RMSE
DCIS Dataset					
DPSM	$\mathbf{0.07399_{\pm 0.00057}}$	$\mathbf{0.09868_{\pm 0.00383}}$	$\mathbf{0.30020_{\pm 0.01366}}$	$\mathbf{0.33587_{\pm 0.00600}}$	$\mathbf{0.08014_{\pm 0.00180}}$
PSD	$0.10700_{\pm 0.00060}$	$0.17246_{\pm 0.02362}$	$0.48283_{\pm 0.01351}$	$0.73153_{\pm 0.00096}$	$0.22886_{\pm 0.00080}$
SI-VAE	$0.11282_{\pm 0.00316}$	$0.19120_{\pm 0.01634}$	$0.49272_{\pm 0.13747}$	$0.56570_{\pm 0.01044}$	$0.19058_{\pm 0.00610}$
CytoSpatio	$0.10631_{\pm 0.00011}$	$0.15078_{\pm 0.00022}$	$0.49389_{\pm 0.00076}$	$0.73906_{\pm 0.00091}$	$0.22751_{\pm 0.00034}$
IMC Dataset					
DPSM	$\mathbf{0.00928_{\pm 0.00184}}$	$\mathbf{0.07868_{\pm 0.00411}}$	$\mathbf{0.09842_{\pm 0.00347}}$	$0.14247_{\pm 0.01496}$	$\mathbf{0.03880_{\pm 0.00131}}$
PSD	$0.11415_{\pm 0.00048}$	$0.16385_{\pm 0.00306}$	$0.25457_{\pm 0.00159}$	$0.24863_{\pm 0.00697}$	$0.07370_{\pm 0.00068}$
SI-VAE	$0.06816_{\pm 0.01061}$	$0.13578_{\pm 0.00526}$	$0.16480_{\pm 0.01165}$	$\mathbf{0.13943_{\pm 0.00973}}$	$0.05609_{\pm 0.00343}$
CytoSpatio	$0.11424_{\pm 0.00136}$	$0.14523_{\pm 0.00010}$	$0.25969_{\pm 0.00259}$	$0.25657_{\pm 0.00136}$	$0.07322_{\pm 0.00089}$
CRC Dataset					
DPSM	$\mathbf{0.02556_{\pm 0.00309}}$	$\mathbf{0.10555_{\pm 0.00310}}$	$\mathbf{0.21622_{\pm 0.01190}}$	$\mathbf{0.10610_{\pm 0.01172}}$	$\mathbf{0.04512_{\pm 0.00606}}$
PSD	$0.07610_{\pm 0.00017}$	$0.12676_{\pm 0.00736}$	$0.39894_{\pm 0.00427}$	$0.19210_{\pm 0.00318}$	$0.10909_{\pm 0.00124}$
SI-VAE	$0.06427_{\pm 0.00301}$	$0.15012_{\pm 0.01545}$	$0.36111_{\pm 0.01473}$	$0.13446_{\pm 0.00201}$	$0.09527_{\pm 0.00028}$
CytoSpatio	$0.07496_{\pm 0.00030}$	$0.14132_{\pm 0.00034}$	$0.40380_{\pm 0.00292}$	$0.18772_{\pm 0.00259}$	$0.10894_{\pm 0.00022}$

5.4 Sampling with a learned noise path versus direct MCMC

We compare generation along the learned noise path with direct MCMC using the same learned energy. In the separate structural comparison, using 32 configurations per condition, annular organization scores are 0.9217 for DPSM, 0.2507 for Local MCMC, and 0.2390 for Collective MCMC (Table 17).

Population allocation. Coherent organization within individual configurations does not ensure that different modes occur with the correct frequencies. We therefore evaluate population allocation by comparing the proportions of generated configurations in different modes with their reference proportions. Table 18 reports these comparisons together with organization within configurations. Collective MCMC improves mode proportion estimates over Local MCMC in the Poisson mixture, collective patches, and learned curved lobes, but retains substantial discrepancies in others. For example, its mode proportion error is lower than DPSM's in the Poisson mixture (0.0066 versus 0.0164), whereas its two-boundary fraction error in the controlled annulus is much higher (0.6141 versus 0.0266). In the separate three-class mixture, DPSM better recovers the class proportions: the total variation error is 0.00547 versus 0.355 for Collective MCMC (Table 19).

6 Discussion

Our results highlight the importance of spatial dependence for generating cellular spatial organization. In controlled settings, DPSM recovers known statistical interactions and, on synthetic cellular patterns with similar marginal spatial structure, coherent spatial organization; real world heldout spatial omics experiments further show that DPSM reproduces multiple spatial statistics beyond the training patients. Generation performance, however, remains dependent on the expressivity of the learned insertion fields and the accuracy of reverse time sampling. Improving both remains an important direction for future work. Overall, DPSM connects statistical modeling of spatial dependence with generative simulation of complete cellular configurations, making spatial organization a direct object of generative modeling rather than only a statistic measured after generation.

GENERATIVE AI USAGE DISCLOSURE

In this work, we used generative AI tools to assist with translation and language polishing, and as a secondary tool for post hoc checking of the correctness and technical consistency of mathematical arguments and algorithms developed by the authors. We used generative AI tools to assist with code implementation. Generative AI tools were not used to develop the theoretical framework, formulate the main mathematical claims or results, provide critical ingredients for proofs, construct or write the proofs, or generate the scientific conclusions of the paper. All AI-assisted code, checks and suggestions were independently evaluated and verified by the authors. All final scientific and editorial decisions were made by the authors, who take full responsibility for the final content of this work.

REFERENCES

Ismaïla Ba and Jean-François Coeurjolly. Inference for low- and high-dimensional inhomogeneous Gibbs point processes. *Scandinavian Journal of Statistics*, 50(3):993–1021, 2023. ISSN 1467-9469. doi: 10.1111/sjos.12616.

Adrian Baddeley and Rolf Turner. Practical Maximum Pseudolikelihood for Spatial Point Patterns: (with Discussion). *Australian & New Zealand Journal of Statistics*, 42(3):283–322, September 2000. ISSN 1369-1473, 1467-842X. doi: 10.1111/1467-842X.00128.

Andrew Campbell, William Harvey, Christian Weilbach, Valentin De Bortoli, Thomas Rainforth, and Arnaud Doucet. Trans-Dimensional Generative Modeling via Jump Diffusion Models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 42217–42257. Curran Associates, Inc., 2023. doi: 10.52202/075280-1830.

Haoran Chen, Yangyuan Zhang, and Robert F. Murphy. CytoSpatio: Learning cell type spatial relationships using multirange, multitype point process models. *PLOS Computational Biology*, 21(8):e1013409, 2025. doi: 10.1371/journal.pcbi.1013409.

Ottmar Cronie, Mehdi Moradi, and Christophe A N Biscio. A cross-validation-based statistical theory for point processes. *Biometrika*, 111(2):625–641, 2024. ISSN 1464-3510. doi: 10.1093/biomet/asad041.

Laurent Decreusefond and Aurélien Vasseur. Stein's method and Papangelou intensity for Poisson or Cox process approximation, 2018. arXiv:1807.02453.

Doron Haviv, Aram-Alexandre Pooladian, Dana Pe'Er, and Brandon Amos. Wasserstein Flow Matching: Generative Modeling Over Families of Distributions. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 22238–22258. PMLR, 2025.

Peter Holderrieth, Marton Havasi, Jason Yim, Neta Shaul, Itai Gat, Tommi Jaakkola, Brian Karrer, Ricky T. Q. Chen, and Yaron Lipman. Generator Matching: Generative modeling with arbitrary Markov processes. In *International Conference on Learning Representations*, volume 2025, pp. 52153–52219, 2025.

Elham Karimi, Miranda W. Yu, Sarah M. Maritan, et al. Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature*, 614:555–563, 2023. doi: 10.1038/s41586-022-05680-3.

Jinwoo Kim, Jaehoon Yoo, Juho Lee, and Seunghoon Hong. SetVAE: Learning Hierarchical Composition for Generative Modeling of Set-Structured Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15059–15068, 2021.

Hechen Li, Ziqi Zhang, Michael Squires, Xi Chen, and Xiuwei Zhang. scMultiSim: simulation of single-cell multi-omics and spatial data guided by gene regulatory networks and cell–cell interactions. *Nature Methods*, 22(5):982–993, May 2025. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-025-02651-0.

Zichong Li, Qunzhi Xu, Zhenghao Xu, Yajun Mei, Tuo Zhao, and Hongyuan Zha. Beyond Point Prediction: Score Matching-based Pseudolikelihood Estimation of Neural Marked Spatio-Temporal Point Process. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 29096–29111. PMLR, 2024.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete Diffusion Modeling by Estimating the Ratios of the Data Distribution. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32819–32848. PMLR, 2024.

David Lüdke, Marin Biloš, Oleksandr Shchur, Marten Lienen, and Stephan Günnemann. Add and Thin: Diffusion for Temporal Point Processes. In *Advances in Neural Information Processing Systems*, volume 36, pp. 56784–56801. Curran Associates, Inc., 2023. doi: 10.52202/075280-2480.

David Lüdke, Enric Rabasseda Raventós, Marcel Kollovieh, and Stephan Günnemann. Unlocking Point Processes through Point Set Diffusion. In *International Conference on Learning Representations*, volume 2025, pp. 82436–82454, 2025.

David Lüdke, Marten Lienen, Marcel Kollovieh, and Stephan Günnemann. Edit-Based Flow Matching for Temporal Point Processes. In *International Conference on Learning Representations*, volume 2026, pp. 111844–111864, 2026.

Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. Log Gaussian Cox Processes. *Scandinavian Journal of Statistics*, 25(3):451–482, 1998. ISSN 1467-9469. doi: 10.1111/1467-9469.00115.

Tyler Risom, David R. Glass, Inna Averbukh, et al. Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell*, 185(2):299–310.e18, 2022. doi: 10.1016/j.cell.2021.12.023.

Christian M. Schürch, Salil S. Bhate, Graham L. Barlow, et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*, 182(5):1341–1359.e19, 2020. doi: 10.1016/j.cell.2020.07.005.

Dongyuan Song, Qingyang Wang, Guanao Yan, Tianyang Liu, Tianyi Sun, and Jingyi Jessica Li. scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics. *Nature Biotechnology*, 42(2):247–252, February 2024. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-023-01772-1.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Yang Song, Jascha Narain Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv preprint arXiv:2011.13456*, 2020.

Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou, and Wei-Ying Ma. Equivariant Flow Matching with Hybrid Probability Transport for 3D Molecule Generation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 549–568. Curran Associates, Inc., 2023. doi: 10.52202/075280-0026.

Dominik Sturm, Hiba Bensalem, and Ivo Sbalzarini. Spatially Informed Autoencoders for Interpretable Visual Representation Learning. In *International Conference on Learning Representations*, volume 2026, pp. 18200–18234, 2026.

Dongyeop Woo, Minsu Kim, Minkyu Kim, Kiyoung Seong, and Sungsoo Ahn. Energy-based generator matching: A neural sampler for general state space. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 110247–110289. Curran Associates, Inc., 2025. doi: 10.52202/085713-3683.

Mengjiao Yang, Bo Dai, Hanjun Dai, and Dale Schuurmans. Energy-Based Processes for Exchangeable Data. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10681–10692. PMLR, 2020.

Yixuan Zhang, Quyu Kong, and Feng Zhou. Integration-free Training for Spatio-temporal Multimodal Covariate Deep Kernel Point Processes. In *Advances in Neural Information Processing Systems*, volume 36, pp. 25031–25049. Curran Associates, Inc., 2023. doi: 10.52202/075280-1088.

Jiaqiang Zhu, Lulu Shang, and Xiang Zhou. SRTsim: spatial pattern preserving simulations for spatially resolved transcriptomics. *Genome Biology*, 24(1):39, March 2023. ISSN 1474-760X. doi: 10.1186/s13059-023-02879-z.

A PAPANGELOU RISK IDENTITIES AND REVERSE GENERATION GUARANTEES

Throughout this appendix, E is a standard Borel marked space equipped with a diffuse σ-finite measure μ. The reference intensity satisfies $0 < \nu(u) < \infty$ for μ-almost every u and $0 < \bar{\nu} = \int_E \nu \, \mathrm{d}\mu < \infty$. We write Π_ν for its Poisson law on finite simple configurations and $a_t = e^{-t}$. Population objectives averaged over time use a common density $\pi(t) > 0$ almost everywhere on $[\varepsilon, T]$, where $0 < \varepsilon < T < \infty$. All insertion fields have jointly measurable versions.

A.1 POSITIVE-TIME REGULARIZATION

Proposition 3 (Positive-time Papangelou fields)**.** *Suppose $P_0 \ll \Pi_\nu$ and $\mathbb{E}_{P_0}|X_0| < \infty$. For every $t > 0$, the law of $X_t = \mathrm{Thin}_{a_t}(X_0) \oplus Z_t$ with independent $Z_t \sim \Pi_{(1-a_t)\nu}$ satisfies $P_t \sim \Pi_\nu$. It admits a Papangelou intensity*

$$\lambda_t(u; X) = \nu(u)\frac{p_t(X \cup \{u\})}{p_t(X)} \geq (1 - a_t)\nu(u) \quad \textit{for } \mu \otimes P_t\textit{-almost every } (u, X), \tag{22}$$

where $p_t = \mathrm{d}P_t/\mathrm{d}\Pi_\nu$ is positive and finite almost everywhere. Moreover,

$$\mathbb{E}_{P_t} \int_E \lambda_t(u; X) \, \mu(\mathrm{d}u) = \mathbb{E}|X_t| = a_t \mathbb{E}|X_0| + (1 - a_t)\bar{\nu}. \tag{23}$$

Proof. Let K_t be the thinning and insertion kernel. Poisson thinning and superposition give $\Pi_\nu K_t = \Pi_\nu$. If $\Pi_\nu(A) = 0$, then $\int K_t(X, A)\Pi_\nu(\mathrm{d}X) = 0$. Consequently $K_t(X, A) = 0$ for Π_ν-almost every X and hence for P_0-almost every X. This proves $P_t = P_0 K_t \ll \Pi_\nu$.

The probability that thinning removes every original point is $\gamma_t = \mathbb{E}[(1 - a_t)^{|X_0|}] > 0$. This event is independent of Z_t, so $P_t \geq \gamma_t \Pi_{(1-a_t)\nu}$ as measures. The Poisson density formula gives

$$\frac{\mathrm{d}\Pi_{(1-a_t)\nu}}{\mathrm{d}\Pi_\nu}(X) = e^{a_t \bar{\nu}}(1 - a_t)^{|X|} > 0. \tag{24}$$

Thus $\Pi_\nu \ll P_t$, and p_t has a positive finite version almost everywhere.

For nonnegative measurable h, Poisson Mecke gives

$$\begin{aligned}\mathbb{E}_{P_t} \sum_{z \in X} h(z, X \setminus \{z\}) &= \mathbb{E}_{\Pi_\nu} \int_E \nu(u) p_t(X \cup \{u\}) h(u, X) \, \mu(\mathrm{d}u) \\ &= \mathbb{E}_{P_t} \int_E \lambda_t(u; X) h(u, X) \, \mu(\mathrm{d}u).\end{aligned}$$

The density ratio therefore satisfies GNZ. Mecke also ensures that exceptional configurations in the density version produce a $\mu \otimes P_t$-null set of insertion queries.

For the lower bound, decompose the reduced point sum into contributions from retained and inserted points. Conditioning on the retained configuration, the contribution from inserted points is

$$\mathbb{E} \int_E (1 - a_t)\nu(u) h(u, X_t) \, \mu(\mathrm{d}u).$$

The retained contribution is nonnegative. Comparison with GNZ proves $\lambda_t \geq (1 - a_t)\nu$ almost everywhere. Finally, thinning and Poisson count expectations give $\mathbb{E}|X_t| = a_t \mathbb{E}|X_0| + (1 - a_t)\bar{\nu}$; GNZ with $h = 1$ gives the integrated-intensity identity. □

The proposition applies to every original data density relative to the finite Poisson reference, including densities supported on prescribed cardinality ranges. The denoising representation below additionally assumes that the original data law admits a Papangelou intensity satisfying GNZ.

A.2 Proof of the Papangelou denoising identity

Assume that the original data distribution has finite mean count and a nonnegative Papangelou field f satisfying GNZ. For $a \in [0,1]$ and an integrable nonnegative intensity η, consider

$$Y = \mathrm{Thin}_a(X_0) \oplus Z_\eta, \qquad Z_\eta \sim \Pi_\eta, \tag{25}$$

with independent thinning and noise. The general identity is

$$\lambda_Y(u;Y) = \eta(u) + a\,\mathbb{E}[f(u;X_0) \mid Y]. \tag{26}$$

A regular conditional law $K(\mathrm{d}\xi \mid Y)$ exists on the standard Borel configuration space. We use the jointly measurable conditional mean $m_a(u,Y) = \int f(u;\xi)K(\mathrm{d}\xi \mid Y)$. GNZ for the original data law gives $\mathbb{E}\int_E f(u;X_0)\,\mu(\mathrm{d}u) = \mathbb{E}|X_0| < \infty$, so m_a is finite $\mu \otimes P_Y$-almost everywhere.

Proof of the identity in Theorem 2. For nonnegative measurable F, Tonelli gives

$$\mathbb{E}\int_E F(u,\omega)\,\mu(\mathrm{d}u) = \int_E \mathbb{E}[F(u,\omega)]\,\mu(\mathrm{d}u). \tag{27}$$

Fix a nonnegative GNZ test function h and split the reduced sum over Y into contributions from retained and inserted points, $I_{\mathrm{ret}} + I_{\mathrm{ins}}$. For deterministic ξ, define

$$H_a(u,\xi) = \mathbb{E}[h(u, \mathrm{Thin}_a(\xi) \oplus Z_\eta)],$$

using fresh independent corruption. Conditioning on retention of each original point and applying GNZ for the original data law yields

$$\begin{aligned} I_{\mathrm{ret}} &= a\,\mathbb{E}\sum_{u\in X_0} H_a(u, X_0 \setminus \{u\}) \\ &= a\,\mathbb{E}\int_E H_a(u,X_0) f(u;X_0)\,\mu(\mathrm{d}u) \\ &= a\,\mathbb{E}\int_E h(u,Y) f(u;X_0)\,\mu(\mathrm{d}u). \end{aligned} \tag{28}$$

The last equality follows from $H_a(u,X_0) = \mathbb{E}[h(u,Y) \mid X_0]$ and the tower property. Conditioning on the retained configuration and applying Poisson Mecke gives

$$I_{\mathrm{ins}} = \mathbb{E}\int_E h(u,Y)\eta(u)\,\mu(\mathrm{d}u). \tag{29}$$

Combining the terms and conditioning on Y gives

$$\begin{aligned} \mathbb{E}\sum_{u\in Y} h(u, Y\setminus\{u\}) &= \mathbb{E}\int_E h(u,Y)\{\eta(u) + a f(u;X_0)\}\,\mu(\mathrm{d}u) \\ &= \mathbb{E}\int_E h(u,Y)\{\eta(u) + a m_a(u,Y)\}\,\mu(\mathrm{d}u). \end{aligned} \tag{30}$$

The factor $h(u,Y)$ is measurable with respect to Y, and Tonelli justifies the integrated conditional expectation identities. The last expression is GNZ with the field in equation 26. Taking $a = a_t$ and $\eta = (1-a_t)\nu$ proves equation 18. □

The argument gives a marked-space proof of the thinning and superposition construction associated with Papangelou intensities (Decreusefond & Vasseur, 2018).

A.3 GNZ risks and Bregman field matching

For fixed $t \in [\varepsilon, T]$, write $\mathcal{R}_t(\theta)$ for the raw population NPL in equation 14 conditional on t. Write $\mathcal{R}_t(\lambda_t)$ for its value at the true field. Assume finite expected integrated intensities and absolutely integrable logarithmic point sums for the candidate and true fields. By GNZ, the latter condition is

$$\mathbb{E}_{P_t}\int_E \lambda_t(u;X)|\log g(u;X)|\,\mu(\mathrm{d}u) < \infty, \qquad g \in \{\lambda_\theta(\cdot;\cdot,t), \lambda_t\}.$$

Proof of Theorem 1. Applying GNZ to the positive and negative parts of the logarithm gives

$$\mathcal{R}_t(\theta) = \mathbb{E}_{P_t} \int_E [\lambda_\theta(u; X, t) - \lambda_t(u; X) \log \lambda_\theta(u; X, t)] \, \mu(\mathrm{d}u). \tag{31}$$

Subtracting the expression at λ_t gives

$$\begin{aligned} \mathcal{R}_t(\theta) - \mathcal{R}_t(\lambda_t) &= \mathbb{E}_{P_t} \int_E D_{\mathrm{P}}(\lambda_t(u; X) \| \lambda_\theta(u; X, t)) \, \mu(\mathrm{d}u) \\ &= \mathcal{E}_t(\theta). \end{aligned} \tag{32}$$

The scalar divergence is nonnegative and vanishes exactly when its arguments agree. The assumed time-integrated absolute integrability makes this identity valid for π-almost every t and permits integration:

$$\mathcal{L}_{\mathrm{NPL}}(\theta) - \mathcal{L}_{\mathrm{NPL}}(\lambda) = \int_\varepsilon^T \pi(t) \mathcal{E}_t(\theta) \, \mathrm{d}t = \mathcal{E}_\pi(\theta).$$

Here λ denotes the true time-indexed field. Every candidate family thus minimizes its integrated field divergence. A family containing λ attains zero excess risk; equality identifies the field almost everywhere under $\pi(t) \, \mathrm{d}t \, P_t(\mathrm{d}X) \, \mu(\mathrm{d}u)$. □

Let $\Phi : (0, \infty) \to \mathbb{R}$ be differentiable and strictly convex, and define $D_\Phi(x \| y) = \Phi(x) - \Phi(y) - \Phi'(y)(x - y)$. For a positive candidate field g, set

$$\begin{aligned} \mathcal{R}_{\Phi,t}(g) = \mathbb{E}_{P_t} \int_E [g(u; X) \Phi'(g(u; X)) - \Phi(g(u; X))] \, \mu(\mathrm{d}u) \\ - \mathbb{E}_{P_t} \sum_{z \in X} \Phi'(g(z; X \setminus \{z\})). \end{aligned} \tag{33}$$

Assume that the displayed integral and sum are absolutely integrable for both g and λ_t, with

$$\mathbb{E}_{P_t} \int_E [\lambda_t |\Phi'(g)| + |\Phi(\lambda_t)|] \, \mathrm{d}\mu < \infty.$$

Proposition 4 (Bregman representations of a Papangelou risk)**.** *Under these conditions,*

$$\mathcal{R}_{\Phi,t}(g) - \mathcal{R}_{\Phi,t}(\lambda_t) = \mathbb{E}_{P_t} \int_E D_\Phi(\lambda_t \| g) \, \mathrm{d}\mu. \tag{34}$$

If a positive random target $\tau_t(u)$ on an augmented probability space satisfies $\mathbb{E}[\tau_t(u) \mid X_t] = \lambda_t(u; X_t)$, and the additional terms below are integrable, then

$$\mathbb{E} \int_E D_\Phi(\tau_t \| g) \, \mathrm{d}\mu = \mathcal{R}_{\Phi,t}(g) + \mathbb{E} \int_E \Phi(\tau_t) \, \mathrm{d}\mu. \tag{35}$$

Proof. Applying GNZ gives

$$\mathcal{R}_{\Phi,t}(g) = \mathbb{E}_{P_t} \int_E [g\Phi'(g) - \Phi(g) - \lambda_t \Phi'(g)] \, \mathrm{d}\mu.$$

At the true field, this reduces to $\mathcal{R}_{\Phi,t}(\lambda_t) = -\mathbb{E}_{P_t} \int_E \Phi(\lambda_t) \, \mathrm{d}\mu$. Subtraction proves equation 34. Expanding the divergence between the denoising target and candidate field and applying $\mathbb{E}[\tau_t \Phi'(g) \mid X_t] = \lambda_t \Phi'(g)$ proves equation 35. □

The choice $\Phi(x) = x \log x - x$ gives NPL exactly. The choice $\Phi(x) = x^2/2$ gives $\mathbb{E} \int_E g^2/2 \, \mathrm{d}\mu - \mathbb{E} \sum_{z \in X} g(z; X \setminus \{z\})$. Strict convexity identifies the population field, and the parametrization determines the attainable field class.

A measurable nonnegative weight $w(u, X, t)$ can be included throughout these identities. The integral then carries $w(u, X, t)$, and the reduced sum carries $w(z, X \setminus \{z\}, t)$. Positive weights identify the field on their support and specify its projection within a restricted candidate class.

A.4 Denoising regression, estimated intensities, and common projections

The target and prediction in equations 16–17 use an independent query $U \sim \nu \, \mathrm{d}\mu/\bar{\nu}$. Write $C = (U, X_t, t)$ and $m(C) = \mathbb{E}[T_t(U; X_0) \mid C]$. For positive predictions $r(C)$, the conditional Bregman decomposition is

$$\mathbb{E}[D_{\mathrm{P}}(T_t \| r) \mid C] = \mathbb{E}[D_{\mathrm{P}}(T_t \| m) \mid C] + D_{\mathrm{P}}(m \| r). \tag{36}$$

It follows by expansion and $\mathbb{E}[T_t - m \mid C] = 0$. The target is positive on $[\varepsilon, T]$, so the conditional minimizer is m.

Proof of the risk identity in Theorem 2. The denoising identity gives $m = r_t(U; X_t)$. The entropy condition on T_t makes $C_{\mathrm{cond}} = \mathbb{E}[D_{\mathrm{P}}(T_t \| m)]$ finite: both T_t and m have finite first moment, their negative $x \log x$ parts are bounded, and conditional Jensen bounds $\mathbb{E}[m \log_+ m]$ by $\mathbb{E}[T_t \log_+ T_t]$. Poisson divergence is homogeneous: $D_{\mathrm{P}}(\nu x \| \nu y) = \nu D_{\mathrm{P}}(x \| y)$. Query averaging and equation 36 therefore give

$$\begin{aligned} \mathcal{L}_{\mathrm{DPSM}}(\theta) &= C_{\mathrm{cond}} + \frac{1}{\bar{\nu}} \int_{\varepsilon}^{T} \pi(t) \mathcal{E}_t(\theta) \, \mathrm{d}t \\ &= \frac{1}{\bar{\nu}} \mathcal{L}_{\mathrm{NPL}}(\theta) + C, \end{aligned} \tag{37}$$

where $C = C_{\mathrm{cond}} - \mathcal{L}_{\mathrm{NPL}}(\lambda)/\bar{\nu}$ is finite and independent of θ. The finite NPL excess risk also establishes finiteness of the regression risk. □

For any common candidate class $\mathcal{F}$, this identity yields

$$\begin{aligned} \operatorname*{arg\,min}_{g \in \mathcal{F}} \mathcal{L}_{\mathrm{DPSM}}(g) &= \operatorname*{arg\,min}_{g \in \mathcal{F}} \mathcal{L}_{\mathrm{NPL}}(g) \\ &= \operatorname*{arg\,min}_{g \in \mathcal{F}} \int_{\varepsilon}^{T} \pi(t) \mathbb{E}_{P_t} \int_{E} D_{\mathrm{P}}(\lambda_t \| g_t) \, \mathrm{d}\mu \, \mathrm{d}t. \end{aligned} \tag{38}$$

This statement covers restricted classes through their best attainable field approximation. The common time law, query measure, and scale specify the projection.

For a fixed intensity estimate $\widehat{f} \geq 0$ with $\mathbb{E}_{P_0} \int_E \widehat{f} \, \mathrm{d}\mu < \infty$, put $\widehat{T}_t = (1 - a_t) + a_t \widehat{f}/\nu$ and assume the analogous entropy and finite-risk conditions. The conditional decomposition gives

$$\begin{aligned} \widehat{r}^*(u; X_t, t) &= \mathbb{E}[\widehat{T}_t(u; X_0) \mid X_t, t], \\ \nu(u) \widehat{r}^*(u; X_t, t) - \lambda_t(u; X_t) &= a_t \, \mathbb{E}[\widehat{f}(u; X_0) - f(u; X_0) \mid X_t, t]. \end{aligned} \tag{39}$$

The conditional expectations use the original data law P_0 and its corruption coupling.

A matched combination of denoising regression using an estimated intensity and NPL has an explicit target. Let $\widehat{\mathcal{L}}_{\mathrm{DPSM}}$ denote regression with $\widehat{T}_t$ and $\beta \in [0, 1]$. Define

$$\mathcal{J}_\beta(\theta) = \beta \widehat{\mathcal{L}}_{\mathrm{DPSM}}(\theta) + (1 - \beta) \frac{\mathcal{L}_{\mathrm{NPL}}(\theta)}{\bar{\nu}}.$$

Expanding the terms linear in $\log r_\theta$ gives

$$\begin{aligned} \lambda_{\beta,t} &= \lambda_t + \beta a_t \mathbb{E}[\widehat{f} - f \mid X_t, t], \\ \mathcal{J}_\beta(\theta) &= C_\beta + \frac{1}{\bar{\nu}} \int_{\varepsilon}^{T} \pi(t) \mathbb{E}_{P_t} \int_{E} D_{\mathrm{P}}(\lambda_{\beta,t} \| \lambda_\theta) \, \mathrm{d}\mu \, \mathrm{d}t. \end{aligned} \tag{40}$$

Here C_β is independent of θ. The unrestricted minimizer is $\lambda_{\beta,t}$; a restricted class projects that target in the displayed divergence.

Whenever the right-hand side is finite, the target bias obeys

$$\mathbb{E}_{P_t} \int_{E} D_{\mathrm{P}}(\lambda_t \| \lambda_{\beta,t}) \, \mathrm{d}\mu \leq \beta a_t \, \mathbb{E}_{P_0} \int_{E} D_{\mathrm{P}}(f \| \widehat{f}) \, \mathrm{d}\mu. \tag{41}$$

The divergence on nonnegative original insertion fields uses its lower semicontinuous extension: $D_{\mathrm{P}}(0\|y) = y$ and $D_{\mathrm{P}}(x\|0) = +\infty$ for $x > 0$. To prove the bound, put $f_\beta = (1-\beta)f + \beta\widehat{f}$. Joint convexity, conditional Jensen, and homogeneity give

$$\begin{aligned} D_{\mathrm{P}}(\lambda_t\|\lambda_{\beta,t}) &\le a_t D_{\mathrm{P}}(\mathbb{E}[f \mid X_t, t]\|\mathbb{E}[f_\beta \mid X_t, t]) \\ &\le a_t \mathbb{E}[D_{\mathrm{P}}(f\|f_\beta) \mid X_t, t] \\ &\le \beta a_t \mathbb{E}[D_{\mathrm{P}}(f\|\widehat{f}) \mid X_t, t]. \end{aligned}$$

Integration proves equation 41.

A.5 Exact reversal and finite-horizon coupling

The forward insertion and removal process has insertion intensity $\nu(u)$ and unit removal rate for each point. Its jump count on any finite interval is finite almost surely: inserts form a finite-rate Poisson process, and each removal acts on an initial or subsequently inserted point. Proposition 3 supplies positive Papangelou fields on $[\varepsilon, T]$.

At forward time t, GNZ equates the forward removal flux with an insertion flux of intensity λ_t:

$$\mathbb{E}_{P_t} \sum_{z\in X} h(z, X\setminus\{z\}) = \mathbb{E}_{P_t} \int_E h(u, X)\lambda_t(u; X)\,\mu(\mathrm{d}u).$$

Applying GNZ to $h(z, X\setminus\{z\})\nu(z)/\lambda_t(z; X\setminus\{z\})$ gives

$$\mathbb{E}_{P_t} \sum_{z\in X} h(z, X\setminus\{z\}) \frac{\nu(z)}{\lambda_t(z; X\setminus\{z\})} = \mathbb{E}_{P_t} \int_E h(u, X)\nu(u)\,\mu(\mathrm{d}u).$$

These identities identify the reversed jump fluxes. In elapsed reverse time $s = T - t$, the generator is

$$\begin{aligned} (\mathcal{A}^{\leftarrow}_s F)(X) = &\int_E \lambda_{T-s}(u; X)[F(X\cup\{u\}) - F(X)]\,\mu(\mathrm{d}u) \\ &+ \sum_{z\in X} \frac{\nu(z)}{\lambda_{T-s}(z; X\setminus\{z\})}[F(X\setminus\{z\}) - F(X)]. \end{aligned} \tag{42}$$

For completeness, regular conditional distributions define the reverse transition kernels by

$$P_t(\mathrm{d}X)\,K^{\leftarrow}_{t,t-h}(X, \mathrm{d}Y) = P_{t-h}(\mathrm{d}Y)\,K_h(Y, \mathrm{d}X), \qquad \varepsilon \le t-h < t \le T.$$

The forward Markov property and disintegration give their Chapman–Kolmogorov identity. Applying the forward jump compensator to this reversed joint law identifies the instantaneous reverse fluxes with the two GNZ expressions above, and hence gives equation 42 for almost every time. The finite jump count of the forward path gives the finite jump count of its reversal. Starting from P_T, the reverse process has marginal P_{T-s} and ends at P_ε. This proves the rate formula in equation 20.

We use $d_{\mathrm{TV}}(P, Q) = \sup_A |P(A) - Q(A)|$. The initialization discrepancy obeys

$$d_{\mathrm{TV}}(P_T, \Pi_\nu) \le \min\{1, a_T(\mathbb{E}|X_0| + \bar{\nu})\}. \tag{43}$$

Draw $C_T = \mathrm{Thin}_{a_T}(X_0)$ and independent $Z_T \sim \Pi_{(1-a_T)\nu}$ and $Z'_T \sim \Pi_{a_T\nu}$. Then $X_T = C_T \oplus Z_T$ and $\widetilde{X}_T = Z'_T \oplus Z_T$ have marginals P_T and Π_ν. They agree whenever C_T and Z'_T are both empty, so

$$\begin{aligned} d_{\mathrm{TV}}(P_T, \Pi_\nu) &\le \Pr(X_T \ne \widetilde{X}_T) \\ &\le \mathbb{E}|C_T| + \mathbb{E}|Z'_T| = a_T(\mathbb{E}|X_0| + \bar{\nu}). \end{aligned} \tag{44}$$

Coupling X_ε directly to X_0 gives the residual corruption bound

$$d_{\mathrm{TV}}(P_\varepsilon, P_0) \le \min\{1, (1-a_\varepsilon)(\mathbb{E}|X_0| + \bar{\nu})\}. \tag{45}$$

The configurations agree whenever every original point is retained and the configuration of inserted points is empty. Their discrepancy probability is bounded by the expected number of removed points plus the expected number of inserted points. This also proves $P_\varepsilon \to P_0$ in total variation as $\varepsilon \downarrow 0$.

A.6 FIELD ERROR AND CONTINUOUS-TIME REVERSE GENERATION ERROR

The following result gives explicit sufficient conditions for controlling reverse generation error through the population field error. A positive candidate field determines inserts and removals through the rates in equation 42. The assumptions concern these rates as functions on configuration space.

Theorem 5 (From field error to reverse generation error)**.** *Assume the hypotheses of Proposition 3. Fix $0 < \varepsilon < T < \infty$ and a positive measurable field $\lambda_\theta(u; X, t)$. Suppose versions of the true and candidate fields satisfy, throughout the interval and configuration space,*

$$c\nu(u) \le \lambda_t(u; X), \lambda_\theta(u; X, t) \le C_+\nu(u), \qquad 0 < c \le 1 - e^{-\varepsilon}, \quad C_+ < \infty, \tag{46}$$

for μ-almost every u. Let $\mathbb{P}^{\leftarrow}$ be the true reverse path law and $\mathbb{Q}_\theta^{P_T}$ the candidate reverse path law with the same initial distribution P_T. Define

$$J_{\varepsilon,T}(\theta) = \left(1 + \frac{1}{c}\right) \int_\varepsilon^T \mathcal{E}_t(\theta)\, \mathrm{d}t. \tag{47}$$

Both reverse processes are nonexplosive, and

$$\mathrm{KL}(\mathbb{P}^{\leftarrow} \| \mathbb{Q}_\theta^{P_T}) \le J_{\varepsilon,T}(\theta). \tag{48}$$

For the candidate terminal law $Q_{\theta,\varepsilon}^{\Pi_\nu}$ initialized from Π_ν,

$$\begin{aligned} d_{\mathrm{TV}}(Q_{\theta,\varepsilon}^{\Pi_\nu}, P_0) \le \min\Big\{1, [e^{-T} + 1 - e^{-\varepsilon}](\mathbb{E}|X_0| + \bar\nu) \\ + \sqrt{J_{\varepsilon,T}(\theta)/2}\Big\}. \end{aligned} \tag{49}$$

Proof. Write $H = T - \varepsilon$ for the elapsed reverse horizon.

Existence and integrability. For either field, the total insertion rate is at most $C_+\bar\nu$, and the removal rate of each present point is at most $1/c$. Inserts can be constructed by thinning a Poisson proposal process of rate $C_+\bar\nu$. Each removal acts on an initial or subsequently inserted point. If B_H counts insertion proposals and $X_0^{\leftarrow}$ is the initial configuration, then

$$\sup_{0 \le s \le H} |X_s| \le |X_0^{\leftarrow}| + B_H, \qquad N_{\mathrm{jumps}} \le |X_0^{\leftarrow}| + 2B_H.$$

The initial laws P_T and Π_ν have finite mean count. These bounds give nonexplosion, finite expected jump count, and finite expected integrated total jump rate.

Path likelihood and relative entropy. Let $q_s(X, \mathrm{d}X')$ and $\widehat{q}_s(X, \mathrm{d}X')$ denote the true and candidate jump kernels. Their insertion rate and removal rate ratios lie between c/C_+ and C_+/c. Conditional on a common initial state, products of jump intensities and holding time tail probabilities define mutually absolutely continuous densities on finite-jump trajectories. Let Λ_s and $\widehat{\Lambda}_s$ be their total rates evaluated along the path. The log likelihood ratio is

$$\log \frac{\mathrm{d}\mathbb{P}^{\leftarrow}}{\mathrm{d}\mathbb{Q}_\theta^{P_T}} = \sum_{s:\, \Delta X_s \neq 0} \log \frac{\mathrm{d}q_s(X_{s-}, \cdot)}{\mathrm{d}\widehat{q}_s(X_{s-}, \cdot)}(X_s) - \int_0^H (\Lambda_s - \widehat{\Lambda}_s)\, \mathrm{d}s. \tag{50}$$

The jump logarithms are bounded in absolute value by $\log(C_+/c)$, and the count bounds above make both terms integrable. Taking expectation under $\mathbb{P}^{\leftarrow}$ and using the jump compensator yields

$$\begin{aligned} &\mathrm{KL}(\mathbb{P}^{\leftarrow} \| \mathbb{Q}_\theta^{P_T}) \\ &\quad = \int_\varepsilon^T \left\{ \mathcal{E}_t(\theta) + \mathbb{E}_{P_t} \sum_{z \in X} D_{\mathrm{P}}\left(\frac{\nu(z)}{\lambda_t(z; X \setminus \{z\})} \,\middle\|\, \frac{\nu(z)}{\lambda_\theta(z; X \setminus \{z\}, t)} \right) \right\} \mathrm{d}t. \end{aligned} \tag{51}$$

The true reverse marginal at forward time t is P_t, which specifies the expectation in this formula.

Comparison with the field error. Apply GNZ to the nonnegative removal summand. Writing $\ell = \lambda_t(u; X)$ and $b = \lambda_\theta(u; X, t)$, its integrated scalar integrand becomes

$$\ell D_{\mathrm{P}}(\nu/\ell \| \nu/b) = \nu[\log(b/\ell) - 1 + \ell/b]. \tag{52}$$

For fixed b and ν, put $k(x) = \nu[\log(b/x) - 1 + x/b]$ and $h(x) = D_{\mathrm{P}}(x\|b)$. Both functions and their first derivatives vanish at $x = b$, and

$$k''(x) = \frac{\nu}{x^2}, \qquad h''(x) = \frac{1}{x}.$$

Every point on the segment joining ℓ and b satisfies $x \geq c\nu$. Hence $k''(x) \leq h''(x)/c$ along the segment. Taylor's integral formula on either side of b gives

$$\nu[\log(b/\ell) - 1 + \ell/b] \leq \frac{1}{c} D_{\mathrm{P}}(\ell\|b). \tag{53}$$

The removal contribution is at most $\mathcal{E}_t(\theta)/c$. The insertion contribution equals $\mathcal{E}_t(\theta)$, proving equation 48.

Terminal comparison and initialization. Let $Q_{\theta,\varepsilon}^{P_T}$ be the candidate terminal law from P_T. The endpoint map and data processing give $\mathrm{KL}(P_\varepsilon\|Q_{\theta,\varepsilon}^{P_T}) \leq J_{\varepsilon,T}(\theta)$. Pinsker then gives

$$d_{\mathrm{TV}}(P_\varepsilon, Q_{\theta,\varepsilon}^{P_T}) \leq \sqrt{J_{\varepsilon,T}(\theta)/2}.$$

For the common candidate reverse transition kernel $\widehat{K}_{\varepsilon,T}$,

$$d_{\mathrm{TV}}(P_T\widehat{K}_{\varepsilon,T}, \Pi_\nu\widehat{K}_{\varepsilon,T}) \leq d_{\mathrm{TV}}(P_T, \Pi_\nu).$$

The triangle inequality, the initialization coupling, and equation 45 prove equation 49. □

To express the result using the training time density, define

$$K_\pi = \operatorname*{ess\,sup}_{t\in[\varepsilon,T]} \frac{1 + c^{-1}}{\pi(t)}.$$

Whenever $K_\pi < \infty$,

$$J_{\varepsilon,T}(\theta) \leq K_\pi \mathcal{E}_\pi(\theta). \tag{54}$$

For a risk normalized by $\bar{\nu}$, the same statement is $J_{\varepsilon,T} \leq K_\pi\bar{\nu}(\mathcal{E}_\pi/\bar{\nu})$. The bound thus uses the measure and scale of the population field matching objective.

B Insertion field parameterizations

B.1 Energy-induced fields and compatibility

Let $E = W \times \{1, \ldots, K\}$ and let Π_ν be the finite Poisson reference from Section 4. A finite, measurable, permutation-invariant set energy E_ϕ induces a positive insertion field

$$\begin{aligned} f_\phi(u; X) &= \nu(u) \exp[E_\phi(X) - E_\phi(X \cup \{u\})], \\ Z_\phi &= \mathbb{E}_{\Pi_\nu} \exp[-E_\phi(X)], \\ p_\phi(X) &= Z_\phi^{-1} \exp[-E_\phi(X)] \quad \text{when } 0 < Z_\phi < \infty. \end{aligned} \tag{55}$$

Under the stated condition on Z_ϕ, p_ϕ is a density relative to Π_ν and f_ϕ is its Papangelou intensity. The same density ratio formula extends to energies with hereditary support, on configurations of positive density. For distinct insertion points $u, v \notin X$, energy differences telescope:

$$f_\phi(u; X) f_\phi(v; X \cup \{u\}) = f_\phi(v; X) f_\phi(u; X \cup \{v\}). \tag{56}$$

This cocycle identity makes the product of insertion ratios independent of insertion order. An explicit sufficient condition for a finite normalizer is $E_\phi(X) \geq -B|X| - C$ for finite $B, C \geq 0$, since

$$Z_\phi \leq e^C \mathbb{E}_{\Pi_\nu} e^{B|X|} = \exp\left[C + \bar{\nu}(e^B - 1)\right].$$

Energy compatibility and this integrability condition specify the two ingredients of the associated density construction.

An energy U_ϕ expressed relative to the unit-intensity Poisson reference can be written relative to Π_ν as

$$E_\phi(X) = U_\phi(X) + \sum_{u \in X} \log \nu(u), \qquad f_\phi(u; X) = \exp[U_\phi(X) - U_\phi(X \cup \{u\})]. \tag{57}$$

The reference-dependent additive constant is absorbed into the normalizer. For example, a symmetric radial pair family has

$$\begin{aligned} U_\phi(X) &= -\sum_i c_{m_i} - \sum_{i<j} \psi_{m_i m_j}(\|x_i - x_j\|), \\ \psi_{ab}(r) &= \sum_{k=1}^{K_\rho} w_{kab} \rho_k(r), \qquad w_{kab} = w_{kba}, \end{aligned} \tag{58}$$

where the ρ_k are measurable radial functions. General permutation-invariant set energies extend this construction to spatially varying and higher-order interactions.

Time-dependent energies induce fields along the corruption path through the same insertion rule. For a spatially homogeneous reference, a radial example is

$$\log \lambda_\theta((x, m); X, t) = c_m(a_t) + \sum_{(y,b) \in X} \psi_{mb,a_t}(\|x - y\|), \qquad \psi_{mb,a} = \psi_{bm,a}. \tag{59}$$

A general endpoint-constrained construction uses a clean energy $\widehat{E}$, a set energy H_θ, and a scalar function η with $\eta(0) = \eta(1) = 0$:

$$E_{\theta,a}(X) = a\widehat{E}(X) + \eta(a) H_\theta(X, a).$$

Its insertion field equals the reference ν at $a = 0$ and the original insertion field induced by $\widehat{E}$ at $a = 1$. The resulting candidate family can be fitted through equation 14. Theorem 5 gives process-level conditions under which field error controls reverse generation error.

Original insertion field estimation risk. When the original data law admits a Papangelou intensity f satisfying GNZ, negative pseudolikelihood (NPL) provides a population criterion for a positive candidate f_ϕ:

$$\mathcal{L}_{\mathrm{PL}}(\phi) = \mathbb{E}_{P_0}\left[\int_E f_\phi(u; X)\, \mu(\mathrm{d}u) - \sum_{u \in X} \log f_\phi(u; X \setminus \{u\})\right]. \tag{60}$$

Its empirical counterpart averages the configuration functional over observed configurations (Ba & Coeurjolly, 2023). Under finite-risk assumptions, GNZ gives

$$\mathcal{L}_{\mathrm{PL}}(\phi) - \mathcal{L}_{\mathrm{PL}}(f) = \mathbb{E}_{P_0} \int_E D_{\mathrm{P}}(f(u; X) \| f_\phi(u; X))\, \mu(\mathrm{d}u).$$

This characterizes the population target of original insertion field estimation.

B.2 Finite Poisson mixtures and exact closure

A finite Poisson mixture provides an explicit family closed under thinning and Poisson inserts. Fix a descriptor $G = g$ and a positive reference intensity ν_g with finite integral $\bar{\nu}_g$. Let $\pi_\ell(g) > 0$, $\sum_{\ell=1}^L \pi_\ell(g) = 1$, and let the component intensities $f_\ell(u \mid g) > 0$ have finite integrals. Write

$$\begin{aligned} P_0(\mathrm{d}X \mid g) &= \sum_{\ell=1}^L \pi_\ell(g) \Pi_{f_\ell(\cdot \mid g)}(\mathrm{d}X), \\ f_\ell((x, m) \mid g) &= \rho_{\ell,g}(x) q_{\ell,g}(m \mid x), \qquad \sum_m q_{\ell,g}(m \mid x) = 1, \\ \Lambda_{\ell,g} &= \int_W \rho_{\ell,g}(x)\, \mathrm{d}x, \qquad \bar{f}_g(u) = \sum_\ell \pi_\ell(g) f_\ell(u \mid g). \end{aligned} \tag{61}$$

Component total intensities may vary with ℓ. Thinning with retention $a_t = e^{-t}$ and inserts with intensity $(1 - a_t)\nu_g$ give

$$\begin{aligned} \lambda_{\ell,t}(u \mid g) &= a_t f_\ell(u \mid g) + (1 - a_t)\nu_g(u), \\ \Lambda_{\ell,t,g} &= a_t \Lambda_{\ell,g} + (1 - a_t)\bar{\nu}_g. \end{aligned} \tag{62}$$

The component densities, mixture density, and posterior component weights are

$$\begin{aligned} D_{\ell,t}(X \mid g) &= \exp(\bar{\nu}_g - \Lambda_{\ell,t,g}) \prod_{u \in X} \frac{\lambda_{\ell,t}(u \mid g)}{\nu_g(u)}, \\ p_{t,g}(X) &= \sum_\ell \pi_\ell(g) D_{\ell,t}(X \mid g), \\ w_{\ell,t}(X, g) &= \frac{\pi_\ell(g) D_{\ell,t}(X \mid g)}{p_{t,g}(X)}. \end{aligned} \tag{63}$$

Every $D_{\ell,t}$ integrates to one relative to Π_{ν_g}, so $p_{t,g}$ is a normalized density. Taking its insertion ratio yields the exact Papangelou field

$$\widehat{\lambda}_t(u; X, g) = \nu_g(u) \frac{p_{t,g}(X \cup \{u\})}{p_{t,g}(X)} = \sum_\ell w_{\ell,t}(X, g) \lambda_{\ell,t}(u \mid g). \tag{64}$$

The hat here denotes the field of this explicit mixture family. Its dependence on X enters through the posterior weights, which incorporate the component-specific integrated intensities in equation 63.

The prior-weighted field is

$$\lambda_t^{\text{off}}(u \mid g) = \sum_\ell \pi_\ell(g) \lambda_{\ell,t}(u \mid g) = (1 - a_t)\nu_g(u) + a_t \bar{f}_g(u). \tag{65}$$

It is the insertion intensity of the Poisson law with the mixture's first-order intensity. Indeed, $\mathbb{E}[w_{\ell,t}(X_t, g) \mid G = g] = \pi_\ell(g)$ implies

$$\mathbb{E}[\widehat{\lambda}_t(u; X_t, g) \mid G = g] = \lambda_t^{\text{off}}(u \mid g), \qquad \widehat{\lambda}_t - \lambda_t^{\text{off}} = \sum_\ell (w_{\ell,t} - \pi_\ell)\lambda_{\ell,t}.$$

Thus prior weighting determines the first-order intensity, and posterior weighting supplies the configuration-dependent insertion field.

The corresponding joint mark law conditional on locations is also explicit. For an enumeration $x_{1:N}$ of the unmarked configuration, let

$$\widetilde{\pi}_\ell(x_{1:N}, g) = \frac{\pi_\ell(g) e^{-\Lambda_{\ell,g}} \prod_{i=1}^N \rho_{\ell,g}(x_i)}{\sum_j \pi_j(g) e^{-\Lambda_{j,g}} \prod_{i=1}^N \rho_{j,g}(x_i)}.$$

Then

$$p(m_{1:N} \mid x_{1:N}, g) = \sum_\ell \widetilde{\pi}_\ell(x_{1:N}, g) \prod_{i=1}^N q_{\ell,g}(m_i \mid x_i). \tag{66}$$

When all components share an unmarked intensity ρ_g, these location-conditional weights reduce to $\pi_\ell(g)$.

C Conditioning on a global variable

Let G be a descriptor held fixed throughout corruption and reversal. The construction applies to observed descriptors and to latent variables in a joint law with the original data configuration. Positive-time insertion fields follow from conditional Poisson absolute continuity; a Papangelou intensity for the original conditional law supplies an additional denoising representation.

C.1 Conditional positive-time fields and denoising

Theorem 6 (DPSM conditional on a fixed descriptor)**.** *Let (X_0, G) have a joint law on $\mathcal{N}_f(E) \times \mathcal{G}$, with $\mathcal{G}$ standard Borel. Let $\nu_g(u) > 0$ be jointly measurable and satisfy $0 < \bar{\nu}_g = \int_E \nu_g \, \mathrm{d}\mu < \infty$ for P_G-almost every g. For such g, assume $P_0(\cdot \mid g) \ll \Pi_{\nu_g}$ and $\mathbb{E}[|X_0| \mid G = g] < \infty$. Conditionally on G, let the insertion and removal process have marginals*

$$X_t = \mathrm{Thin}_{a_t}(X_0) \oplus Z_t, \qquad Z_t \mid G \sim \Pi_{(1-a_t)\nu_G}, \qquad a_t = e^{-t}, \tag{67}$$

where points are retained independently and Z_t is independent of X_0 and the thinning randomness conditionally on G. For every $t > 0$, $P_t(\cdot \mid g)$ is equivalent to Π_{ν_g} and admits a conditional GNZ field satisfying $\lambda_t(u; X, g) \geq (1 - a_t)\nu_g(u)$ almost everywhere. If, additionally, the original conditional laws admit a jointly measurable Papangelou intensity $f(u; X, g)$ satisfying GNZ, then

$$\lambda_t(u; X_t, G) = (1 - a_t)\nu_G(u) + a_t \mathbb{E}[f(u; X_0, G) \mid X_t, G], \tag{68}$$

for $\mu \otimes P_{(X_t,G)}$-almost every (u, X_t, G).

Proof. Fix a descriptor g satisfying the assumptions. The conditional corruption kernel $K_{t,g}$ preserves Π_{ν_g}, so conditional absolute continuity gives $P_t(\cdot \mid g) \ll \Pi_{\nu_g}$. The event that every original point is removed has probability $\gamma_{t,g} = \mathbb{E}[(1 - a_t)^{|X_0|} \mid G = g] > 0$ and yields the measure inequality

$$P_t(\cdot \mid g) \geq \gamma_{t,g} \Pi_{(1-a_t)\nu_g}.$$

The two finite positive Poisson references are equivalent, which gives the reverse absolute continuity. Poisson Mecke applied to a positive conditional density defines its GNZ insertion ratio. The contribution from inserted points to the reduced Campbell measure is $(1 - a_t)\nu_g(u)\mu(\mathrm{d}u)P_t(\mathrm{d}X \mid g)$; the retained contribution is nonnegative. This proves the asserted lower bound, as in Appendix A.1. The conditional expected cardinality is $a_t \mathbb{E}[|X_0| \mid G = g] + (1 - a_t)\bar{\nu}_g < \infty$.

Under the additional GNZ assumption for the original conditional law, take a nonnegative measurable h and set $H_g(u, \xi) = \mathbb{E}[h(u, \mathrm{Thin}_{a_t}(\xi) \oplus Z_t) \mid G = g]$, using fresh conditional corruption. The retained contribution is

$$\begin{aligned} I_{\mathrm{ret}}(g) &= a_t \mathbb{E}\left[\sum_{u \in X_0} H_g(u, X_0 \setminus \{u\}) \mid G = g \right] \\ &= a_t \mathbb{E}\left[\int_E h(u, X_t) f(u; X_0, g) \, \mu(\mathrm{d}u) \mid G = g \right]. \end{aligned}$$

Adding the conditional Mecke contribution and conditioning on X_t proves equation 68. Nonnegativity permits Tonelli, and regular conditional laws give jointly measurable versions, as in Appendix A.2. □

C.2 Conditional population risks and reversal

Draw $t \sim \pi$ independently of (X_0, G), where π has positive density on $[\varepsilon, T]$ with $0 < \varepsilon < T$. For a positive candidate field, define the configuration functional

$$\ell_{\theta,t,g}(X) = \int_E \lambda_\theta(u; X, t, g) \, \mu(\mathrm{d}u) - \sum_{u \in X} \log \lambda_\theta(u; X \setminus \{u\}, t, g).$$

The conditional version of equation 14 is

$$\mathcal{L}^{\mathrm{cond}}_{\mathrm{NPL}}(\theta) = \mathbb{E}_{G,t,X_t} \ell_{\theta,t,G}(X_t). \tag{69}$$

The same descriptor is used in every insertion and reduced-configuration evaluation. Under finite-risk assumptions, conditional GNZ gives

$$\begin{aligned} &\mathcal{L}^{\mathrm{cond}}_{\mathrm{NPL}}(\theta) - \mathcal{L}^{\mathrm{cond}}_{\mathrm{NPL}}(\lambda) \\ &\quad = \mathbb{E}_{G,t,X_t} \int_E D_{\mathrm{P}}(\lambda_t(u; X_t, G) \| \lambda_\theta(u; X_t, t, G)) \, \mu(\mathrm{d}u). \end{aligned}$$

This conclusion uses the positive-time part of Theorem 6.

Denoising representation. Under its additional GNZ assumption for the original conditional law, let $\widehat{f}$ be a fixed, nonnegative, jointly measurable intensity estimate and set $r_\theta(u; X_t, t, G) = \lambda_\theta(u; X_t, t, G)/\nu_G(u)$. The associated target is

$$T_t(u; X_0, G; \widehat{f}) = (1 - a_t) + a_t \frac{\widehat{f}(u; X_0, G)}{\nu_G(u)}. \tag{70}$$

Conditionally on G, draw $U \sim \nu_G \mu/\bar{\nu}_G$ independently of X_0, t and the corruption randomness. For $\widehat{T} = T_t(U; X_0, G; \widehat{f})$ under this joint query law, assume $\mathbb{E}\widehat{T} < \infty$ and $\mathbb{E}[\widehat{T} \log_+ \widehat{T}] < \infty$, where $\log_+ x = \max\{\log x, 0\}$, together with the finite prediction risks in Appendix A.4. The decomposition in equation 36, conditioned on $C = (U, X_t, t, G)$, yields $r^*(C) = \mathbb{E}[T_t(U; X_0, G; \widehat{f}) \mid C]$ and

$$\begin{aligned} &\nu_G(u) r^*(u; X_t, t, G) - \lambda_t(u; X_t, G) \\ &\qquad = a_t \mathbb{E}[\widehat{f}(u; X_0, G) - f(u; X_0, G) \mid X_t, t, G]. \end{aligned} \tag{71}$$

Using the true intensity $\widehat{f} = f$ gives a representation of the same conditional noisy field.

The query normalization also determines the weighting across descriptors. Define

$$\mathcal{L}_{\mathrm{NPL}}^{\mathrm{cond},\nu}(\theta) = \mathbb{E}_G \left[\frac{1}{\bar{\nu}_G} \mathbb{E}_{t, X_t \mid G} \ell_{\theta, t, G}(X_t) \right].$$

For the denoising target based on the true intensity and the finite expectations required in Appendix A.4, homogeneity of Poisson divergence and conditional GNZ imply

$$\mathbb{E} D_{\mathrm{P}}(T_t(U; X_0, G; f) \| r_\theta(U; X_t, t, G)) = \mathcal{L}_{\mathrm{NPL}}^{\mathrm{cond},\nu}(\theta) + C_f, \tag{72}$$

where C_f is independent of θ. This equality holds for every candidate field and therefore preserves risk ordering within a common candidate family. For descriptor-dependent $\bar{\nu}_G$, the normalized regression corresponds to the displayed $1/\bar{\nu}_G$ weighting. Multiplying each descriptor's regression contribution by $\bar{\nu}_G$ gives an equivalent representation of the raw risk in equation 69, under the corresponding finite-expectation conditions.

Conditional reversal. Conditional GNZ transforms the forward probability fluxes into

$$b_t(u; X, G) = \lambda_t(u; X, G), \qquad d_t(z; X, G) = \frac{\nu_G(z)}{\lambda_t(z; X \setminus \{z\}, G)}.$$

The derivation in Appendix A.5 applies at each fixed descriptor. Using a common estimated field in both rates gives

$$b_{\theta,t}(u; X, G) = \nu_G(u) r_\theta(u; X, t, G), \qquad d_{\theta,t}(z; X, G) = \frac{1}{r_\theta(z; X \setminus \{z\}, t, G)}. \tag{73}$$

The exact reversal from $P_T(\cdot \mid g)$ has endpoint $P_\varepsilon(\cdot \mid g)$ on $[\varepsilon, T]$. Under the hypotheses of Theorem 5 for almost every g, its field error and initialization bounds apply conditionally, together with

$$d_{\mathrm{TV}}(P_\varepsilon(\cdot \mid g), P_0(\cdot \mid g)) \le (1 - e^{-\varepsilon}) \left(\mathbb{E}[|X_0| \mid G = g] + \bar{\nu}_g \right).$$

Integrating the conditional bounds over P_G controls the joint law of (X, G) and its configuration marginal.

C.3 Descriptors obtained from stochastic encoders

Proposition 7 (Conditional laws induced by a stochastic encoder)**.** *Suppose $P_0 \ll \Pi_\nu$ and $\mathbb{E}_{P_0}|X_0| < \infty$. Let $q_\xi(g \mid X)$ be a nonnegative measurable conditional density relative to a sigma-finite measure $\zeta(\mathrm{d}g)$, with $\int q_\xi(g \mid X)\, \zeta(\mathrm{d}g) = 1$. Define the joint law $P_0(\mathrm{d}X) q_\xi(g \mid X) \zeta(\mathrm{d}g)$ and $m(g) = \mathbb{E}_{P_0} q_\xi(g \mid X_0)$. For P_G-almost every g, the conditional law is Poisson absolutely continuous and has finite expected cardinality. If, additionally, P_0 has a hereditary density p_0 and the encoder density is strictly positive and finite, its original conditional insertion field is*

$$f(u; X, g) = f(u; X) \frac{q_\xi(g \mid X \cup \{u\})}{q_\xi(g \mid X)} \tag{74}$$

on the support of the conditional law, with the same g in both encoder evaluations.

Proof. For $0 < m(g) < \infty$, the conditional density relative to Π_ν is $p_g(X) = p_0(X)q_\xi(g \mid X)/m(g)$. The identity $\int m(g)\,\zeta(\mathrm{d}g) = 1$ gives this representation for P_G-almost every g. Conditional cardinalities are finite almost everywhere because $\mathbb{E}_G\mathbb{E}[|X_0| \mid G] = \mathbb{E}|X_0| < \infty$. Under the additional assumptions, positivity of the encoder preserves hereditary support. Taking the density ratio $\nu(u)p_g(X \cup \{u\})/p_g(X)$ proves equation 74. □

For example, the deterministic descriptor $G = |X_0|$ has a conditional density with respect to counting measure on the nonnegative integers. For every n with $P(|X_0| = n) > 0$, $P_0(\cdot \mid G = n)$ is Poisson absolutely continuous and has cardinality n. Its corrupted laws therefore satisfy the positive-time conclusion of Theorem 6. The denoising representation uses the additional conditional GNZ hypothesis stated in that theorem.

Conditional NPL on (X_t, G) targets the insertion field of the conditional law through its reduced-configuration risk. An encoder representation is held fixed throughout each path. If the available descriptor is $S = \mathrm{decode}(G)$, the applicable construction uses the conditional laws given S. The stronger relation $P_0(\cdot \mid G) = P_0(\cdot \mid S)$ holds when X_0 and G are conditionally independent given S.

C.4 Marginal generation and conditional observations

The descriptor prior and the conditional laws determine the original data marginal by disintegration:

$$P_0(\mathrm{d}X) = \int P_0(\mathrm{d}X \mid g)\, P_G(\mathrm{d}g). \tag{75}$$

For encoder-derived descriptors, P_G is the aggregate encoder law with density $m(g)$. More generally, if Q_g approximates $P_0(\cdot \mid g)$, then

$$d_{\mathrm{TV}}\left(\int Q_g\, P_G(\mathrm{d}g), P_0\right) \le \int d_{\mathrm{TV}}(Q_g, P_0(\cdot \mid g))\, P_G(\mathrm{d}g).$$

This gives the marginal counterpart of the conditional reverse error bounds.

For an observation $Y = X_t$ at a known $t > 0$ from equation 67, disintegration also gives

$$P(\mathrm{d}X_0 \mid Y) = \int P(\mathrm{d}X_0 \mid Y, g)\, P(\mathrm{d}g \mid Y). \tag{76}$$

The exact fixed-g reversal from Y has terminal law $P(\mathrm{d}X_0 \mid Y, g)$ whenever the reversal extends to the clean endpoint. Combining it with the descriptor posterior realizes equation 76. On $[\varepsilon, t]$, the same construction recovers the corresponding conditional law of X_ε.

At complete corruption, the conditional law is Π_{ν_G}. A common reference $\nu_G = \nu$ makes the corrupted configuration independent of G. For a descriptor-dependent reference, the joint initialization law is $P_G(\mathrm{d}g)\Pi_{\nu_g}(\mathrm{d}X)$. Keeping the descriptor as part of the state gives a Markov construction throughout corruption and reversal.

D Scores, local transitions, and configuration space risks

The score of a probability law describes its variation under local changes of state. Coordinate perturbations lead to a differential score, while point insertions lead to a finite density ratio. In both settings, the forward transition structure determines the score required by time reversal.

D.1 Local density variation and reverse transitions

For a smooth positive density q_t on $\mathbb{R}^q$ and a direction $v \in \mathbb{R}^q$,

$$\log \frac{q_t(z + hv)}{q_t(z)} = hv^\top \nabla \log q_t(z) + o(h). \tag{77}$$

Thus the gradient score describes the first-order change of log density under a coordinate displacement. Consider the diffusion

$$\mathrm{d}Z_t = b_t(Z_t)\,\mathrm{d}t + \sqrt{2\kappa_t}\,\mathrm{d}B_t, \tag{78}$$

where $\kappa_t > 0$ is deterministic. Assume smooth positive marginal densities and the regularity required for nonexplosive time reversal. With elapsed reverse time $\tau = T - t$, the reverse drift is

$$\widetilde{b}_\tau(z) = -b_t(z) + 2\kappa_t \nabla \log q_t(z). \tag{79}$$

Indeed, reversing the forward probability current $b_t q_t - \kappa_t \nabla q_t$ gives this expression (Song et al., 2020).

On marked configuration space, let $p_t = \mathrm{d}P_t/\mathrm{d}\Pi_\nu$ be a positive density. The insertion log density difference is

$$\log r_t(u; X) = \log p_t(X \cup \{u\}) - \log p_t(X), \qquad r_t(u; X) = \lambda_t(u; X)/\nu(u). \tag{80}$$

These identities are understood on the support of the relevant measures. At positive corruption times, Proposition 3 supplies a positive density version. An insertion connects configurations whose cardinalities differ by one. The GNZ flux calculation in Appendix A.5 then gives reverse insertion intensity νr_t and per-point removal rate $1/r_t$ given the surrounding configuration. The same insertion ratio therefore enters the two directions of each configuration space transition.

D.2 Compatibility and density reconstruction

Both scores are invariant under multiplication of a density by a positive constant. For a smooth density, integrating its gradient score along a path recovers a log density difference. For a configuration $X = \{u_1, \ldots, u_N\}$, set $\Xi_i = \{u_1, \ldots, u_i\}$ and $\Xi_0 = \emptyset$. On an insertion path with positive density,

$$\log p_t(X) - \log p_t(\emptyset) = \sum_{i=1}^{N} \log r_t(u_i; \Xi_{i-1}). \tag{81}$$

The local condition for invariance under exchanging adjacent insertions is

$$r_t(u; X)\, r_t(v; X \cup \{u\}) = r_t(v; X)\, r_t(u; X \cup \{v\}). \tag{82}$$

For completeness, suppose a positive finite measurable candidate ratio r satisfies this cocycle for all admissible insertions. Define $w(\emptyset) = 1$ and

$$w(X) = \prod_{i=1}^{N} r(u_i; \Xi_{i-1}).$$

Every permutation is a composition of adjacent exchanges, so the cocycle makes $w(X)$ independent of the ordering. With a measurable ordering of finite configurations, w is measurable. If

$$0 < Z := \int w(X)\, \Pi_\nu(\mathrm{d}X) < \infty, \tag{83}$$

then $p(X) = w(X)/Z$ is a probability density and $p(X \cup \{u\})/p(X) = r(u; X)$. Poisson Mecke gives its Papangelou field νr. For example, $r(u; X) \le C$ implies $w(X) \le C^{|X|}$ and hence $Z \le \exp[\bar{\nu}(C-1)] < \infty$. This construction states separately the insertion compatibility and integrability conditions for a density representation. The population risks in Appendix A.3 apply to positive field estimators, and their use in continuous-time generation is governed by the conditions of Theorem 5.

D.3 Integration identities and population score matching

The Euclidean and point process risks arise from integration identities for their respective local transitions. For a continuously differentiable vector field s and a smooth density q, assume finite required integrals and boundary conditions giving $\int \mathrm{div}(qs)\, \mathrm{d}z = 0$. Then

$$\mathbb{E}_q[\mathrm{div}\, s(Z)] = -\mathbb{E}_q[s(Z)^\top \nabla \log q(Z)]. \tag{84}$$

Consequently,

$$\begin{aligned} \mathbb{E}_q\big[\tfrac{1}{2}\|s(Z)\|^2 + \mathrm{div}\, s(Z)\big] = \tfrac{1}{2}\mathbb{E}_q\|s(Z) - \nabla \log q(Z)\|^2 \\ - \tfrac{1}{2}\mathbb{E}_q\|\nabla \log q(Z)\|^2. \end{aligned} \tag{85}$$

Integration by parts thus represents a score matching risk through expectations under q.

For point configurations, GNZ supplies the corresponding integration identity. At a fixed positive noise level, it gives

$$\mathbb{E}_{P_t} \sum_{z \in X} \log \lambda_\theta(z; X \setminus \{z\}, t) = \mathbb{E}_{P_t} \int_E \lambda_t(u; X) \log \lambda_\theta(u; X, t)\, \mu(\mathrm{d}u),$$

under absolute integrability. Subtracting the true field risk yields

$$\mathcal{R}_t(\lambda_\theta) - \mathcal{R}_t(\lambda_t) = \mathbb{E}_{P_t} \int_E D_{\mathrm{P}}(\lambda_t \| \lambda_\theta)\, \mathrm{d}\mu. \tag{86}$$

Here $\mathcal{R}_t$ is the NPL risk in Appendix A.3. The Poisson divergence measures errors in positive insertion intensities, and GNZ represents its mixed term by a reduced point sum.

D.4 Conditional mean denoising representations

Let $k_t(z \mid z_0)$ be a differentiable Euclidean corruption density on a common positive support, with differentiation interchangeable with integration. For $q_t(z) = \int k_t(z \mid z_0) P_0^{\mathrm{Euc}}(\mathrm{d}z_0)$,

$$\begin{aligned} \nabla_z \log q_t(z) &= \frac{\int \nabla_z k_t(z \mid z_0) P_0^{\mathrm{Euc}}(\mathrm{d}z_0)}{q_t(z)} \\ &= \mathbb{E}[\nabla_z \log k_t(z \mid Z_0) \mid Z_t = z]. \end{aligned} \tag{87}$$

With finite second moments, the corresponding squared-error denoising risk equals the marginal score risk plus a prediction-independent conditional variance term.

For point configurations, the GNZ identity for the original data law yields

$$r_t(u; X_t) = \mathbb{E}\left[(1 - a_t) + a_t \frac{f(u; X_0)}{\nu(u)} \,\middle|\, X_t \right]. \tag{88}$$

The denoising target is defined through the original insertion field, and the prediction depends on the corrupted configuration. Conditional expectation and the Bregman decomposition for Poisson divergence give

$$\mathcal{L}_{\mathrm{DPSM}}(\theta) = \bar{\nu}^{-1} \mathcal{L}_{\mathrm{NPL}}(\theta) + C$$

under the assumptions of Theorem 2. This identity holds for each candidate in a fixed family, so it also preserves the optimum under model restriction. The two population representations evaluate the same denoising criterion using either denoising targets from original data configurations or reduced sums over corrupted configurations.

E Models and baseline implementations

The general insertion field constructions are given in Appendix B. The interacting-process experiments use a symmetric radial insertion field; the synthetic spatial structure experiments combine a learned spatial field, a global configuration energy, and a local neural energy.

E.1 DPSM for interacting point processes

For candidate type a, we use a positive insertion field estimator with symmetric radial terms and an intermediate-time composition term:

$$\log \lambda_{q,\theta}(u, a; X) = c_a(q) + \sum_{(x,b) \in X} \left[\sum_{k=1}^{K} w_{kab}(q) \rho_k(\|u - x\|) + g_{ab}(q) \right], \tag{89}$$

where $q \in [0, 1]$ is the retention probability and

$$\rho_k(r) = \exp\left[-\tfrac{1}{2}((r - \mu_k)/\sigma)^2 \right] \left[1 - (r/R)^2 \right]_+^2. \tag{90}$$

Both w_{kab} and g_{ab} are symmetric in a, b. A two-hidden-layer, width-64 SiLU MLP receives seven time features. Its bounded residual is gated by $4q(1-q)$, preserving the learned radial endpoint and training-estimated marked Poisson endpoint. The global pair term vanishes at $q = 1$. The resulting time-indexed fields provide the insertion rate estimates used by the numerical generator.

Learning the noisy path. Training uses 4,000 updates, learning rate 0.002, batch size four, and 32 query locations. Retention is uniform on $[e^{-8}, 1]$. Independent thinning with probability q and superposed marked Poisson noise of intensity $(1-q)\nu$ produce corrupted configurations for empirical negative pseudolikelihood:

$$\mathcal{L} = \mathbb{E}_{q,X_q}\left[\int_{[0,1]^2}\sum_a \lambda_{q,\theta}(u,a;X_q)\,du - \sum_{(u,a)\in X_q}\log\lambda_{q,\theta}(u,a;X_q\setminus\{(u,a)\})\right]. \tag{91}$$

Noise path and sampling. Numerical generation uses 200 reverse steps and a 300-point capacity.

E.2 DPSM FOR SYNTHETIC SPATIAL STRUCTURES

Spatial field fitting. The conditional Poisson VAE takes mark-summed counts and the dataset label, with a 16-dimensional latent, four stride-two convolutional stages of widths $(32, 64, 128, 128)$, group normalization, and SiLU activations. Its decoder predicts normalized spatial mass and expected total count.

Global configuration energy. The global energy is induced by a mixture of 48 Poisson processes in 24 pairs, each with positive intensity. The shared network has two width-64 SiLU hidden layers, 48 outputs, an eight-dimensional dataset embedding, coordinates with Fourier frequencies $(1, 2, 4, 8)$, and learned spatial fields. We also introduce a separate channel to embed the mark distribution into the global configuration energy.

Local neural energy. We add a continuous-coordinate neighborhood energy to the configuration energy, $U_q(X) = U_q^{\mathrm{global}}(X) + R_{\theta,q}(X)$. For radii $r \in \{0.025, 0.05, 0.10\}$, define

$$C_{irm} = \sum_{j\neq i}\left[1 - \frac{\|x_i - x_j\|}{r}\right]_+^2 \mathbf{1}\{m_j = m\}, \qquad h_{irm} = \frac{C_{irm}}{1+\sum_b C_{irb}}.$$

Let $\psi_\theta(h, m, q)$ be the hyperbolic tangent of a shared two-hidden-layer, width-16 SiLU network, with inputs h, the central mark's one-hot encoding, and $(q, q^2, \sin(\pi q), \cos(\pi q))$. At retention q, the added energy is

$$R_{\theta,q}(X) = \frac{q}{2}\sum_i\big[\psi_\theta(h_i, m_i, q) - \psi_\theta(0, m_i, q)\big].$$

Noise path and sampling. Numerical generation uses 240 equal intervals from time six to zero, midpoint retention $q = e^{-t}$, 2,048 uniform insertion candidates per interval, and a 2,600-point capacity. Each interval applies removals, then a Poisson number of inserts; candidate locations are selected according to current intensity, followed by marks.

E.3 DPSM FOR REAL BIOLOGICAL DATA

The real biological data experiments use the same DPSM architecture as the synthetic spatial structure experiments (Appendix E.2). The cell type vocabularies and biological conditions for DCIS, IMC, and CRC are described in Appendix H.1.

E.4 DPSM TRAINING AND GENERATION

Algorithm 1 summarizes the common workflow for interacting point processes, synthetic spatial structures, and real biological data. The condition c is omitted in unconditional experiments.

Given a data configuration X_0, corruption at time t produces

$$X_t = \mathrm{Thin}_{a_t}(X_0)\cup Z_t, \qquad a_t = e^{-t}, \qquad Z_t \sim \mathrm{PPP}((1-a_t)\nu). \tag{92}$$

The noisy insertion field is learned using noisy negative pseudolikelihood (NPL),

$$\mathcal{L}_{\mathrm{NPL}}(\theta) = \mathbb{E}\left[\int_E \lambda_\theta(u; X_t, t, c)\,\mu(du) - \sum_{u\in X_t}\log\lambda_\theta(u; X_t\setminus\{u\}, t, c)\right], \tag{93}$$

where E is the marked observation space and the expectation is over training configurations, noise levels, and corruptions. The integral is estimated numerically, and every observed point contributes a leave-one-out term.

Algorithm 1 DPSM: learning and generation

Require: Training data $\mathcal{D}_{\mathrm{tr}}$; reference intensity ν; generation condition $c^\star$; time grid $0 = t_0 < \cdots < t_K = T$

Ensure: Generated marked configuration $\widehat{X}$

Stage 1: Estimate the clean energy

Fit a clean energy model $\widehat{U}_0(X; c)$ using $\mathcal{D}_{\mathrm{tr}}$

Stage 2: Learn the noise path

Construct the time-dependent insertion field $\lambda_\theta(u; X, t, c)$ from the clean model and the Poisson reference

for each training iteration **do**

  Sample a minibatch of data configurations and their conditions

  Sample noise levels and independently corrupt each configuration by thinning and Poisson superposition

  Estimate the minibatch loss in equation equation 93

  Update the trainable path parameters

end for

Stage 3: Reverse generation

Prepare the conditioning information for $c^\star$

Initialize $X \sim \mathrm{PPP}(\nu)$

for $k = K, K-1, \ldots, 1$ **do**

  Set $\Delta t = t_k - t_{k-1}$ and $\bar{t}_k = (t_k + t_{k-1})/2$

  Evaluate the reverse insertion intensity

$$b_k(u; X) = \lambda_\theta(u; X, \bar{t}_k, c^\star)$$

  Evaluate the reverse removal rate for each $u \in X$

$$d_k(u; X) = \frac{\nu(u)}{\lambda_\theta(u; X \setminus \{u\}, \bar{t}_k, c^\star)}$$

  Numerically evolve X for duration Δt using these insertion and removal rates

end for

return $\widehat{X} = X$

E.5 Baseline methods and training budgets

We compare with PSD, EBP, SI-VAE, and SetVAE in the interacting-process and synthetic spatial structure experiments, and with PSD, SI-VAE, and CytoSpatio in the real biological data experiments. Let M denote the number of marks: $M = 1, 3, 4$ for the three interacting processes, $M = 3$ for structural data, and $M = 6, 6, 8$ for DCIS, IMC, and CRC. The interacting-process datasets use separate models; the six structural datasets share one conditional model per method and seed. For biological data, PSD is shared across conditions within each dataset, whereas SI-VAE and CytoSpatio are fitted separately per condition.

PSD. We extend the point set diffusion construction of Lüdke et al. (2025) with an $M \times h$ mark embedding table and an M-category mark distribution for each spatial Gaussian mixture component. For conditional datasets, a learned h-dimensional label embedding is added to the pooled context. The interaction benchmarks additionally use relative-coordinate messages with eight distance RBFs and mark pair embeddings of dimension $\max(8, h/4)$; message MLPs have width h. The hidden width h is selected as 48. Training uses Adam with learning rate 10^{-3}. The thinning and superposition kernel and reverse posterior are retained.

EBP. We retain the released posterior encoder, set energy network, and recurrent generator of Yang et al. (2020). Coordinates are mapped to $[-1, 1]^2$; marked data append M one-hot channels, decoded by argmax. The Gaussian latent dimension is 64, encoder widths are $(128, 64)$, energy network

widths are $(128, 128)$, and recurrent input dimension is 64. For variable-cardinality interaction data, a standardized scalar count conditions the encoder, energy, and generator, giving recurrent state dimension 65 and block size one. Generation draws the count from its training empirical distribution. For structural data, a six-category one-hot label instead conditions these modules, giving state dimension 70 and block size 50; training subsamples 1,000 points per configuration and generation returns 1,000 points. The learning rate is 10^{-4}, with one energy/posterior update and three generator updates per sampled minibatch. Posterior noise scale, entropy weight, and gradient penalty weight are 0.1, 10^{-5}, and one.

SI-VAE. We use the image-free conditional intensity model in Appendix B of Sturm et al. (2026). Coordinates are mapped to $[-1, 1]^2$. To accommodate marks, the unary network receives coordinates and an M-dimensional one-hot mark; the pair network receives distance and an $M(M+1)/2$-dimensional unordered-mark-pair encoding. Both networks have four width-64 ReLU hidden layers; pair contributions are weighted by $\exp(-8d)$. For joint structural training, a six-dimensional one-hot dataset label is appended to both inputs. Fitting uses count-normalized negative pseudolikelihood, 64 integration locations summed over all marks, and Adam at 5×10^{-4}. Generation retains the native insertion and removal sampler with 30,000 proposals per configuration.

SetVAE. We retain the hierarchical encoder, decoder, Gaussian priors, and Chamfer–KL objective of Kim et al. (2021), appending unit one-hot marks to coordinates and decoding generated marks by argmax. The five latent set sizes are $(2, 4, 8, 16, 32)$, hidden dimension 64, latent dimension 16, attention head count four, and inducing point count 16. Structural conditioning adds a 64-dimensional label embedding to encoder and initial decoder features; five $64 \to 32$ linear maps shift the hierarchical prior means and log variances. Output counts are drawn from the training empirical distribution, conditional on the requested dataset for structural generation. Adam uses learning rate 10^{-3} and KL weight 0.01.

CytoSpatio. The official multirange, multitype implementation of Chen et al. (2025) is adapted to independent tissue windows and arbitrary mark vocabularies. Quadrature is computed within each window before pooling design rows. The interaction range is 50 μm in four 12.5 μm bands; the hard-core radius is the smaller of 0.1 μm and half the minimum training separation. The quasi-Poisson GLM retains its 50-iteration limit. Absent training types are omitted locally and mapped back to the common vocabulary. The native simulator draws Poisson locations at the training density and performs N single-cell mark updates, keeping locations fixed.

Equivalent budgets. We define

$$E_{\text{eff}} = \frac{\sum_{s=1}^{S} B_s}{N_{\text{train}}},$$

where B_s is the number of real training configurations drawn in minibatch s, including repeated draws, and N_{train} is the full training set size. Gradient accumulation counts all constituent configurations; EBP's repeated optimization on the same minibatch counts that draw once. In the interacting-process experiments, all four baselines are evaluated at the same equivalent-epoch budget as DPSM: 16,000 draws for each Strauss dataset and Gibbs. In the structural experiments, all methods receive 200 equivalent epochs. Local and collective MCMC protocols are given in Appendix I.

F Interacting point process experiments

F.1 Data-generating processes and independent-chain splits

We model finite unordered configurations $X = \{(x_i, m_i)\}_{i=1}^{N}$ on the unit square with Euclidean distances. Densities are relative to a unit-intensity Poisson reference per mark; positive pair log density terms favor pairs and negative terms penalize them. Fitting uses configurations only, withholding simulator activities, radii, pair matrices, and hard-core parameters.

Unmarked Strauss. In evaluation coordinates, the density is

$$\widetilde{p}(X) = 160^N \, (0.4)^{s_{0.1}(X)}, \qquad s_r(X) = \sum_{i<j} \mathbf{1}\{\|x_i - x_j\| < r\}. \tag{94}$$

Simulation on $[-1,1]^2$ with activity 40 and radius 0.2 is transformed by $x \mapsto (x+1)/2$, giving activity 160 per unit area. Each of 100 independent insertion and removal chains uses 20,000 burn-in proposals and retains 50 states at 500-proposal intervals. Training/validation/test contain 80/10/10 complete chains and 4,000/500/500 configurations.

Three-type Strauss. Locations and marks are simulated jointly:

$$\widetilde{p}(X) = \prod_{i=1}^{N} \beta_{m_i} \prod_{i<j} \Gamma_{m_i m_j}^{\mathbf{1}\{\|x_i - x_j\| < 0.1\}}, \quad \beta = (64, 56, 40), \quad \Gamma = \begin{pmatrix} 0.25 & 0.60 & 0.45 \\ 0.60 & 0.35 & 0.55 \\ 0.45 & 0.55 & 0.40 \end{pmatrix}. \tag{95}$$

Chain numbers, burn-in, retention intervals, and splits match unmarked Strauss. Insertion, removal, relocation, and mark flip probabilities are $0.3, 0.3, 0.2, 0.2$.

Four-type Gibbs. With $d_{ij} = \|x_i - x_j\|$, the density is

$$\widetilde{p}(X) = \mathbf{1}\{\min_{i<j} d_{ij} \geq 0.018\} \exp\left\{ \sum_i \alpha_{m_i} - 0.005(N-80)^2 + \sum_{i<j} \phi_{m_i m_j}(d_{ij}) \right\}, \tag{96}$$

where the calibrated log activities, rounded here to two decimals, are $\alpha \approx (3.15, 2.94, 3.02, 2.75)$ and

$$\phi_{ab}(r) = \frac{1}{2} \sum_{\ell=1}^{3} B_{ab}^{(\ell)} \mathbf{1}\{l_\ell \leq r < u_\ell\}, \quad ([l_\ell, u_\ell))_{\ell=1}^{3} = ([0.018, 0.05), [0.05, 0.12), [0.12, 0.25)). \tag{97}$$

The interaction matrices are

$$B^{(1)} = \begin{pmatrix} -0.55 & -0.35 & -0.70 & -0.40 \\ -0.35 & -0.50 & -0.25 & -0.65 \\ -0.70 & -0.25 & -0.45 & -0.30 \\ -0.40 & -0.65 & -0.30 & -0.50 \end{pmatrix}, \quad B^{(2)} = \begin{pmatrix} 0.18 & 0.42 & -0.38 & 0 \\ 0.42 & 0.12 & 0.08 & -0.32 \\ -0.38 & 0.08 & 0.15 & 0.36 \\ 0 & -0.32 & 0.36 & 0.10 \end{pmatrix}, \tag{98}$$

$$B^{(3)} = \begin{pmatrix} 0.05 & 0.12 & -0.10 & 0 \\ 0.12 & 0.04 & 0.02 & -0.08 \\ -0.10 & 0.02 & 0.05 & 0.10 \\ 0 & -0.08 & 0.10 & 0.03 \end{pmatrix}. \tag{99}$$

The factor $1/2$ applies in addition to counting unordered pairs once. Starting from log target mean counts $(26, 22, 18, 14)$, two damped pilot simulation rounds calibrate α; production uses the archived full-precision values. Counts vary between configurations. Production uses 64 chains, 30,000 burn-in proposals, and 1,000-proposal retention intervals. Training/validation/test contain 44/8/12 whole chains and 2,048/256/512 configurations. Insertion, removal, relocation, mark flip, and mark swap probabilities are $0.25, 0.25, 0.20, 0.15, 0.15$.

Model architecture, fitting, and generation settings are given in Appendix E.1; baseline settings are summarized in Appendix E.5.

F.2 Evaluation metrics

Each method generates 512 configurations per dataset and seed, compared with fixed test populations of 500 for each Strauss dataset and 512 for Gibbs. Generated and test configurations are unpaired. Results report mean and sample standard deviation ($\mathrm{ddof} = 1$) over seeds 20261001, 20261002, and 20261003.

Let $\mathcal{G}$ and $\mathcal{R}$ denote the generated and reference populations, and let $\mathcal{D} \in \{\mathcal{G}, \mathcal{R}\}$. For $X = \{(x_i, m_i)\}_{i=1}^{N(X)}$ on $[0,1]^2$, define $d_{ij} = \|x_i - x_j\|_2$ and $N_a(X) = \sum_i \mathbf{1}\{m_i = a\}$, where $a \in \{1, \ldots, M\}$ indexes mark types. Write W_1 for the one-dimensional Wasserstein distance and δ_t for a unit point mass at t.

Nearest-neighbor Wasserstein distances. These metrics compare pooled nearest-neighbor distances, weighting points equally and ignoring marks. Let $d_i^{\mathrm{NN}}(X) = \min_{j \neq i} d_{ij}$ and $L_{\mathcal{D}} = \sum_{X \in \mathcal{D}: N(X) \geq 2} N(X)$. Define

$$\begin{aligned} \widehat{\mu}_{\mathcal{D}}^{(\gamma)} &= \frac{1}{L_{\mathcal{D}}} \sum_{X \in \mathcal{D}: N(X) \geq 2} \sum_{i=1}^{N(X)} \delta_{N(X)^{\gamma} d_i^{\mathrm{NN}}(X)}, \\ \mathrm{NN}_{\gamma} &= W_1\left(\widehat{\mu}_{\mathcal{G}}^{(\gamma)}, \widehat{\mu}_{\mathcal{R}}^{(\gamma)}\right). \end{aligned} \tag{100}$$

Here $\gamma = 0$ gives NN W_1, while $\gamma = 1/2$ gives scaled NN W_1.

Neighbor mark RMSE. For $\mathcal{R}_M = \{0.05, 0.10, 0.20\}$, directed edges are pooled across configurations, excluding self-pairs and zero distances. Row normalization by source mark gives neighboring mark distributions:

$$\begin{aligned} C_{ab}^{\mathcal{D}}(r) &= \sum_{X \in \mathcal{D}} \sum_{i \neq j} \mathbf{1}\{m_i = a, m_j = b,\, 0 < d_{ij} \leq r\}, \\ P_{ab}^{\mathcal{D}}(r) &= \frac{C_{ab}^{\mathcal{D}}(r)}{\max\{\sum_{c=1}^{M} C_{ac}^{\mathcal{D}}(r), 1\}}, \\ \mathrm{RMSE}_{\mathrm{NM}} &= \left[\frac{\sum_{r \in \mathcal{R}_M} \sum_{a,b=1}^{M} (P_{ab}^{\mathcal{G}}(r) - P_{ab}^{\mathcal{R}}(r))^2}{M^2 |\mathcal{R}_M|}\right]^{1/2}. \end{aligned} \tag{101}$$

This metric is omitted for unmarked data.

Spectral error. Let A_X be the 32×32 count grid with marks pooled, and let $\widetilde{A}_X = A_X - N(X)/32^2$ be its centered version. For integer Fourier frequencies $\omega \in \{-16, \ldots, 15\}^2$, define $B_k = \{\omega : \lfloor \|\omega\|_2 \rfloor = k\}$. Using the discrete Fourier transform $\mathcal{F}$, compute

$$\begin{aligned} \overline{P}_{\mathcal{D}}(k) &= \frac{1}{|\mathcal{D}||B_k|} \sum_{X \in \mathcal{D}} \sum_{\omega \in B_k} \frac{|\mathcal{F}[\widetilde{A}_X](\omega)|^2}{\max\{N(X), 1\}}, \\ \mathrm{Err}_{\mathrm{spec}} &= \frac{\sum_{k=1}^{16} |\overline{P}_{\mathcal{G}}(k) - \overline{P}_{\mathcal{R}}(k)|}{\sum_{k=1}^{16} |\overline{P}_{\mathcal{R}}(k)|}. \end{aligned} \tag{102}$$

This relative L^1 error uses nonzero frequency bins 1–16 and weights configurations equally.

Enrichment RMSE. We compare pooled log ratio enrichment at radii $\mathcal{R}_E = \{0.025, 0.05, 0.10, 0.20\}$ over unordered mark pairs $a \leq b$. For each configuration, let $O_{ab}(X, r)$ count distinct pairs $i < j$ with marks $\{a, b\}$ and $d_{ij} \leq r$, and set $Q_{aa}(X) = \binom{N_a(X)}{2}$ and $Q_{ab}(X) = N_a(X) N_b(X)$ for $a < b$. For marked data, random labeling conditional on coordinates and mark counts gives

$$E_{ab}(X, r) = \frac{Q_{ab}(X)}{\binom{N(X)}{2}} \sum_{c \leq d} O_{cd}(X, r). \tag{103}$$

For unmarked data, the reference is independent uniform locations with the same count, giving $E_{11}(X, r) = \binom{N(X)}{2}(\pi r^2 - 8r^3/3 + r^4/2)$ on the unit square. Both expectations are zero when $N(X) < 2$. We pool observed and expected counts before taking the ratio:

$$\begin{aligned} L_{ab}^{\mathcal{D}}(r) &= \log_2 \frac{\sum_{X \in \mathcal{D}} O_{ab}(X, r) + 1/2}{\sum_{X \in \mathcal{D}} E_{ab}(X, r) + 1/2}, \\ \mathrm{RMSE}_{\mathrm{enrich}} &= \left[\frac{1}{|\mathcal{S}|} \sum_{(a,b,r) \in \mathcal{S}} (L_{ab}^{\mathcal{G}}(r) - L_{ab}^{\mathcal{R}}(r))^2\right]^{1/2}, \end{aligned} \tag{104}$$

where $\mathcal{S}$ contains entries with pooled reference expectation at least 20 and is fixed across methods. All reported entries have positive generated expectations. This measures spatial association relative to the specified null.

Moran RMSE. For a scalar signal y on n spatial units, we compute

$$I(y, W) = \frac{n}{S_0} \frac{\sum_{i,j} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}, \qquad S_0 = \sum_{i,j} w_{ij}. \tag{105}$$

Weights are binary and symmetric, with no self-edges or row normalization. For marked data, units are points and $y_i = \mathbf{1}\{m_i = a\}$ for each type a. The graph is the symmetric union of four-nearest-neighbor relations ($k = \min\{4, N(X) - 1\}$), including all ties at the kth distance. For unmarked data, the signals are counts on $g \times g$ grids for $g \in \{8, 16, 32\}$; units include empty cells and edges join cells sharing a side. Let $\mathcal{H}$ index types for marked data or grid resolutions for unmarked data, and let $\overline{I}_h^{\mathcal{D}}$ average the corresponding Moran coefficient equally over configurations on which it is defined. We report

$$\mathrm{RMSE}_{\mathrm{Moran}} = \left[\frac{1}{|\mathcal{H}|} \sum_{h \in \mathcal{H}} (\overline{I}_h^{\mathcal{G}} - \overline{I}_h^{\mathcal{R}})^2 \right]^{1/2}. \tag{106}$$

Zero-variance signals or graphs with no edges have undefined coefficients and are excluded, rather than assigned zero. All reference profiles satisfy the 80% validity threshold used to select $\mathcal{H}$. All generated profiles are fully valid except for EBP: its minimum valid fractions across types and seeds are 72.07% on three-type Strauss and 88.09% on Gibbs, so these entries compare conditional means over valid configurations. Marked-data rankings refer to the specified four-neighbor graph.

F.3 Interaction and conditional intensity recovery

Pair function recovery. We compare $\widehat{\phi}_{ab}(r) = \sum_k w_{kab}(1)\rho_k(r)$ with Strauss truth $\log \Gamma_{ab}\, \mathbf{1}\{r < 0.1\}$, or $\log(0.4)\, \mathbf{1}\{r < 0.1\}$ for unmarked data. This compares functions despite their different parameterizations. RMSE uses 1,501 uniform radii from zero to 0.3, computed per unordered mark pair and then averaged. Gibbs comparisons exclude $r < 0.018$ and assess only the finite local term ϕ_{ab}, excluding hard-core and count potentials.

Exploratory chain bootstrap ranges. Twelve refits per dataset resample whole training chains with multiplicity, redrawing 128 integration queries per configuration while fixing radial family and ridge weight. Unequal Gibbs chain lengths vary bootstrap sample sizes. All 36 refits converge with objective-gradient infinity norms below 10^{-3}; the maximum is 3.22×10^{-6}. Pointwise percentile ranges reflect chain and quadrature resampling and remain exploratory with 12 resamples.

Gibbs intensity and admissible insertions. With $C(N) = -0.005(N - 80)^2$, admissible Gibbs insertions have log intensity $\alpha_a + C(N + 1) - C(N) + \sum_i \phi_{am_i}(\|u - x_i\|)$. Inserting a type-$b$ context point changes this by $\phi_{ab}(r) + C(N + 2) - 2C(N + 1) + C(N) = \phi_{ab}(r) - 0.01$, distinct from the local pair term. Forbidden insertions have zero intensity and are excluded from finite log error calculations.

Conditional intensity probes. Each test window uses 64 uniform queries, 64 near-neighbor queries, and up to 32 observed points sampled without replacement for leave-one-out evaluation. Near-neighbor queries select a point uniformly, then a uniform angle and a radius uniform in $[0, 0.12]$, reflecting into the square. Spatial queries evaluate all marks; leave-one-out queries use the observed mark. True Gibbs intensities include count and hard-core terms. Errors average queries and marks within windows, then windows equally; log-RMSE takes the square root last and excludes inadmissible Gibbs queries.

We also evaluate Poisson divergence, which remains finite on forbidden insertions:

$$d(\widehat{\lambda}, \lambda^*) = \widehat{\lambda} - \lambda^* + \lambda^* \log(\lambda^*/\widehat{\lambda}), \qquad d(\widehat{\lambda}, 0) = \widehat{\lambda}. \tag{107}$$

Table 4 averages Poisson divergence over the specified queries and equally over configurations. Near-neighbor and leave-one-out probes stress different contexts. Forbidden-query fractions depend on the reference and query distribution; predicted intensity there measures support mismatch, since the fitted continuous intensity has no exact hard core.

Table 4: Selected-endpoint recovery. Pair RMSE is the mean of per-pair uniform-radius RMSEs. Log intensity RMSE is restricted to admissible queries for Gibbs; the Poisson divergence includes forbidden queries. These are single shared-endpoint estimates.

Dataset	Query distribution	Pair RMSE	Log intensity RMSE	Poisson divergence
Unmarked Strauss	Uniform	0.051	0.148	0.432
	Near neighbor	—	0.156	0.340
	Leave one out	—	0.124	0.576
Three-type Strauss	Uniform	0.060	0.147	0.139
	Near neighbor	—	0.158	0.127
	Leave one out	—	0.116	0.171
Four-type Gibbs	Uniform	0.068	0.161	0.341
	Near neighbor	—	0.191	0.449
	Leave one out	—	0.157	0.317

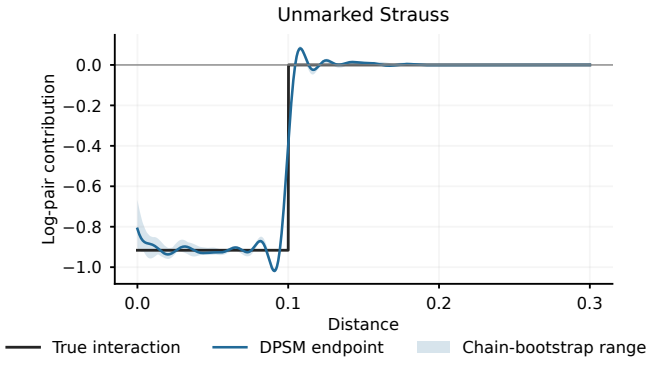

Figure 3: Unmarked Strauss pair function recovery with the same bootstrap convention as Figure 1. The discontinuous repulsive interaction is approximated by a smooth radial function.

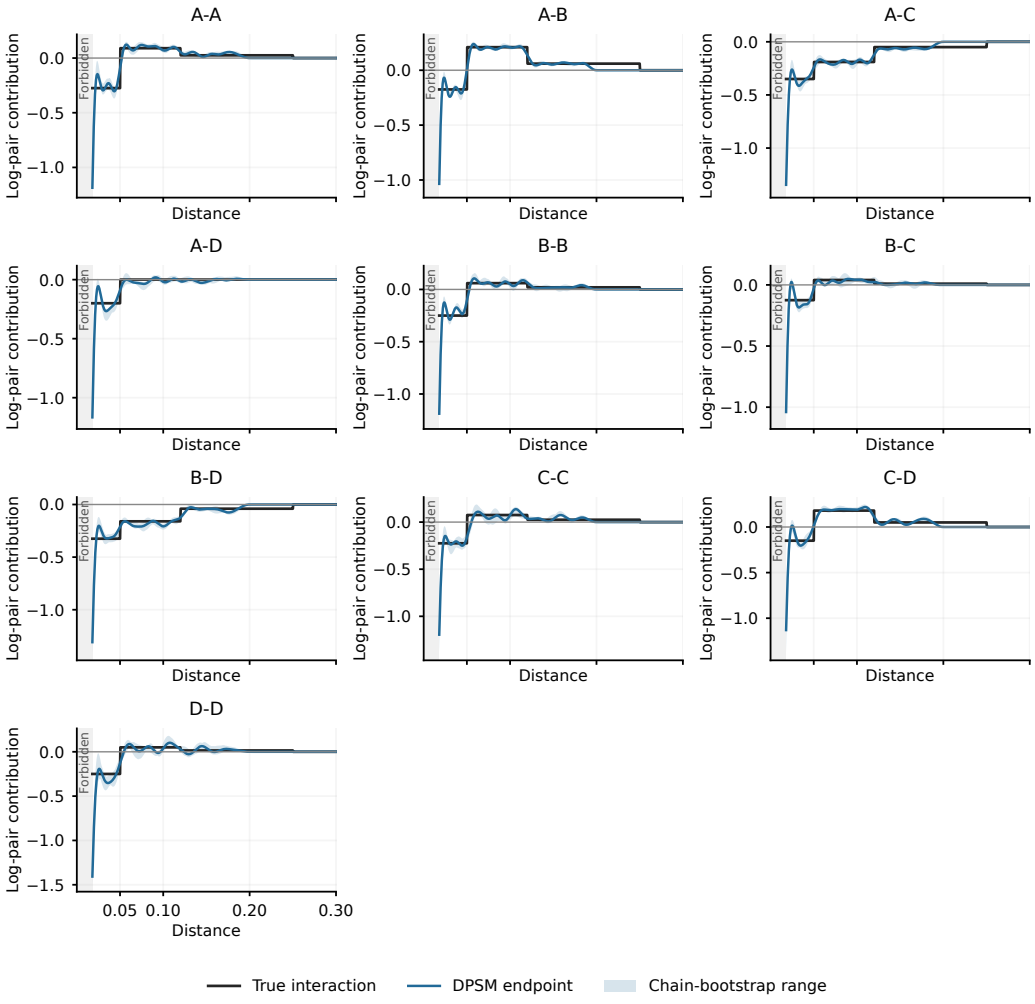

Figure 4: All distinct local pair functions of the four-type Gibbs simulator and the fitted DPSM endpoint, restricted to admissible separations $r \geq 0.018$. Gray marks the forbidden region, which is not assigned a finite ground-truth value. The global count contribution is assessed in the intensity probes. Blue shading denotes exploratory pointwise chain bootstrap ranges.

F.4 Complete spatial generation comparison

Table 5 adds unscaled nearest-neighbor distance, enrichment, and LGCP to the main comparison. DPSM has the lowest displayed mean errors on all three datasets.

Table 5: Complete spatial comparison (lower is better): mean ± sample SD over three seeds. Bold marks the lowest mean. All means and SDs are reported in units of 10^{-2}. Metrics are defined in Appendix F.2; neighbor mark RMSE is omitted for unmarked data.

Method	NN W_1	Scaled NN W_1	Neighbor mark RMSE	Spectral error	Enrichment RMSE	Moran RMSE
Unmarked Strauss Dataset						
DPSM	$\mathbf{0.199}_{\pm 0.057}$	$\mathbf{0.638}_{\pm 0.171}$	—	$\mathbf{1.066}_{\pm 0.284}$	$\mathbf{9.786}_{\pm 3.091}$	$\mathbf{0.383}_{\pm 0.184}$
PSD	$1.024_{\pm 0.139}$	$9.617_{\pm 0.298}$	—	$14.844_{\pm 0.355}$	$72.385_{\pm 1.697}$	$6.673_{\pm 0.088}$
EBP	$1.669_{\pm 0.301}$	$12.356_{\pm 2.229}$	—	$22.492_{\pm 5.529}$	$96.000_{\pm 17.988}$	$12.456_{\pm 5.193}$
SI-VAE	$0.405_{\pm 0.020}$	$2.945_{\pm 0.247}$	—	$4.123_{\pm 1.080}$	$48.178_{\pm 2.340}$	$1.097_{\pm 0.473}$
SetVAE	$1.705_{\pm 0.020}$	$12.587_{\pm 0.152}$	—	$16.522_{\pm 0.177}$	$92.781_{\pm 1.340}$	$7.678_{\pm 0.094}$
LGCP	$1.572_{\pm 0.023}$	$11.471_{\pm 0.154}$	—	$17.425_{\pm 0.155}$	$97.494_{\pm 0.994}$	$5.722_{\pm 0.294}$
Three-type Strauss Dataset						
DPSM	$\mathbf{0.098}_{\pm 0.039}$	$\mathbf{0.382}_{\pm 0.253}$	$\mathbf{1.889}_{\pm 0.537}$	$\mathbf{1.128}_{\pm 0.161}$	$\mathbf{11.952}_{\pm 2.298}$	$\mathbf{0.404}_{\pm 0.129}$
PSD	$1.270_{\pm 0.201}$	$7.716_{\pm 0.211}$	$6.188_{\pm 0.196}$	$14.043_{\pm 0.314}$	$30.187_{\pm 2.479}$	$3.006_{\pm 0.202}$
EBP	$1.707_{\pm 0.863}$	$13.122_{\pm 6.652}$	$20.588_{\pm 11.000}$	$36.429_{\pm 28.878}$	$41.761_{\pm 14.864}$	$7.033_{\pm 5.456}$
SI-VAE	$0.386_{\pm 0.016}$	$2.908_{\pm 0.112}$	$2.201_{\pm 0.200}$	$5.427_{\pm 0.210}$	$18.170_{\pm 4.624}$	$1.049_{\pm 0.134}$
SetVAE	$1.846_{\pm 0.085}$	$14.210_{\pm 0.610}$	$11.586_{\pm 1.228}$	$20.191_{\pm 1.100}$	$36.133_{\pm 2.640}$	$4.197_{\pm 0.704}$
LGCP	$1.103_{\pm 0.034}$	$8.428_{\pm 0.156}$	$8.432_{\pm 0.344}$	$14.922_{\pm 0.163}$	$46.088_{\pm 2.471}$	$3.398_{\pm 0.122}$
Four-type Gibbs Dataset						
DPSM	$\mathbf{0.187}_{\pm 0.184}$	$\mathbf{0.326}_{\pm 0.016}$	$\mathbf{1.539}_{\pm 0.484}$	$\mathbf{1.386}_{\pm 0.316}$	$\mathbf{8.307}_{\pm 0.815}$	$\mathbf{0.388}_{\pm 0.186}$
PSD	$0.483_{\pm 0.049}$	$4.573_{\pm 0.236}$	$3.185_{\pm 0.323}$	$8.404_{\pm 0.152}$	$16.781_{\pm 1.990}$	$1.760_{\pm 0.116}$
EBP	$1.164_{\pm 0.244}$	$10.514_{\pm 2.236}$	$13.790_{\pm 9.566}$	$39.133_{\pm 17.751}$	$20.805_{\pm 8.344}$	$1.991_{\pm 1.093}$
SI-VAE	$0.349_{\pm 0.141}$	$1.297_{\pm 0.212}$	$2.101_{\pm 1.350}$	$3.694_{\pm 1.174}$	$10.447_{\pm 1.253}$	$0.863_{\pm 0.168}$
SetVAE	$1.379_{\pm 0.175}$	$12.335_{\pm 1.522}$	$12.602_{\pm 7.842}$	$21.846_{\pm 1.434}$	$18.925_{\pm 1.009}$	$0.989_{\pm 0.093}$
LGCP	$0.611_{\pm 0.005}$	$5.361_{\pm 0.039}$	$2.882_{\pm 0.068}$	$10.496_{\pm 0.420}$	$21.154_{\pm 0.506}$	$1.083_{\pm 0.285}$

F.5 Configuration statistics and support

Joint oracle summary distribution. Summary vectors contain mark counts and unordered pair counts within true interaction bands; Gibbs adds quadratic count and forbidden-pair statistics. Dimensions are two, nine, and 36 for unmarked Strauss, three-type Strauss, and Gibbs. Features are divided by training standard deviations, lower-bounded by one. Oracle summary SW1 averages Wasserstein discrepancies over 32 fixed unit-normalized Gaussian projections. It compares selected sufficient statistic distributions within each dataset.

Table 6 adds count Wasserstein and mean composition L^1 errors. Independent-location controls uniformly randomize test coordinates while preserving marks, counts, and population size, hence have zero count/composition errors. DPSM has the lowest displayed mean summary SW1, including on Gibbs (0.320 ± 0.165 versus 0.604 ± 0.303 for SI-VAE), where it also has lower count error than SI-VAE.

Hard-core support. Gibbs forbids pairs below 0.018. Table 7 reports violating configurations, pairs per configuration, and pooled pair fractions after final correction. DPSM violates support in 21.81% of configurations, versus 92.45% for PSD, 85.83% for SI-VAE.

F.6 Reference variation

Conditional ranges use 200 resamples of whole test chains and generated configurations within each seed, retaining source chain blocks for independent-location controls. The 2.5%–97.5% percentiles condition on fitted checkpoints. Across 50 partitions of test chains into disjoint halves, median reference SW1 is 0.101, 0.109, and 0.103 for unmarked Strauss, three-type Strauss, and Gibbs. These half-sized populations describe reference variation.

Table 6: Configuration diagnostics: mean ± sample SD over three seeds; lower is better. Oracle summary SW_1 is comparable only within datasets. Independent locations preserves test counts and marks, giving zero count and composition errors. Bold marks the lowest mean among the listed methods and controls.

Method	Summary SW_1	Count W_1	Mean composition L^1
Unmarked Strauss Dataset			
DPSM	$0.390_{\pm 0.093}$	$2.099_{\pm 0.630}$	—
PSD	$1.656_{\pm 0.290}$	$7.185_{\pm 3.675}$	—
EBP	$3.295_{\pm 0.725}$	$0.513_{\pm 0.196}$	—
SI-VAE	$0.734_{\pm 0.019}$	$0.779_{\pm 0.416}$	—
SetVAE	$3.175_{\pm 0.054}$	$0.456_{\pm 0.081}$	—
LGCP	$2.552_{\pm 0.045}$	$1.904_{\pm 0.304}$	—
Three-type Strauss Dataset			
DPSM	$0.168_{\pm 0.056}$	$0.966_{\pm 0.222}$	$0.040_{\pm 0.017}$
PSD	$1.764_{\pm 0.404}$	$5.936_{\pm 3.666}$	$0.027_{\pm 0.017}$
EBP	$7.184_{\pm 6.697}$	$0.317_{\pm 0.038}$	$0.554_{\pm 0.307}$
SI-VAE	$0.338_{\pm 0.082}$	$1.202_{\pm 1.661}$	$0.010_{\pm 0.005}$
SetVAE	$2.687_{\pm 0.203}$	$0.499_{\pm 0.078}$	$0.254_{\pm 0.006}$
LGCP	$1.321_{\pm 0.059}$	$2.329_{\pm 0.120}$	$0.008_{\pm 0.004}$
Four-type Gibbs Dataset			
DPSM	$0.320_{\pm 0.165}$	$5.548_{\pm 5.284}$	$0.045_{\pm 0.012}$
PSD	$0.595_{\pm 0.057}$	$6.897_{\pm 5.640}$	$0.036_{\pm 0.009}$
EBP	$2.298_{\pm 1.463}$	$0.443_{\pm 0.015}$	$0.465_{\pm 0.393}$
SI-VAE	$0.604_{\pm 0.303}$	$7.041_{\pm 3.698}$	$0.031_{\pm 0.022}$
SetVAE	$1.569_{\pm 0.642}$	$0.674_{\pm 0.215}$	$0.430_{\pm 0.282}$
LGCP	$0.644_{\pm 0.034}$	$1.998_{\pm 0.173}$	$0.015_{\pm 0.003}$

Table 7: Hard-core violations on the Four-type Gibbs Dataset in the final generated populations. Values show mean and sample standard deviation over three seeds. The held-out reference contains no violations. Lower is better; bold marks the lowest mean. For pair fractions, each power of ten applies to the mean and SD.

Method	Violating config. fraction	Pairs / config.	Pair fraction
DPSM	$\mathbf{0.2181_{\pm 0.0536}}$	$\mathbf{0.242_{\pm 0.062}}$	$\mathbf{6.49_{\pm 0.75} \times 10^{-5}}$
PSD	$0.9245_{\pm 0.0056}$	$2.816_{\pm 0.434}$	$1.022_{\pm 0.017} \times 10^{-3}$
EBP	$0.9928_{\pm 0.0011}$	$5.051_{\pm 0.691}$	$1.5457_{\pm 0.2252} \times 10^{-3}$
SI-VAE	$0.8583_{\pm 0.0139}$	$2.304_{\pm 0.246}$	$5.918_{\pm 0.160} \times 10^{-4}$
SetVAE	$0.9922_{\pm 0.0020}$	$4.767_{\pm 0.344}$	$1.4563_{\pm 0.1001} \times 10^{-3}$
LGCP	$0.9622_{\pm 0.0049}$	$3.690_{\pm 0.134}$	$1.1169_{\pm 0.0336} \times 10^{-3}$
Independent locations	$0.9596_{\pm 0.0126}$	$3.165_{\pm 0.037}$	$9.728_{\pm 0.114} \times 10^{-4}$

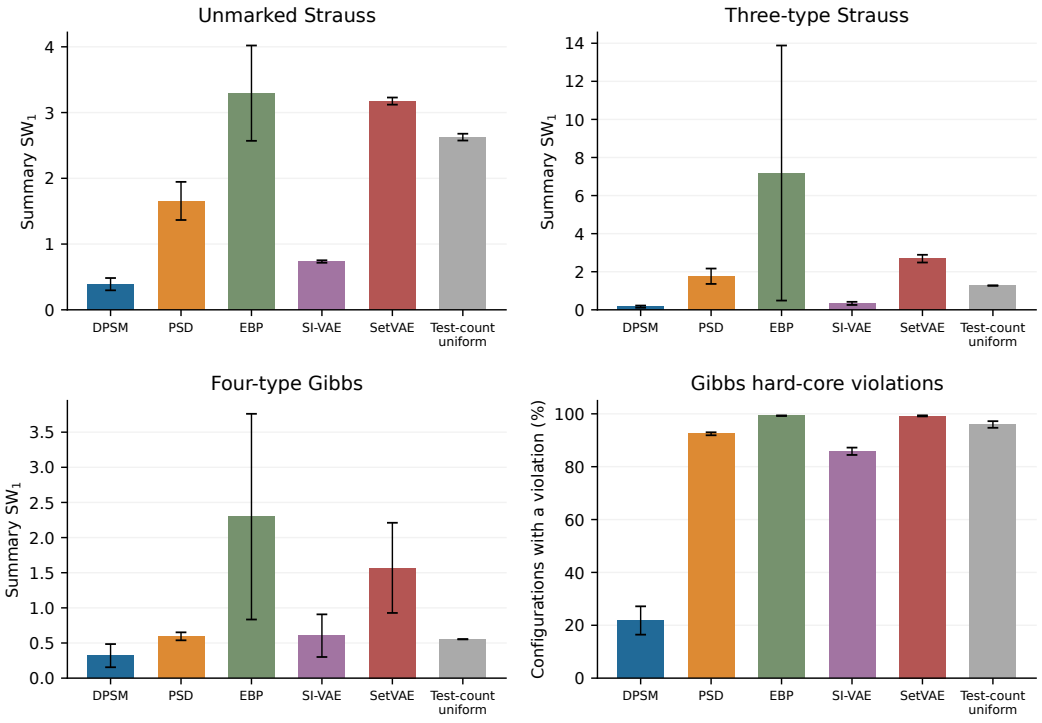

Figure 5: Interaction summary discrepancies and Gibbs support violations. Error bars: sample SD over three seeds. Compare SW1 within panels only. Test-count uniform is the Independent locations control in Table 6, preserving test counts and marks.

Oracle-statistic diagnostics include split rank-normalized/folded $\widehat{R}$, autocorrelations through lag ten, and a within-chain initial-positive monotone lag-pair ESS approximation. Chains are truncated to the shortest within each split. Maximum test $\widehat{R}$ values are 1.004, 1.020, and 1.041 for the three datasets; Gibbs validation reaches 1.060. Minimum test ESS fractions are 1.000, 0.827, and 0.608.

F.7 Noisy insertion field fit diagnostics

Table 8 evaluates the held-out negative pseudolikelihood difference from the Poisson reference intensity. For a corrupted configuration $Y = X_q$, we compute

$$\widehat{\Delta}_q(Y) = \frac{1}{\max(1, \bar{N}_\nu)} \left[\frac{|W|}{Q} \sum_{j=1}^{Q} \sum_{a} \{\lambda_{q,\theta}(V_j, a; Y) - \nu_a\} - \sum_{(u,a)\in Y} \log \frac{\lambda_{q,\theta}(u, a; Y \setminus \{(u,a)\})}{\nu_a} \right]. \tag{108}$$

Here $Q = 128$, $V_j \sim \mathrm{Unif}(W)$, and $\bar{N}_\nu = |W| \sum_a \nu_a$ is the expected reference count (greater than one in all datasets). The integral estimate sums over all marks; every observed point in the corrupted configuration supplies a leave-one-out term. At each of 11 retentions, each checkpoint's value averages equally over 128 test configurations and two independent corruptions. The table reports the mean and sample SD ($\mathrm{ddof} = 1$) across three checkpoints per dataset. Configurations, corruptions, and queries are shared across checkpoints. Lower values indicate improved held-out noisy pseudolikelihood at a fixed retention. NPL values are close to the Poisson reference at high noise and improve relative to it as retention increases.

Table 8: Held-out negative pseudolikelihood difference from the Poisson reference intensity, normalized by the expected reference count. Negative values favor the learned insertion field. Values are means with sample standard deviations across three checkpoints, after averaging two corruptions of 128 test configurations.

Retention	Unmarked Strauss Dataset	Three-type Strauss Dataset	Four-type Gibbs Dataset
0.00	$0.000000_{\pm 0.000000}$	$0.000000_{\pm 0.000000}$	$0.000000_{\pm 0.000000}$
0.10	$0.000924_{\pm 0.002041}$	$0.001338_{\pm 0.000386}$	$0.000586_{\pm 0.000544}$
0.35	$0.000721_{\pm 0.000354}$	$0.001856_{\pm 0.001108}$	$0.010373_{\pm 0.004618}$
0.50	$-0.008763_{\pm 0.001121}$	$-0.006092_{\pm 0.001963}$	$0.005548_{\pm 0.001810}$
0.65	$-0.039609_{\pm 0.000738}$	$-0.027793_{\pm 0.000768}$	$0.000890_{\pm 0.002742}$
0.80	$-0.102574_{\pm 0.000327}$	$-0.084463_{\pm 0.001635}$	$-0.012054_{\pm 0.001800}$
0.90	$-0.180428_{\pm 0.000430}$	$-0.160171_{\pm 0.001203}$	$-0.048196_{\pm 0.000751}$
1.00	$-0.296534_{\pm 0.000000}$	$-0.278220_{\pm 0.000000}$	$-0.124348_{\pm 0.000000}$

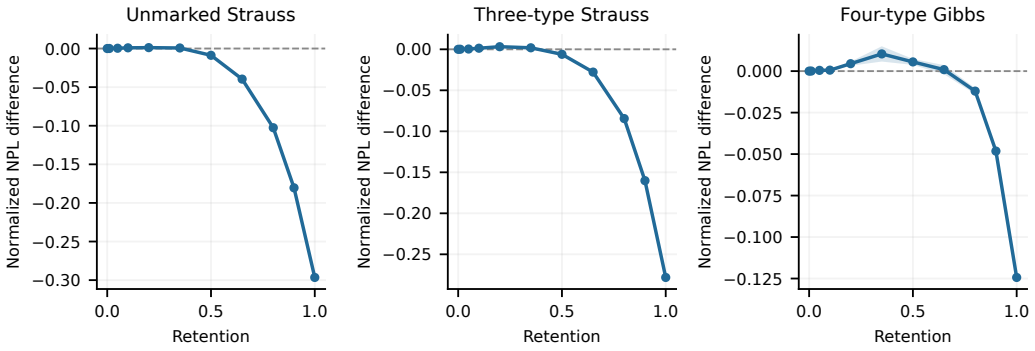

Figure 6: Held-out noisy negative pseudolikelihood relative to the Poisson reference (zero); negative values favor DPSM.

G ADDITIONAL SPATIAL STRUCTURE RESULTS

This appendix gives the datasets, protocols, metrics, and supplementary results for Section 5.2.

G.1 DATA-GENERATING PROCESSES AND CONFIGURATION SPLITS

Observations are unordered configurations on $[0,1]^2$ with marks $m \in \{A, B, \text{background}\}$. Each conditional model receives the dataset label and is fitted jointly on all six datasets.

Clustered patterns and hollow clusters. Six nominal cluster centers are $c_j = (1/2, 1/2) + 0.285(\cos\theta_j, \sin\theta_j)$, with $\theta_j = 0.18 + 2\pi j/6$ and $j = 0, \ldots, 5$. Per configuration, a shared rotation $\omega \sim \mathcal{N}(0, 0.035^2)$ about $(1/2, 1/2)$ and independent $\mathcal{N}(0, 0.007^2 I_2)$ perturbations give centers $\widetilde{c}_j$. A rounded draw from $\mathcal{N}(1500, 210^2)$ is clamped to $[1000, 2000]$; $N_{\text{bg}} \sim \text{Binomial}(N, 0.025)$ points are reserved for spatial background. Allocate the rest multinomially with $\text{Dirichlet}(45, \ldots, 45)$ cluster weights. Proposals are $x = \widetilde{c}_j + 0.055z$, $z \sim \mathcal{N}(0, I_2)$. With $r = \|z\|$, clustered patterns allow $r \geq 0$ and have radial mark boundaries $(0.62, 1.18)$; hollow clusters reject $r < 0.90$ and use $(1.25, 1.65)$. Both reject points within 0.09 of either boundary or outside the square. The three radial regions have nominal marks A, B, and background, retained with probability 0.96; each alternative has probability 0.02. Spatial background locations are uniform, excluding the six disks of radius 0.055×0.90 about perturbed centers for hollow clusters. Their marks are B with probability 0.18 and background otherwise, so spatial background allocation differs from final background-mark count.

Double spirals. Draw a $\text{Poisson}(1500)$ count clamped to $[1000, 2000]$, with $N_{\text{bg}} \sim \text{Binomial}(N, 0.04)$ uniform background points. Each foreground point draws arm $a \sim \text{Bernoulli}(1/2)$ and progress $u \sim \text{Uniform}(0, 1)$:

$$\theta = 0.35 + 3.65\pi u + \pi a + \omega, \qquad r = 0.045 + 0.365u + \epsilon_r,$$
$$x = c + r(\cos\theta, \sin\theta) + \epsilon_x.$$

The shared rotation is $\omega \sim \mathcal{N}(0, 0.025^2)$, the shared center $c \sim \mathcal{N}((1/2, 1/2), 0.006^2 I_2)$, and pointwise noises $\epsilon_r \sim \mathcal{N}(0, 0.009^2)$ and $\epsilon_x \sim \mathcal{N}(0, 0.006^2 I_2)$. Arms 0 and 1 carry marks A and B without mark noise; background points carry the background mark. All final coordinates are clamped to $[10^{-5}, 1 - 10^{-5}]^2$.

Annular domains and elliptic domains. These datasets and paired microcolonies draw $N \sim \text{Poisson}(1500)$ without truncation, with $N_{\text{bg}} \sim \text{Binomial}(N, 0.05)$ independent uniform points carrying the background mark. Foreground locations are uniform in area on the annulus $\{x : 0.26 \leq \|x - (1/2, 1/2)\| \leq 0.46\}$ or inside the ellipse centered at $(1/2, 1/2)$ with horizontal and vertical semiaxes 0.40 and 0.27. Independently per configuration, draw $\phi \sim \text{Uniform}(0, 2\pi)$. Negative

and nonnegative values of $(x - (1/2, 1/2))^\top(\cos\phi, \sin\phi)$ assign marks A and B, followed by independent flips with probability 0.015.

Paired microcolonies. Pair centers are $(0.28, 0.28)$, $(0.72, 0.28)$, $(0.28, 0.72)$, and $(0.72, 0.72)$, each with colonies displaced horizontally by -0.06 and $+0.06$. Each foreground point chooses one of eight colonies uniformly and receives isotropic Gaussian displacement with standard deviation 0.025; out-of-square proposals are resampled. Each pair independently draws a fair binary state assigning opposite nominal marks to its colonies, followed by independent mark flips with probability 0.015. The four pair assignments give $2^4 = 16$ nominal mark configurations. Averaging over hidden states in all three relational datasets gives equal foreground mark probabilities at each location, despite coherent organization within individual configurations.

Splits and reproducibility. Each dataset has 144 training, 16 validation, and 32 test configurations (pooled sizes 864, 96, and 192), all independently simulated. The first two datasets retain the first 192 of 512 source configurations; spirals retain all 192. Indices 0–143, 144–159, and 160–191 define the splits without shuffling. Source seeds are 20260902 for clusters and 20260905 for spirals, with a separate NumPy `SeedSequence` child per configuration. For the three relational datasets, training, validation, and test streams use seeds 907601, 907602, and 907603 through `SeedSequence([split seed, dataset ID])`, with IDs 3, 4, and 5. All methods share these archives, containing only coordinates, marks, offsets, and condition labels.

Model architecture, fitting, and generation settings are given in Appendix E.2; baseline settings are summarized in Appendix E.5.

G.2 Evaluation metrics

For dataset c, let $X = \{(x_i, m_i)\}_{i=1}^N$, with $m_i \in \mathcal{M} = \{A, B, \text{background}\}$. Write $N_m = \sum_i \mathbf{1}\{m_i = m\}$ and $N_F = N_A + N_B$. Geometry and mark scores use foreground points; count and composition use all marks. Thresholds and normalizations are fixed using training data only, and landmarks follow nominal geometry. Define $[u]_+ = \max(u, 0)$.

Geometric precision (Precision). Precision measures leakage outside the intended geometry:

$$\text{Precision}(X) = \frac{1}{\max(N_F, 1)} \sum_{i:m_i \in \{A,B\}} \mathbf{1}\{d_c(x_i) \le \tau_c\}.$$

Here $d_c(x)$ is distance outside the nearest nominal disk or annulus, or distance to the nearest sampled spiral curve. For elliptic domains, $d_c(x) = 0.27[\|v(x)\|_2 - 1]_+$, where $v(x) = ((x_1 - 0.5)/0.40, (x_2 - 0.5)/0.27)$. Microcolony disks have radius $0.025\sqrt{-2\log(0.05)}$. Tolerance τ_c is the training 95th percentile of deviations, weighting configurations equally and foreground points equally within each. It is zero for annular domains, elliptic domains, and microcolonies.

Geometric coverage (Coverage). Coverage detects missing components or gaps while limiting the influence of point count and duplicate locations. Let $\{z_{c\ell}\}_{\ell=1}^{384}$ be fixed landmarks with equal component weights. Draw five subsamples S_b, each containing at most 256 distinct foreground locations sampled without replacement:

$$\text{Coverage}(X) = \frac{1}{5 \cdot 384} \sum_{b=1}^{5} \sum_{\ell=1}^{384} \mathbf{1}\left\{ \min_{x \in S_b} \|z_{c\ell} - x\|_2 \le r_c \right\}.$$

All locations are retained when fewer than 256 are available; empty foregrounds have zero coverage. Radius r_c is the training 95th percentile of landmark-to-subsample distances, using equal configuration, subsample, and landmark weights and the upper observed order statistic.

Mark alignment (Alignment). Alignment tests whether foreground marks match their intended spatial roles in clustered patterns, hollow clusters, and spirals:

$$\text{Alignment}(X) = \frac{1}{2} \sum_{m \in \{A,B\}} \frac{\sum_{i:m_i=m} \mathbf{1}\{g_c(x_i) = m\}}{\max(N_m, 1)}.$$

The evaluation-only map $g_c(x)$ gives the nearest cluster's nominal radial layer mark or the nearest spiral's identity. Absent marks contribute zero, and all foreground points enter regardless of their precision score.

Mark organization (Organization). Organization tests coordinated marks within a configuration while allowing different valid partitions or colony pair assignments. For annular and elliptic domains, let a^* be the maximum balanced A/B accuracy over 720 central partition axes $\theta_j = j\pi/720$, $j = 0, \ldots, 719$, and both label assignments. Balanced accuracy averages the two per-mark correct-classification fractions. Define

$$h = [2a^* - 1]_+, \qquad b = \frac{2\min(N_A, N_B)}{\max(N_F, 1)}, \qquad \text{Organization}(X) = hb.$$

The score is zero if either mark is absent and is invariant to global A/B reversal. For microcolonies, assign foreground points to their nearest nominal colony. Let n_k be its foreground count and q_k its A fraction, setting $q_k = 1/2$ when $n_k = 0$. Then

$$\text{Organization}(X) = \frac{1}{4}\sum_{j=1}^{4} |q_{2j-1} - q_{2j}|.$$

This rewards opposite marks within each pair. Alignment and organization have different chance levels and are reported separately.

Configuration pass rate (Pass rate). Pass rate measures how often an entire configuration meets all organization requirements, complementing average scores. Annular and elliptic configurations require $h \geq 0.8$ and $b \geq 0.8$. Microcolonies require $n_k \geq 10$ for every colony and $|q_{2j-1} - q_{2j}| \geq 0.5$ for every pair. For generated population $\mathcal{G}$,

$$\text{PassRate} = \frac{1}{|\mathcal{G}|}\sum_{X \in \mathcal{G}} \mathbf{1}\{X \text{ satisfies all criteria}\}.$$

Count and composition Wasserstein distances. These metrics test variation in total point count and mark proportions. For generated and test populations $\mathcal{G}$ and $\mathcal{T}$, let $\widehat{P}^f_{\mathcal{D}} = |\mathcal{D}|^{-1}\sum_{X\in\mathcal{D}} \delta_{f(X)}$, where δ_t is a unit point mass. With $p_m(X) = N_m/\max(N, 1)$ and $\overline{N}_{\mathcal{T}} = |\mathcal{T}|^{-1}\sum_{X\in\mathcal{T}} N(X)$, define

$$\text{Rel. count } W_1 = \frac{W_1(\widehat{P}^N_{\mathcal{G}}, \widehat{P}^N_{\mathcal{T}})}{\overline{N}_{\mathcal{T}}},$$

$$\text{Composition } W_1 = \frac{1}{2}\sum_{m\in\mathcal{M}} W_1(\widehat{P}^{p_m}_{\mathcal{G}}, \widehat{P}^{p_m}_{\mathcal{T}}).$$

Here W_1 is the one-dimensional Wasserstein distance.

Configuration summary distance (Summary SW_1). Summary SW_1 tests the joint distribution of selected geometric and mark statistics beyond their individual means. Let $\phi_c(X) \in \mathbb{R}^d$ be the dataset-specific summary. For the first three datasets, it includes foreground coordinate means/covariances, component weights and centroids, radial means/standard deviations, and mean sine/cosine of angular or curve parameters. For annular and elliptic domains, it includes foreground coordinate means/covariances, radial means/standard deviations, all-mark proportions, and 32-bin angular and signed mark occupancy. Microcolonies additionally include eight colony weights and eight signed weights. Empty components contribute zeros; signed occupancy is the A-minus-B count divided by $\max(N_F, 1)$.

Using training means μ_{cj} and sample standard deviations s_{cj}, define

$$\widetilde{\phi}_{cj}(X) = \frac{\phi_{cj}(X) - \mu_{cj}}{\max(s_{cj}, \epsilon_c)}, \qquad \epsilon_c = \max\left\{10^{-4}, 0.1\ \text{median}_{j:s_{cj}>10^{-8}}\ s_{cj}\right\}.$$

For 128 fixed unit-normalized Gaussian directions $\theta_\ell \in \mathbb{R}^d$, let $f_\ell(X) = \theta_\ell^\top \widetilde{\phi}_c(X)$. Then

$$\text{Summary } \text{SW}_1 = \frac{1}{128}\sum_{\ell=1}^{128} W_1(\widehat{P}^{f_\ell}_{\mathcal{G}}, \widehat{P}^{f_\ell}_{\mathcal{T}}).$$

Table 9: Complete per-dataset evaluation: position-dependent datasets. Values are means with sample standard deviations over three independent fits, each generating 32 configurations per dataset. Mark alignment is reported for these position-dependent datasets. Metric definitions and aggregation follow Appendix G.2. Bold marks the best model mean.

Method	Precision ↑	Coverage ↑	Alignment ↑	Composition W_1 ↓	Rel. count W_1 ↓	Summary SW_1 ↓
Clustered patterns Dataset						
Reference	0.950	0.946	0.856	–	–	–
DPSM	$0.955_{\pm 0.010}$	$\mathbf{0.953_{\pm 0.004}}$	$0.868_{\pm 0.008}$	$0.018_{\pm 0.008}$	$\mathbf{0.033_{\pm 0.010}}$	$\mathbf{0.37_{\pm 0.04}}$
PSD	$0.900_{\pm 0.013}$	$0.938_{\pm 0.003}$	$0.687_{\pm 0.013}$	$\mathbf{0.009_{\pm 0.005}}$	$0.094_{\pm 0.021}$	$0.66_{\pm 0.06}$
EBP	$0.227_{\pm 0.036}$	$0.378_{\pm 0.064}$	$0.492_{\pm 0.018}$	$0.132_{\pm 0.058}$	$0.346_{\pm 0.000}$	$10.07_{\pm 0.89}$
SI-VAE	$0.336_{\pm 0.028}$	$0.455_{\pm 0.015}$	$0.503_{\pm 0.005}$	$0.267_{\pm 0.030}$	$0.452_{\pm 0.090}$	$6.33_{\pm 0.25}$
SetVAE	$\mathbf{0.956_{\pm 0.012}}$	$0.729_{\pm 0.029}$	$\mathbf{0.904_{\pm 0.002}}$	$0.333_{\pm 0.002}$	$0.039_{\pm 0.012}$	$2.67_{\pm 0.16}$
Hollow clusters Dataset						
Reference	0.953	0.953	0.825	–	–	–
DPSM	$\mathbf{0.943_{\pm 0.009}}$	$\mathbf{0.959_{\pm 0.002}}$	$\mathbf{0.822_{\pm 0.004}}$	$0.030_{\pm 0.003}$	$\mathbf{0.040_{\pm 0.002}}$	$\mathbf{0.38_{\pm 0.02}}$
PSD	$0.803_{\pm 0.033}$	$0.943_{\pm 0.006}$	$0.649_{\pm 0.019}$	$\mathbf{0.021_{\pm 0.010}}$	$0.170_{\pm 0.004}$	$0.78_{\pm 0.08}$
EBP	$0.308_{\pm 0.042}$	$0.450_{\pm 0.039}$	$0.501_{\pm 0.027}$	$0.262_{\pm 0.104}$	$0.314_{\pm 0.000}$	$10.85_{\pm 2.74}$
SI-VAE	$0.452_{\pm 0.018}$	$0.612_{\pm 0.110}$	$0.512_{\pm 0.016}$	$0.372_{\pm 0.099}$	$0.297_{\pm 0.165}$	$5.24_{\pm 0.04}$
SetVAE	$0.908_{\pm 0.020}$	$0.774_{\pm 0.030}$	$0.817_{\pm 0.030}$	$0.203_{\pm 0.003}$	$0.048_{\pm 0.015}$	$2.52_{\pm 0.17}$
Double spirals Dataset						
Reference	0.945	0.951	1.000	–	–	–
DPSM	$\mathbf{0.921_{\pm 0.005}}$	$\mathbf{0.944_{\pm 0.003}}$	$\mathbf{0.996_{\pm 0.001}}$	$0.017_{\pm 0.018}$	$0.049_{\pm 0.035}$	$\mathbf{0.69_{\pm 0.25}}$
PSD	$0.749_{\pm 0.034}$	$0.906_{\pm 0.008}$	$0.956_{\pm 0.021}$	$\mathbf{0.015_{\pm 0.001}}$	$0.141_{\pm 0.025}$	$0.88_{\pm 0.15}$
EBP	$0.433_{\pm 0.018}$	$0.634_{\pm 0.030}$	$0.536_{\pm 0.033}$	$0.056_{\pm 0.018}$	$0.330_{\pm 0.000}$	$6.90_{\pm 1.37}$
SI-VAE	$0.519_{\pm 0.021}$	$0.827_{\pm 0.016}$	$0.575_{\pm 0.034}$	$0.093_{\pm 0.009}$	$0.235_{\pm 0.044}$	$5.60_{\pm 1.25}$
SetVAE	$0.918_{\pm 0.025}$	$0.859_{\pm 0.033}$	$0.953_{\pm 0.016}$	$0.232_{\pm 0.002}$	$\mathbf{0.008_{\pm 0.002}}$	$7.83_{\pm 0.54}$

Feature maps and normalizations differ across datasets and from Appendix F.5; comparisons are therefore restricted to the same dataset.

Per-configuration scores are averaged equally within each population. Reported uncertainty is the sample standard deviation across three fits.

G.3 Complete per-dataset evaluation

Tables 9 and 10 give all per-dataset scores.

Table 10: Complete per-dataset evaluation: relational datasets. Values are means with sample standard deviations over three independent fits, each generating 32 configurations per dataset. Mark organization is reported for these relational datasets. Metric definitions and aggregation follow Appendix G.2. Bold marks the best model mean.

Method	Precision ↑	Coverage ↑	Organization ↑	Composition W_1 ↓	Rel. count W_1 ↓	Summary SW_1 ↓
Annular domains Dataset						
Reference	1.000	0.951	0.948	–	–	–
DPSM	$0.972_{\pm 0.002}$	$\mathbf{0.947_{\pm 0.001}}$	$\mathbf{0.896_{\pm 0.024}}$	$0.016_{\pm 0.010}$	$0.020_{\pm 0.004}$	$\mathbf{0.39_{\pm 0.06}}$
PSD	$\mathbf{0.972_{\pm 0.008}}$	$0.943_{\pm 0.004}$	$0.152_{\pm 0.161}$	$\mathbf{0.007_{\pm 0.001}}$	$0.047_{\pm 0.018}$	$0.59_{\pm 0.08}$
EBP	$0.549_{\pm 0.049}$	$0.551_{\pm 0.068}$	$0.488_{\pm 0.312}$	$0.071_{\pm 0.049}$	$0.331_{\pm 0.000}$	$8.56_{\pm 0.58}$
SI-VAE	$0.768_{\pm 0.050}$	$0.902_{\pm 0.006}$	$0.106_{\pm 0.014}$	$0.183_{\pm 0.027}$	$0.717_{\pm 0.010}$	$5.95_{\pm 0.83}$
SetVAE	$0.960_{\pm 0.029}$	$0.845_{\pm 0.020}$	$0.377_{\pm 0.057}$	$0.229_{\pm 0.002}$	$\mathbf{0.006_{\pm 0.002}}$	$5.30_{\pm 0.36}$
Elliptic domains Dataset						
Reference	1.000	0.950	0.950	–	–	–
DPSM	$0.982_{\pm 0.001}$	$\mathbf{0.947_{\pm 0.001}}$	$\mathbf{0.915_{\pm 0.012}}$	$0.015_{\pm 0.013}$	$0.021_{\pm 0.007}$	$\mathbf{0.34_{\pm 0.08}}$
PSD	$0.982_{\pm 0.001}$	$0.945_{\pm 0.002}$	$0.163_{\pm 0.159}$	$\mathbf{0.008_{\pm 0.005}}$	$0.047_{\pm 0.017}$	$0.41_{\pm 0.03}$
EBP	$0.757_{\pm 0.071}$	$0.650_{\pm 0.078}$	$0.412_{\pm 0.265}$	$0.027_{\pm 0.023}$	$0.337_{\pm 0.000}$	$5.11_{\pm 1.32}$
SI-VAE	$0.934_{\pm 0.045}$	$0.938_{\pm 0.009}$	$0.100_{\pm 0.006}$	$0.120_{\pm 0.035}$	$0.642_{\pm 0.062}$	$2.39_{\pm 0.68}$
SetVAE	$\mathbf{0.994_{\pm 0.002}}$	$0.856_{\pm 0.036}$	$0.386_{\pm 0.084}$	$0.222_{\pm 0.005}$	$\mathbf{0.006_{\pm 0.003}}$	$4.63_{\pm 0.64}$
Paired microcolonies Dataset						
Reference	0.956	0.949	0.955	–	–	–
DPSM	$\mathbf{0.946_{\pm 0.005}}$	$\mathbf{0.952_{\pm 0.004}}$	$\mathbf{0.856_{\pm 0.129}}$	$0.013_{\pm 0.009}$	$0.017_{\pm 0.007}$	$\mathbf{0.44_{\pm 0.05}}$
PSD	$0.946_{\pm 0.009}$	$0.948_{\pm 0.002}$	$0.079_{\pm 0.015}$	$\mathbf{0.009_{\pm 0.001}}$	$0.040_{\pm 0.016}$	$0.47_{\pm 0.04}$
EBP	$0.233_{\pm 0.070}$	$0.414_{\pm 0.083}$	$0.297_{\pm 0.064}$	$0.063_{\pm 0.042}$	$0.332_{\pm 0.000}$	$14.36_{\pm 3.11}$
SI-VAE	$0.208_{\pm 0.039}$	$0.135_{\pm 0.023}$	$0.290_{\pm 0.039}$	$0.597_{\pm 0.057}$	$0.914_{\pm 0.012}$	$21.28_{\pm 1.15}$
SetVAE	$0.937_{\pm 0.019}$	$0.818_{\pm 0.006}$	$0.112_{\pm 0.019}$	$0.225_{\pm 0.002}$	$\mathbf{0.009_{\pm 0.003}}$	$4.71_{\pm 0.53}$

G.4 INTERACTION ABLATION

For each selected checkpoint, we disable configuration interactions throughout the noise path, holding spatial conditions, initial noise, reverse seed, and sampling budget fixed. Table 11 reports organization and pass rates; Figure 7 shows configurations.

Table 11: Interaction ablation: means with sample standard deviations over three fits and 32 configurations per dataset and fit. Higher is better; bold marks the best model mean, excluding the reference. Pass criteria are defined in Appendix G.2.

	Annular domains		Elliptic domains		Paired microcolonies	
Method	Organization ↑	Pass rate ↑	Organization ↑	Pass rate ↑	Organization ↑	Pass rate ↑
Reference	0.948	1.000	0.950	1.000	0.955	1.000
DPSM	$\mathbf{0.896}_{\pm 0.024}$	$\mathbf{0.990}_{\pm 0.018}$	$\mathbf{0.915}_{\pm 0.012}$	$\mathbf{1.000}_{\pm 0.000}$	$\mathbf{0.856}_{\pm 0.129}$	$\mathbf{0.688}_{\pm 0.384}$
Without interactions	$0.084_{\pm 0.019}$	$0.000_{\pm 0.000}$	$0.065_{\pm 0.009}$	$0.000_{\pm 0.000}$	$0.047_{\pm 0.009}$	$0.000_{\pm 0.000}$

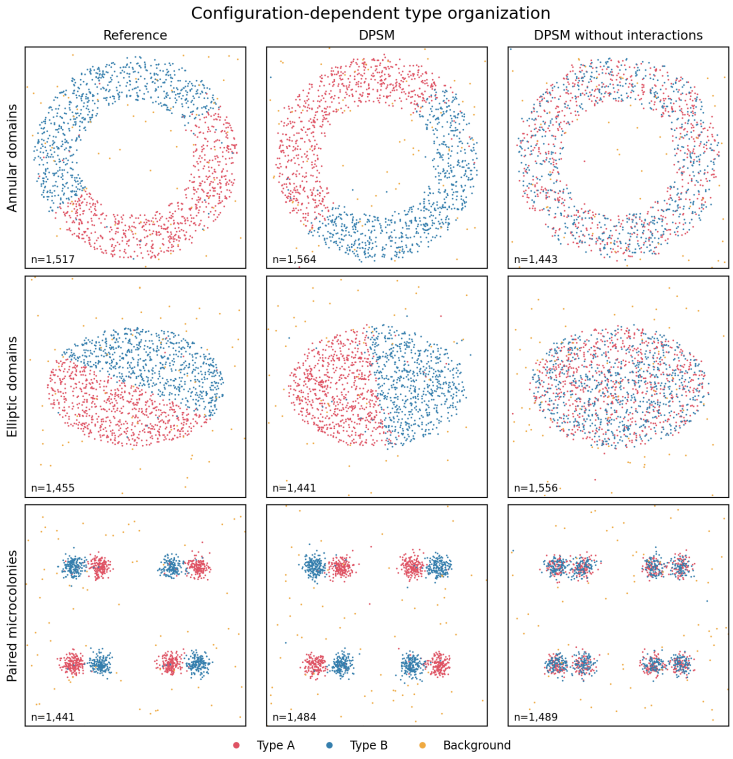

Figure 7: First saved configurations of the first listed fit, with matched spatial conditions and noise. Rows: annular domains, elliptic domains, and paired microcolonies. Disabling interactions retains spatial support but mixes foreground marks. Reference: first held-out configuration.

G.5 CONDITIONAL GENERATION TRAJECTORIES

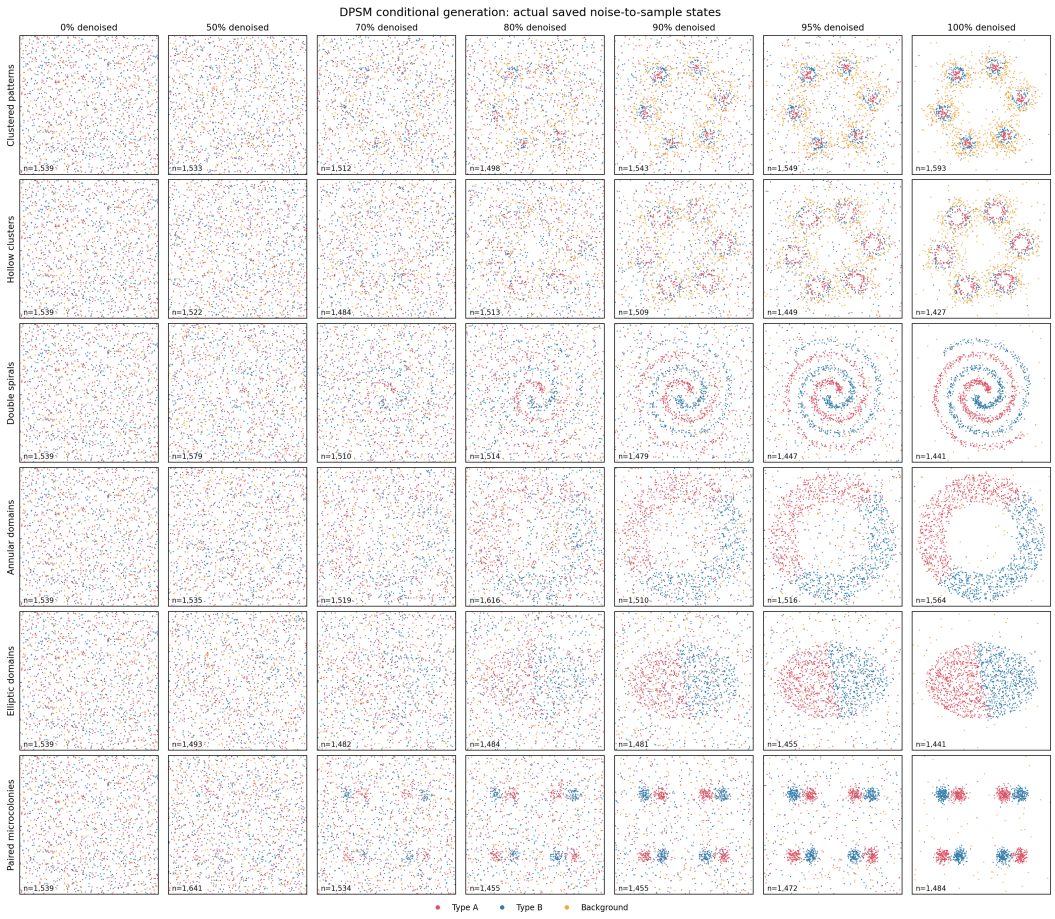

Figure 8: Actual saved sampler states for the first configuration of the first prespecified DPSM fit. Columns show the stages labeled 0%, 50%, 70%, 80%, 90%, 95%, and 100% denoised. From top to bottom, rows are clustered patterns, hollow clusters, double spirals, annular domains, elliptic domains, and paired microcolonies. Every row starts from the same homogeneous marked Poisson reference configuration; only the structure condition and its resulting learned spatial field differ. All six conditions are jointly learned.

H REAL BIOLOGICAL DATA EXPERIMENTS

The DPSM architecture is described in Appendix E.3. Baseline methods and their training budget convention are summarized in Appendix E.5.

H.1 DATA, CELL TYPES, AND PATIENT-HELD-OUT WINDOWS

Source populations. DCIS uses the centroid and cell type table associated with Risom et al. (2022), with 79 fields of view of 1024×1024 pixels, approximately $500 \times 500\,\mu$m. Conditions are initial DCIS followed by ipsilateral invasive recurrence (case), initial DCIS without recurrence during follow-up (ctrl), normal breast, and subsequent ipsilateral invasive cancer (Ipsinv). Author-provided `Patient_ID` values are recovered through the `PointNumber` mapping in `Table_S1_Patient_Feature_Table.csv`. Initial and subsequent specimens from the same patient remain in one split.

IMC uses the single-cell annotations and segmentation objects released by Karimi et al. (2023). We restrict the cohort to primary, IDH-wildtype glioblastoma and tumor-core regions. Each coordinate is the centroid of a complete segmented cell object, computed from its unique pixels. MATLAB one-based, column-major pixel indices are converted using the actual image height. The spatial calibration is 1 μm per pixel. Empty and `NONE` annotations are excluded.

CRC uses the cell coordinate and annotation table from Schürch et al. (2020), containing 140 regions from 35 patients. Conditions are Crohn's-like lymphoid reaction (CLR) and diffuse inflammatory infiltration (DII).

Patient splits and window extraction. For each dataset, patients are assigned to training, validation, and test splits before window extraction. These splits remain fixed across methods and training seeds. Training windows use a stride equal to half the window side length; validation and test windows are non-overlapping. Coordinates are normalized to the unit square using the window bounds.

Cell type aggregation and filtering. Table 13 gives the complete mapping to six, six, and eight coarse types for DCIS, IMC, and CRC. The windows have side 512, 250, 480 pixels. A window is retained when its remaining count is at least $\max\{64, \lceil Q_{0.1}^{\text{train}} \rceil\}$. The resulting thresholds are 104, 102, and 64 cells. Retained sizes are shown in Tables 12.

Table 12: Retained data after preprocessing. Counts of patients and windows are train/validation/test.

Dataset	Types	Side (μm)	Min. cells	Patients	Windows
DCIS	6	250.00	104	44/12/13	416/46/50
IMC	6	250.00	102	64/21/21	4,736/429/374
CRC	8	181.17	64	23/6/6	2,643/250/239

Table 13: Cell type aggregation. Original annotation strings are retained here; mixed CRC labels describe annotation uncertainty.

Coarse type	Original annotations
DCIS Dataset	
Epithelial/tumor	TUMOR
Myoepithelial	MYOEP
Fibroblast	FIBROBLAST
Endothelial	ENDO
Immune	APC, BCELL, CD4T, CD8T, DC, IMMUNEOTHER, MACS, MAST, MONO, MONODC, NEUT, TCELL
Other	OTHER
IMC Dataset	
Cancer	Cancer
Astrocytes	Astrocytes

Coarse type	Original annotations
Endothelial	Endothelial cell
Lymphoid	B cell, NK cell, Tc, Th, Treg, T other
Myeloid	DCs cell, Cl BMDM, Alt BMDM, Cl MG, Alt MG, Macrophages other, Cl Mo, Non-Cl Mo, Int Mo
Granulocytes	Neutrophils, Mast cell
CRC Dataset	
Tumor	tumor cells
Immune	B cells, CD11b+ monocytes, CD11b+CD68+ macrophages, CD11c+ DCs, CD163+ macrophages, CD3+ T cells, CD4+ T cells, CD4+ T cells CD45RO+, CD4+ T cells GATA3+, CD68+ macrophages, CD68+ macrophages GzmB+, CD68+CD163+ macrophages, CD8+ T cells, NK cells, Tregs, granulocytes, immune cells, plasma cells
Stromal	adipocytes, smooth muscle, stroma
Vascular	lymphatics, vasculature
Neural	nerves
Tumor/immune (mixed)	tumor cells / immune cells
Immune/vascular (mixed)	immune cells / vasculature
Other	undefined

H.2 Evaluation statistics and complete condition-level results

Every method generates 512 raw windows per condition and seed, for 43,008 windows in total. References are the same fixed test windows for all methods and seeds. Each statistic is first computed within a biological condition; dataset scores are the arithmetic mean over its conditions. Tables report the arithmetic mean and sample SD ($\mathrm{ddof} = 1$) across the three seed-level scores. Lower is better for all five errors. Tables 14 and 15 report every biological condition in DCIS, IMC, and CRC. Dataset-level rankings describe averages across conditions.

Table 14: Per-condition DCIS results: mean with sample SD over three seeds; lower is better.

Method	Scaled NN W_1	Neighbor mark RMSE	Spectral error	Enrichment RMSE	Moran RMSE
DCIS Dataset: DCIS-case					
DPSM	$\mathbf{0.04113_{\pm 0.00065}}$	$\mathbf{0.06308_{\pm 0.00364}}$	$\mathbf{0.18945_{\pm 0.00769}}$	$\mathbf{0.23336_{\pm 0.01800}}$	$\mathbf{0.07674_{\pm 0.00351}}$
PSD	$0.10712_{\pm 0.00141}$	$0.16269_{\pm 0.02927}$	$0.42570_{\pm 0.00866}$	$0.72889_{\pm 0.00458}$	$0.27981_{\pm 0.00159}$
SI-VAE	$0.11479_{\pm 0.00765}$	$0.19589_{\pm 0.00387}$	$0.32352_{\pm 0.00792}$	$0.59133_{\pm 0.00853}$	$0.24830_{\pm 0.00192}$
CytoSpatio	$0.10564_{\pm 0.00028}$	$0.16939_{\pm 0.00012}$	$0.43555_{\pm 0.00357}$	$0.73441_{\pm 0.00048}$	$0.27907_{\pm 0.00041}$
DCIS Dataset: DCIS-ctrl					
DPSM	$\mathbf{0.04272_{\pm 0.00111}}$	$\mathbf{0.07379_{\pm 0.01154}}$	$\mathbf{0.28720_{\pm 0.02188}}$	$\mathbf{0.25228_{\pm 0.01129}}$	$\mathbf{0.06332_{\pm 0.00783}}$
PSD	$0.12571_{\pm 0.00031}$	$0.17194_{\pm 0.01239}$	$0.54586_{\pm 0.00876}$	$0.65231_{\pm 0.00066}$	$0.24163_{\pm 0.00120}$
SI-VAE	$0.12777_{\pm 0.01177}$	$0.19978_{\pm 0.04098}$	$0.65804_{\pm 0.34440}$	$0.49742_{\pm 0.06082}$	$0.20092_{\pm 0.01078}$
CytoSpatio	$0.12519_{\pm 0.00037}$	$0.18523_{\pm 0.00004}$	$0.55203_{\pm 0.00256}$	$0.65644_{\pm 0.00176}$	$0.24065_{\pm 0.00050}$
DCIS Dataset: Normal					
DPSM	$0.14871_{\pm 0.00341}$	$\mathbf{0.06609_{\pm 0.00503}}$	$\mathbf{0.51039_{\pm 0.01883}}$	$\mathbf{0.46863_{\pm 0.00303}}$	$\mathbf{0.09072_{\pm 0.00561}}$
PSD	$0.11974_{\pm 0.00030}$	$0.13564_{\pm 0.01150}$	$0.67091_{\pm 0.00226}$	$0.86361_{\pm 0.00828}$	$0.21901_{\pm 0.00072}$
SI-VAE	$0.13228_{\pm 0.00679}$	$0.19029_{\pm 0.04211}$	$0.64309_{\pm 0.02215}$	$0.66030_{\pm 0.02402}$	$0.19099_{\pm 0.00260}$
CytoSpatio	$\mathbf{0.11970_{\pm 0.00029}}$	$0.12840_{\pm 0.00046}$	$0.67706_{\pm 0.00083}$	$0.87572_{\pm 0.00267}$	$0.21811_{\pm 0.00039}$
DCIS Dataset: Ipsinv					
DPSM	$\mathbf{0.06340_{\pm 0.00409}}$	$0.19174_{\pm 0.00555}$	$\mathbf{0.21374_{\pm 0.02888}}$	$\mathbf{0.38921_{\pm 0.02466}}$	$\mathbf{0.08979_{\pm 0.00899}}$
PSD	$0.07541_{\pm 0.00108}$	$0.21957_{\pm 0.04301}$	$0.28887_{\pm 0.03455}$	$0.68132_{\pm 0.00417}$	$0.17499_{\pm 0.00049}$
SI-VAE	$0.07645_{\pm 0.00748}$	$0.17886_{\pm 0.01295}$	$0.34622_{\pm 0.23389}$	$0.51377_{\pm 0.01340}$	$0.12212_{\pm 0.01427}$
CytoSpatio	$0.07470_{\pm 0.00007}$	$\mathbf{0.12009_{\pm 0.00036}}$	$0.31092_{\pm 0.00027}$	$0.68966_{\pm 0.00108}$	$0.17223_{\pm 0.00083}$

Table 15: Per-condition IMC and CRC results: mean with sample SD over three seeds; lower is better.

Method	Scaled NN W_1	Neighbor mark RMSE	Spectral error	Enrichment RMSE	Moran RMSE
IMC Dataset: Primary IDH-wildtype glioblastoma					
DPSM	$\mathbf{0.00928}_{\pm 0.00184}$	$\mathbf{0.07868}_{\pm 0.00411}$	$\mathbf{0.09842}_{\pm 0.00347}$	$0.14247_{\pm 0.01496}$	$\mathbf{0.03880}_{\pm 0.00131}$
PSD	$0.11415_{\pm 0.00048}$	$0.16385_{\pm 0.00306}$	$0.25457_{\pm 0.00159}$	$0.24863_{\pm 0.00697}$	$0.07370_{\pm 0.00068}$
SI-VAE	$0.06816_{\pm 0.01061}$	$0.13578_{\pm 0.00526}$	$0.16480_{\pm 0.01165}$	$\mathbf{0.13943}_{\pm 0.00973}$	$0.05609_{\pm 0.00343}$
CytoSpatio	$0.11424_{\pm 0.00136}$	$0.14523_{\pm 0.00010}$	$0.25969_{\pm 0.00259}$	$0.25657_{\pm 0.00136}$	$0.07322_{\pm 0.00089}$
CRC Dataset: CLR					
DPSM	$\mathbf{0.02332}_{\pm 0.00119}$	$\mathbf{0.08758}_{\pm 0.00305}$	$\mathbf{0.18524}_{\pm 0.01113}$	$\mathbf{0.07831}_{\pm 0.00395}$	$\mathbf{0.02630}_{\pm 0.00311}$
PSD	$0.06262_{\pm 0.00065}$	$0.10581_{\pm 0.00694}$	$0.37132_{\pm 0.00554}$	$0.15293_{\pm 0.00426}$	$0.08235_{\pm 0.00176}$
SI-VAE	$0.06178_{\pm 0.00302}$	$0.14426_{\pm 0.02815}$	$0.34981_{\pm 0.02080}$	$0.09330_{\pm 0.00762}$	$0.06663_{\pm 0.00281}$
CytoSpatio	$0.06135_{\pm 0.00046}$	$0.13307_{\pm 0.00027}$	$0.37785_{\pm 0.00209}$	$0.14660_{\pm 0.00350}$	$0.08107_{\pm 0.00017}$
CRC Dataset: DII					
DPSM	$\mathbf{0.02780}_{\pm 0.00528}$	$\mathbf{0.12351}_{\pm 0.00423}$	$\mathbf{0.24720}_{\pm 0.01573}$	$\mathbf{0.13390}_{\pm 0.01973}$	$\mathbf{0.06394}_{\pm 0.00909}$
PSD	$0.08958_{\pm 0.00041}$	$0.14772_{\pm 0.00787}$	$0.42656_{\pm 0.00315}$	$0.23128_{\pm 0.01023}$	$0.13583_{\pm 0.00088}$
SI-VAE	$0.06675_{\pm 0.00609}$	$0.15597_{\pm 0.00327}$	$0.37241_{\pm 0.00997}$	$0.17563_{\pm 0.00638}$	$0.12391_{\pm 0.00308}$
CytoSpatio	$0.08856_{\pm 0.00025}$	$0.14957_{\pm 0.00043}$	$0.42976_{\pm 0.00386}$	$0.22884_{\pm 0.00288}$	$0.13681_{\pm 0.00057}$

I Comparison with direct MCMC

We compare spatial accuracy on the main datasets and population allocation against Local MCMC and Collective MCMC in controlled studies.

I.1 Interacting point process datasets

On the three interacting point process datasets, DPSM and MCMC share the fitted energy. MCMC draws joint mark counts from the empirical training distribution and updates positions conditional on them. Each method generates 512 configurations using 200 sampling steps. Both methods have small errors on the reported metrics, with metric-dependent rankings (Table 16).

Table 16: Interacting point process datasets, same fitted energy. Metrics follow Appendix F.2. Lower is better. Bold marks the smaller error within each dataset.

Dataset	Method	Scaled NN W_1	Neighbor mark RMSE	Spectral error	Enrichment RMSE	Moran RMSE
Unmarked Strauss	MCMC	$\mathbf{2.98 \times 10^{-3}}$	—	9.09×10^{-3}	$\mathbf{8.97 \times 10^{-2}}$	3.29×10^{-3}
	DPSM	5.80×10^{-3}	—	$\mathbf{7.02 \times 10^{-3}}$	9.54×10^{-2}	$\mathbf{1.83 \times 10^{-3}}$
Three-type Strauss	MCMC	4.25×10^{-3}	$\mathbf{9.65 \times 10^{-3}}$	8.95×10^{-3}	6.70×10^{-2}	$\mathbf{1.90 \times 10^{-3}}$
	DPSM	$\mathbf{2.16 \times 10^{-3}}$	1.01×10^{-2}	$\mathbf{7.47 \times 10^{-3}}$	$\mathbf{4.77 \times 10^{-2}}$	3.77×10^{-3}
Four-type Gibbs	MCMC	$\mathbf{2.47 \times 10^{-3}}$	$\mathbf{6.33 \times 10^{-3}}$	1.61×10^{-2}	1.21×10^{-1}	$\mathbf{4.70 \times 10^{-3}}$
	DPSM	3.44×10^{-3}	1.20×10^{-2}	$\mathbf{1.44 \times 10^{-2}}$	$\mathbf{6.78 \times 10^{-2}}$	9.03×10^{-3}

I.2 Spatial structure and mark organization

We compare DPSM with Local MCMC and Collective MCMC on the six structural datasets, using 32 generated configurations per condition.

Table 17 shows comparable geometric recovery on the first three conditions, with higher alignment scores for MCMC. In contrast, DPSM better recovers coherent mark partitions and complementary colony pairs in the last three conditions. Figure 9 illustrates this distinction: the MCMC samples reproduce the spatial outlines but retain substantial mark mixing.

Table 17: Precision, coverage, and alignment are averaged over the clustered patterns, hollow clusters, and double spirals datasets. The three organization scores correspond individually to the annular domains, elliptic domains, and paired microcolonies datasets. Relative count W_1 is averaged over all six datasets.

Method	Precision ↑	Coverage ↑	Alignment ↑	Annular org. ↑	Elliptic org. ↑	Paired org. ↑	Rel. count W_1 ↓
DPSM	0.9390	0.9494	0.8916	0.9217	0.9154	0.8655	0.0314
Local MCMC	0.9457	0.9494	0.9243	0.2507	0.3484	0.3798	0.1362
Collective MCMC	0.9442	0.9512	0.9243	0.2390	0.3537	0.3666	0.1394

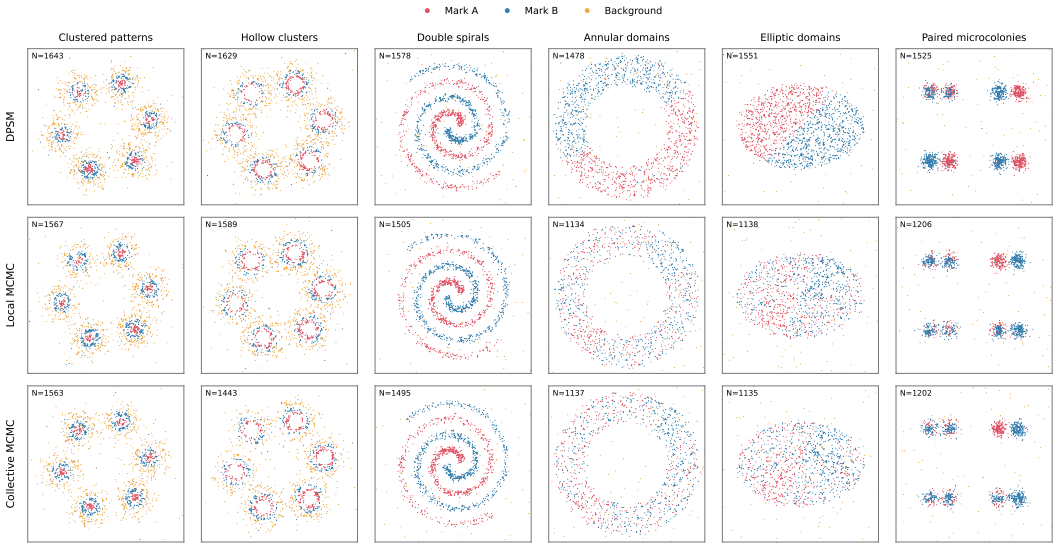

Figure 9: Structural samples from DPSM (top), Local MCMC (middle), and Collective MCMC (bottom). Columns show the six structural conditions.

I.3 Population allocation and configuration-level organization

Local MCMC uses pointwise updates; Collective MCMC also proposes transformations of the full configuration, accepted against the target distribution. Such moves help when they connect modes. DPSM exposes mixture weights along a noisy path with overlapping component laws, paralleling the Langevin motivation of Song & Ermon (2019).

Table 18 uses directly sampled references; their construction is given in Section I.4. The first six studies supply spatial bases or latent structure; curved lobes learns component maps and weights from unlabeled data.

Local MCMC misallocates Poisson mixture and collective-patch populations; Collective MCMC recovers their mode proportions reasonably well. Global flips and geometric symmetries leave substantial organization errors in the chain, annulus and tissue.

Figure 10 separates orientation from organization magnitude in the local chain. Collective MCMC reduces the positive-orientation overrepresentation of Local MCMC, but still underrepresents organized configurations. DPSM more closely matches both complete distributions.

Table 18: Organization studies comparing DPSM, Local MCMC and Collective MCMC. Method columns report absolute deviations $|\hat{s} - s_{\mathrm{ref}}|$, computed before rounding (lower is better). 512 outputs in the first three studies and 256 in the remainder.

Dataset and statistic	Target/ref.	DPSM	Local MCMC	Collective MCMC
Poisson mixture: mode A[a]	0.80	0.01641	0.28438	0.00664
Local chain: coherent fraction[b]	0.50	0.00005	0.23637	0.20708
Collective patches: positive phase[b]	0.80	0.01250	0.29219	0.02031
Annulus: two-boundary fraction[c]	0.83	0.02655	0.61407	0.61407
Tissue: disagreeing edges[c]	4.68	0.06719	15.05781	14.94063
Spots/bands: profile TV[c]	0.20	0.01172	0.42969	0.41016
Learned curved lobes: mode A[a]	0.50	0.00391	0.19922	0.02734
Learned curved lobes: mode A[a]	0.80	0.03828	0.16484	0.02266

Collective kernels combine local updates with global MH moves: [a]transpose; [b]mark/sign exchange; [c]rotations/reflections and, where applicable, mark exchange using the domain geometry.

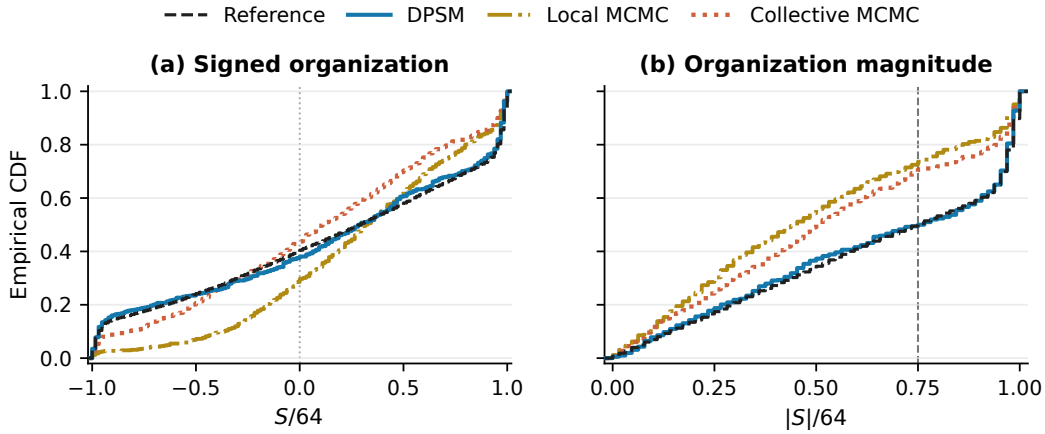

Figure 10: **Empirical organization distributions for the local chain.** All methods use 128 steps (512 configurations); reference: 8,192 independent draws. Here $S = \sum_j (n_{j+} - n_{j-})$, where $n_{j\pm}$ count points with the two local orientations defined in Section I.4. The panels show signed organization $S/64$ and magnitude $|S|/64$; the dashed vertical line marks the coherence threshold 0.75.

These symmetry-based moves connect some of the global modes above. To examine allocation when they cannot connect shape classes, we consider a two-mark Poisson mixture of rings, disconnected rectangles, and a T-shape paired with four islands, with weights $(0.6, 0.3, 0.1)$ and identical mark count distributions. Each class includes its full rotation/reflection/mark-exchange orbit; these actions cannot connect classes. The 28 template components are supplied; mixture parameters and a constrained noisy time function are learned from unlabeled data.

DPSM and Collective MCMC share the learned target model f and Poisson starts, producing 512 configurations each in 200 steps. All $102{,}400$ collective proposals are accepted, yet none changes the class posterior: $p_k(gX) = p_k(X)$. After the first step, no further class changes appear in the recorded states. DPSM recovers the class proportions (Table 19). Independent draws and DPSM using the exact noise path from the same fitted f recover reasonable allocations; the learned path is already close to that analytic path.

Table 19: Three-class mixture dataset: absolute class proportion errors $|\hat{p}_k - p_k^\star|$ and Class TV (lower is better). Class errors and Class TV use weights $(0.6, 0.3, 0.1)$.

Method	Class A	Class B	Class C	Class TV
DPSM	5.47×10^{-3}	3.13×10^{-3}	2.34×10^{-3}	5.47×10^{-3}
Collective MCMC	3.23×10^{-1}	3.24×10^{-2}	3.55×10^{-1}	3.55×10^{-1}

I.4 CONSTRUCTION OF THE CONTROLLED DATASETS

The first five Table 18 settings use marks $m \in \{0, 1\}$, whereas spots/bands is unmarked. For the Poisson constructions below, μ_{jm} denotes the integrated intensity in spatial cell C_j: draw independent $N_{jm} \sim \text{Poisson}(\mu_{jm})$ conditional on the latent state, then place points independently and uniformly in C_j.

Poisson mixture. On $[0, 1]^2$, choose $Z \in \{A, B\}$ with probabilities $(0.8, 0.2)$ and set

$$\lambda_m(x \mid Z) = 64\{1 + 0.95(-1)^m g_Z(x)\},$$
$$g_A(x) = 1 - 2\mathbf{1}\{x_1 \geq 1/2\}, \qquad g_B(x) = 1 - 2\mathbf{1}\{x_2 \geq 1/2\}.$$

The four quadrants provide exact integration cells. Conditional on Z, both marks have total intensity 64; their preferred halves differ between the two layouts. Reference configurations are independent draws using categorical sampling followed by Poisson sampling, with continuous within-quadrant locations.

Local chain. Arrange 64 cells of an 8×8 unit-square grid in a serpentine open chain. Each cell has state $z_j \in \{0, +1, -1\}$, indicating vacancy or one point, with discrete law

$$p(z) \propto \exp\left\{ 3\sum_{j=1}^{63} z_j z_{j+1} - 1.5\sum_{j=1}^{64} |z_j| + 0.005\sum_{j=1}^{64} z_j \right\}.$$

There is at most one point per cell, located uniformly within it. A positive state carries mark 0 in the left half and mark 1 in the right half; negative states reverse these assignments. Exact backward messages followed by forward categorical sampling generate the chain directly. The table's coherence statistic is $\Pr(|\sum_j z_j| > 48)$.

Collective patches. Use the same 8×8 hard-cell support and physical mark conversion, but couple all cell states through $S = \sum_j z_j$ and $N = \sum_j |z_j|$:

$$p(z) \propto \exp\{\eta N + JS^2/(2K) + hS\}, \qquad K = 64, \quad \eta = 0, \quad J = 4, \quad h \simeq 0.011.$$

The field is calibrated by finite summation to $\Pr(S > 0) = 0.8$. Direct simulation draws (n_+, n_-) with weight $\frac{64!}{n_+!n_-!(64-n_+-n_-)!}\exp\{J(n_+ - n_-)^2/128 + h(n_+ - n_-)\}$, assigns these states uniformly to distinct cells, and draws uniform locations.

Annulus. Divide $\{x : 0.26 \leq \|x - (1/2, 1/2)\| \leq 0.46\}$ into twelve equal angular sectors. Draw $z \in \{-1, +1\}^6$ from

$$p(z) \propto \exp\left\{ 0.7\left(\sum_{j=1}^{5} z_j z_{j+1} - z_6 z_1 \right) \right\}, \qquad s = (z, -z), \qquad \mu_{jm} = 32\{1 + 0.97(-1)^m s_j\}.$$

Enumerating the 64 latent states permits exact categorical sampling, followed by independent Poisson sector counts. Locations are uniform in area: angle is uniform within the sector and squared radius is uniform on $[0.26^2, 0.46^2]$. Expected total count is 768. This two-mark latent-sector Cox model differs from the three-mark annular spatial structure benchmark.

Tissue compartments. For the disk D of radius 0.49 centered at $(1/2, 1/2)$, define cells $C_{rc} = D \cap ([(r-1)/8, r/8) \times [(c-1)/4, c/4))$, with $r = 1, \ldots, 8$ along x and $c = 1, \ldots, 4$ along y. Cell masses M_{rc} sum to 640 and are proportional to disk intersection areas approximated by midpoint quadrature with 1024 locations along x per strip. Latent signs $z_{rc} \in \{-1, +1\}$ have law

$$p(z) \propto \exp\left\{ 0.95\sum_{\langle u,v \rangle} z_u z_v + 1.2\sum_{c=1}^{4}(z_{1c} - z_{8c}) \right\}, \qquad \mu_{rcm} = \frac{M_{rc}}{2}\{1 + 0.94(-1)^m z_{rc}\},$$

where the sum counts each horizontal or vertical nearest-neighbor grid edge once. Exact transfer messages over the sixteen possible four-sign row states give direct latent draws, followed by Poisson counts and uniform disk intersection locations. The boundary field favors opposite marks at the two ends.

Spots/bands. This unmarked Cox model uses a fixed bank of 64 profiles on a periodic 20×20 unit-square grid. Independent real standard Gaussian white-noise fields are Fourier-filtered with multiplier proportional to $[0.12 + (\|k\|^2 - 2.5^2)^2/2.5^4]^{-1/2}$, setting the zero mode to zero and normalizing the multiplier to unit mean square. For each resulting field H, define

$$b_j = 0.01 + \mathrm{sigmoid}\{5(H_j - 0.3)\}, \qquad g_j = b_j/\bar{b}, \qquad \mu_j = 2\{1 + 0.97(g_j - 1)\}.$$

A configuration chooses one frozen profile uniformly, then samples its Poisson counts and uniform cell locations, giving expected total count 800.

Curved lobes with two mixing weights. Configurations lie on $[0,1]^2$ with marks $m \in \{0,1\}$ and a 12×12 grid. For zero-based cell indices (i,j), set $x_j = (j+1/2)/12-1/2$, $y_i = (i+1/2)/12-1/2$, and $b(y) = 0.135 + 0.08(y/0.25)^2$. The first orientation has integrated cell means

$$\ell_{mij} = -\frac{1}{2}\left(\frac{x_j + (-1)^m b(y_i)}{0.065}\right)^2 - \left(\frac{y_i}{0.27}\right)^4, \qquad \mu^{(0)}_{mij} = 64\left(0.994\frac{e^{\ell_{mij}}}{\sum_{u,v} e^{\ell_{muv}}} + \frac{0.006}{144}\right).$$

The second orientation transposes the spatial indices, $\mu^{(1)}_{mij} = \mu^{(0)}_{mji}$. A shared orientation is selected with probabilities $(\pi, 1-\pi)$, using $\pi = 0.5$ and 0.8 in the two experiments. Each mark therefore has an independent $\mathrm{Poisson}(64)$ total in either orientation. Each experiment contains 4096 training configurations and an independent bank of 2048 reference configurations.

Three-class mixture: explicit shape templates. This family uses the same square, grid, and marks. Write $R(r,c;h,w) = \{r,\dots,r+h-1\} \times \{c,\dots,c+w-1\}$ for a set of row and column indices, and ∂R for its perimeter cells. The three canonical pairs of foreground masks are

$$\begin{aligned} M_{00} &= \partial R(1,1;5,5), & M_{01} &= \partial R(6,6;5,5),\\ M_{10} &= R(1,1;2,4) \cup R(8,1;2,4), & M_{11} &= R(1,7;2,4) \cup R(8,7;2,4),\\ M_{20} &= R(2,2;1,9) \cup R(3,6;7,1), & M_{21} &= \bigcup_{r\in\{3,8\},\, c\in\{2,9\}} R(r,c;2,2). \end{aligned}$$

These are paired rings, separated rectangles, and a T paired with four islands; each mask contains 16 cells. Draw the shape class with probabilities $(0.6, 0.3, 0.1)$, then choose uniformly among its distinct transforms under D_4 rotations/reflections and whole-configuration mark exchange. The respective orbit sizes are 4, 8, and 16. For the resulting mask $M_{c,g,m}$, set

$$\mu_{c,g,mij} = 96\left(\frac{0.97}{16}\mathbf{1}\{(i,j) \in M_{c,g,m}\} + \frac{0.03}{144}\right).$$

Direct sampling and splits. The curved-lobe and three-class mixture families use the same conditional Poisson and uniform-within-cell sampling rule as above, with continuous intensity 144μ. The three-class mixture archives contain 4096 training, 512 validation, 2048 test, and 4096 independent reference configurations; hidden class labels are reserved for evaluation. The final comparison uses the fixed 2048-configuration test bank.