Submitted to the Annals of Applied Statistics

STATISTICAL CURVE MODELS FOR INFERRING 3D CHROMATIN ARCHITECTURE (APPENDIX)

BY ELENA TUZHILINA[1,a], TREVOR HASTIE[2,b] AND MARK SEGAL[3,c]

[1]*Department of Statistical Sciences, University of Toronto, Toronto ON M5S3G3 Canada,* [a]*elena.tuzhilina@utoronto.ca*

[2]*Department of Statistics, Stanford University, Stanford CA 94305 USA,* [b]*hastie@stanford.edu*

[3]*Department of Epidemiology and Biostatistics, University of California, San Francisco CA 94143 USA,* [c]*Mark.Segal@ucsf.edu*

APPENDIX A: PROPERTY OF THE DEMMLER-REINSCH PENALTY MATRIX

LEMMA 1. *If* $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^n$ *is* n*-dimensional vector of ones and* K *is the smoothing spline penalty matrix, then*

- $K\mathbf{1} = 0$*;*
- $\mathbf{1}^\top K = 0$*;*
- $\sum_{1 \le i,j, \le n} K_{ij} = \mathbf{1}^\top K \mathbf{1} = 0.$

PROOF. Denote $K = U\Sigma U^\top$ the eigendecomposition of the penalty matrix. One can show that K has two zero eigenvalues and the corresponding eigenvectors span the subspace of linear functions (see, for example, Green and Silverman, 1994). For simplicity, we assume that the eigenvalues are sorted in increasing order in Σ, so $\sigma_1 = \sigma_2 = 0$ and matrix U has the following structure: $U = (U_0, U_0^\perp)$, where $U_0 \in \mathbb{R}^{n\times 2}$ corresponds to the null-space of K and $U_0^\perp \in \mathbb{R}^{n\times(n-2)}$ is the orthogonal space. Therefore, if $g(t)$ is some linear function of t and $G = (g(t_1), \dots, g(t_n))^\top$, we get $G \in span(U_0)$ as well as $(U_0^\perp)^\top G = 0$. This leads us to the relation

$$KG = U\Sigma \begin{pmatrix} U_0^\top G \\ (U_0^\perp)^\top G \end{pmatrix} = U \begin{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} U_0^\top G \\ 0 \end{pmatrix} = 0.$$

In particular, setting $g(t) = 1$ implies $KG = K\mathbf{1} = 0$. The second relation $\mathbf{1}^\top K = 0$ automatically follows form the fact that K is positive semi-definite (PSD), which immediately implies $\sum_{1 \le i,j, \le n} K_{ij} = \mathbf{1}^\top K \mathbf{1} = 0$. $\square$

APPENDIX B: DEGREES-OF-FREEDOM

LEMMA 2. *The projection step of SPoisMS solves the following penalized PCMS problem for some* $\tilde{S}$ *independent of* λ*:*

$$\underset{X \in \mathbb{R}^{n\times 3}}{\text{minimize}} \frac{1}{n^2} \|\tilde{S} - XX^\top\|_F^2 + \lambda \operatorname{tr}(X^\top K X).$$

PROOF. Recall, that *SPoisMS* is equivalent to *PoisMS* with C replaced by

$$C^\lambda = C - \frac{\lambda n^2}{2} K.$$

Suppose that the current reconstruction guess at the beginning of a new *WPCMS* loop is X_0. Thus, the working response in the *WPCMS* problem is $Z^\lambda = D^2(X_0) - \frac{C^\lambda}{W} + 1$.

Next, each gradient step within the *WPCMS* loop updates the similarity matrix as

$$S = XX^\top - \Phi(W * (Z^\lambda - D^2(X))) =$$

$$= \underbrace{XX^\top - \Phi(W * (Z - D^2(X)))}_{\tilde{S}} + \frac{\lambda n^2}{2} K.$$

Here we use the linearity of the operator $\Phi(\cdot)$ and Lemma 1 implying

$$\Phi(K) = K - \operatorname{diag}(K \cdot 1) = K.$$

As a result, the projection step will minimize the *PCMS* loss as

$$\|S - XX^\top\|_F^2 = \|\tilde{S} + \frac{\lambda n^2}{2} K - XX^\top\|_F^2 =$$

$$= \|\tilde{S} - XX^\top\|_F^2 - \lambda n^2 \operatorname{tr}(X^\top K X) + \frac{\lambda^2 n^4}{4} \|K\|_F^2.$$

Removing the terms independent of X and dividing by n^2 this leads us to the optimization problem from the lemma. $\square$

LEMMA 3. *The degrees-of-freedom for one-dimensional penalized PCMS*

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{n^2}\|\tilde{S}-xx^\top\|_F^2+\lambda x^\top Kx$$

can be calculated as

$$df=\sum_{i=1}^{n}\frac{1}{1+\frac{\lambda n^2}{\|x_0\|^2}\sigma_i},$$

where $x_0\in\mathbb{R}^n$ is the solution to the problem.

PROOF. Again, note that the one-dimensional problem can be solved in two ways. One can rewrite it as

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \big\|\big(\tilde{S}-\frac{\lambda n^2}{2}K\big)-xx^\top\big\|_F^2$$

and find x via the eigendecomposition of $\tilde{S}-\frac{\lambda n^2}{2}K$. Alternatively, one can apply alternating algorithm. If the current solution guess is x_0 then updated x can be found via solving a spline problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{n^2}\|\tilde{S}-xx_0^\top\|_F^2+\lambda x^\top Kx \Longleftrightarrow$$

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{n^2}\|y-x\|_F^2+\frac{\lambda}{\|x_0\|^2}x^\top Kx. \tag{1}$$

Here we set $y=\tilde{S}\cdot\frac{x_0}{\|x_0\|}$ and reparameterize the problem as $x:=\|x_0\|\cdot x$. The second approach implies that at the convergence point we can replace the original *PCMS* problem by (1). Thus, we can borrow the definition of degrees-of-freedom from regular smoothing splines to compute *df* for one-dimensional penalized *PCMS*. The last step is to plug-in the corresponding values to formula (14).

□

APPENDIX C: *POISMS* AND *SPOISMS* RECONSTRUCTION COMPARISON

To understand how well formula (17) aligns our *SPoisMS* and *PoisMS* reconstructions from Section 8 we introduce the notion of *discrete curvature* as follows. For a 3D conformation $X=\begin{pmatrix}-x_1^\top-\\ \vdots\\ -x_n^\top-\end{pmatrix}$ we calculate directed edges $\vec{e}_i=\vec{x}_{i+1}-\vec{x}_i$ and compute the *average angle deficit* as

$$\kappa(X)=\frac{1}{n-2}\sum_{i=1}^{n-2}\angle(\vec{e}_i;\vec{e}_{i+1}).$$

This quantity measures how wiggly the reconstruction X is, and can serve as an approximation to the formal curvature for the corresponding smooth one-dimensional curve γ.

Now we take the *SPoisMS* reconstructions calculated for the grid of penalty factors

$$\lambda=10^4,10^3,100,10,1,0.1$$

and the accordant *PoisMS* reconstructions with

$$df=5,9,15,27,47,84.$$

We plot the average angle deficit for each reconstruction in Figure 1. As expected, the reconstruction wiggliness increases with the growth of the model complexity (increase in *df* or decrease in λ). Moreover, the plot suggests that the corresponding *SPoisMS* and *PoisMS* reconstructions have similar curvature.

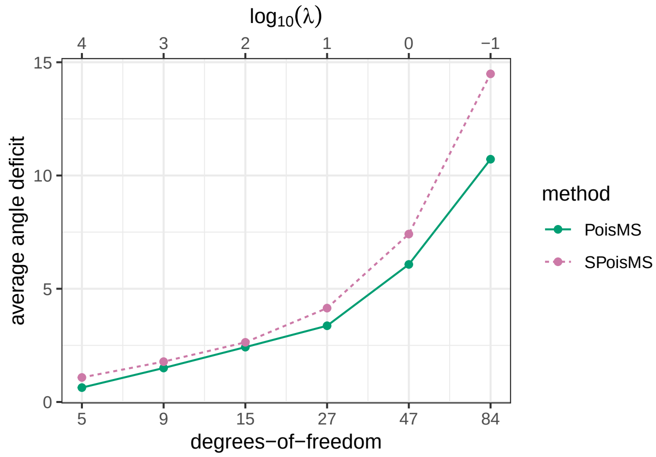


FIG 1. *Dependence between the average angle deficit (measuring how wiggly is a reconstruction) and the model complexity (controlled by df in PoisMS and λ in SPoisMS). The plot demonstrates that the corresponding SPoisMS and PoisMS reconstructions have similar curvature, which implies the accuracy of the proposed formula for estimating degrees-of-freedom.*

APPENDIX D: CONSISTENCY OF THE *SPOISMS* RECONSTRUCTIONS ACROSS DIFFERENT RESOLUTIONS

Below we provide plots validating the consistency of the *SPoisMS* reconstruction across 100kb, 50kb, 25kb, and 10kb resolutions. Figure 2 displays the aligned 3D conformations produced by *SPoisMS* that exhibit the degree of smoothness $df \approx 25$ and $df \approx 50$. Figure 3 presents pairwise plots for expected counts estimated using each of these reconstructions.

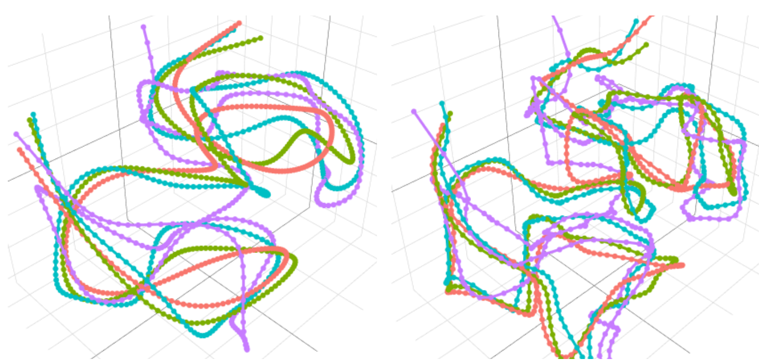


FIG 2. *The 3D reconstructions corresponding to approximately 25 (left) and 50 (right) degrees-of-freedom. For better visualization, the reconstructions were aligned using the Procrustes method before plotting. The plot reveals a high degree of consistency between the SPoisMS reconstructions across different resolutions.*

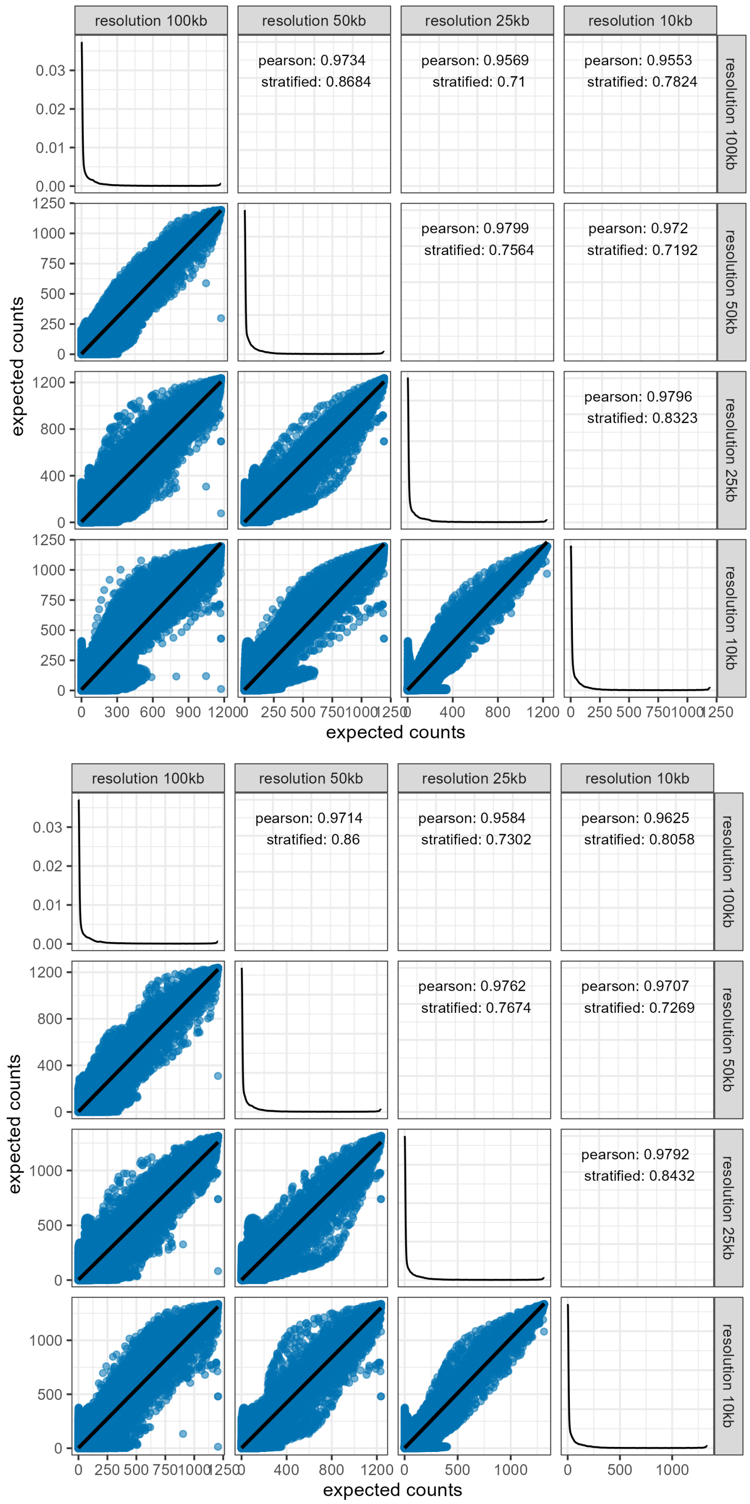


FIG 3. *Pairwise plots for expected counts computed for the 3D reconstructions corresponding to approximately 25 (top) and 50 (bottom) degrees-of-freedom. Before plotting, the expected counts for the 50kb, 25kb, and 10kb resolutions were summed to align with the 100kb resolution. The correlation values are close to one, thus validating the significant similarity of the reconstructions across various resolutions.*

APPENDIX E: *SPOISMS* PERFORMANCE EVALUATION

LEMMA 4. *The optimization problem*

$$\underset{\gamma\in\Gamma,\ \beta\in\mathbb{R}}{\text{minimize}}\, \ell^{train}_{SPoisMS}(\gamma,\beta;C,\lambda).$$

is equivalent to solving the reduced PoisMS problem

$$\underset{X_{\mathcal{O}}\in\mathbb{R}^{n\times 3},\ \beta\in\mathbb{R}}{\text{minimize}}\, \ell_{PoisMS}\big(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}}-\frac{\lambda|\mathcal{O}|^2}{2}(K/K_{\mathcal{U}\mathcal{U}})\big). \tag{2}$$

Here $K/K_{\mathcal{U}\mathcal{U}} = K_{\mathcal{O}\mathcal{O}} - K_{\mathcal{O}\mathcal{U}}K^{-1}_{\mathcal{U}\mathcal{U}}K_{\mathcal{U}\mathcal{O}}$ *is the Schur complement of the penalty matrix, which has block structure similar to* C, *i.e.* $K = \left(\begin{smallmatrix} K_{\mathcal{O}\mathcal{O}} & K_{\mathcal{O}\mathcal{U}} \\ K_{\mathcal{U}\mathcal{O}} & K_{\mathcal{U}\mathcal{U}}\end{smallmatrix}\right)$.

PROOF. First we note that Denote the solution by $X = \left(\begin{smallmatrix} X_{\mathcal{O}} \\ X_{\mathcal{U}}\end{smallmatrix}\right)$. We begin with rewriting the loss function in matrix form as

$$\ell^{train}_{SPoisMS}(X,\beta;C,K,\lambda) = \ell_{PoisMS}(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}}) + \lambda\,\mathrm{tr}(X^\top K X).$$

Next, we notice that the first part of the loss does not depend on $X_{\mathcal{U}}$. If $X_{\mathcal{O}}$ is fixed then $X_{\mathcal{U}}$ can be found as a minimum of

$$\mathrm{tr}(X^\top K X) = \mathrm{tr}(X_{\mathcal{O}}^\top K_{\mathcal{O}\mathcal{O}}X_{\mathcal{O}} + 2X_{\mathcal{U}}^\top K_{\mathcal{U}\mathcal{O}}X_{\mathcal{O}} + X_{\mathcal{U}}^\top K_{\mathcal{U}\mathcal{U}}X_{\mathcal{U}}).$$

Taking the derivative w.r.t. $X_{\mathcal{U}}$ and setting it to zero leads us to the stationary point $X_{\mathcal{U}} = -K^{-1}_{\mathcal{U}\mathcal{U}}K_{\mathcal{U}\mathcal{O}}X_{\mathcal{O}}$. Plugging it back to the original loss function implies

$$\begin{aligned}
&\ell^{train}_{SPoisMS}(X,\beta;C,K,\lambda) = \\
&= \ell_{PoisMS}(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}}) + \lambda\,\mathrm{tr}(X_{\mathcal{O}}^\top K_{\mathcal{O}\mathcal{O}}X_{\mathcal{O}} + 2X_{\mathcal{U}}^\top K_{\mathcal{U}\mathcal{O}}X_{\mathcal{O}} + X_{\mathcal{U}}^\top K_{\mathcal{U}\mathcal{U}}X_{\mathcal{U}}) = \\
&= \ell_{PoisMS}(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}}) + \lambda\,\mathrm{tr}\left(X_{\mathcal{O}}^\top(K_{\mathcal{O}\mathcal{O}} - K_{\mathcal{O}\mathcal{U}}K^{-1}_{\mathcal{U}\mathcal{U}}K_{\mathcal{U}\mathcal{O}})X_{\mathcal{O}}\right) = \\
&= \ell_{SPoisMS}(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}},K/K_{\mathcal{U}\mathcal{U}},\lambda).
\end{aligned}$$

Therefore, the optimal $X_{\mathcal{O}}$ can be found by solving the *SPoisMS* problem for the contact matrix $C_{\mathcal{O}\mathcal{O}}$ and the penalty matrix $K/K_{\mathcal{U}\mathcal{U}}$, i.e.

$$\underset{X_{\mathcal{O}}\in\mathbb{R}^{|\mathcal{O}|\times 3},\ \beta\in\mathbb{R}}{\text{minimize}}\, \ell_{SPoisMS}(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}},K/K_{\mathcal{U}\mathcal{U}},\lambda). \tag{3}$$

Note that in the main paper we demonstrated that the *SPoisMS* problem with the Demmler-Reinsch penalty matrix can be efficiently solved via the *PoisMS* algorithm. We extend this result to the optimization (3) as well. From the penalty matrix property $K\mathbf{1} = 0$ one can derive the system on equations involving the blocks of K, i.e. $\begin{cases} K_{\mathcal{O}\mathcal{O}}\mathbf{1} = -K_{\mathcal{O}\mathcal{U}}\mathbf{1} \\ K_{\mathcal{U}\mathcal{O}}\mathbf{1} = -K_{\mathcal{U}\mathcal{U}}\mathbf{1}\end{cases}$, leading us to analogous property of the Schur complement

$$(K/K_{\mathcal{U}\mathcal{U}})\mathbf{1} = K_{\mathcal{O}\mathcal{O}}\mathbf{1} - K_{\mathcal{O}\mathcal{U}}K^{-1}_{\mathcal{U}\mathcal{U}}K_{\mathcal{U}\mathcal{O}}\mathbf{1} = K_{\mathcal{O}\mathcal{O}}\mathbf{1} + K_{\mathcal{O}\mathcal{U}}\mathbf{1} = 0.$$

Following the proof in the main paper, one can show that the smoothing penalty in the loss function $\ell_{SPoisMS}(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}},K/K_{\mathcal{U}\mathcal{U}},\lambda)$ can be absorbed into the *PoisMS* part thereby proving the equivalence between (3) and the *PoisMS* problem

$$\underset{X_{\mathcal{O}}\in\mathbb{R}^{|\mathcal{O}|\times 3},\ \beta\in\mathbb{R}}{\text{minimize}}\, \ell_{PoisMS}\big(X_{\mathcal{O}},\beta;C_{\mathcal{O}\mathcal{O}}-\frac{\lambda}{2}(K/K_{\mathcal{U}\mathcal{U}})\big). \tag{4}$$

$\square$

APPENDIX F: DISTRIBUTION BASED METRIC SCALING

Denote Γ the matrix with element $\Gamma_{ij} = \mathbb{1}_{C_{ij}=0}$ and note that the derivatives of the log-link $\Lambda = e^{-D^2(X)+\beta}$ w.r.t. $D^2(X)$ and β are

$$\nabla_{D^2}\Lambda = -\Lambda \text{ and } \nabla_\beta \Lambda = \Lambda.$$

We will use these relations to derive the update steps for the optimization algorithm.

F.1. Hurdle Poisson.

Recall that the loss function has the following form

$$\ell_{\textit{HPoisMS}}(X, \beta, \pi; C) = \frac{1}{n^2}\Big[- \sum_{(i,j)\in\mathcal{N}} \log(\pi) - \sum_{(i,j)\notin\mathcal{N}} \log(1-\pi) + \\ + \sum_{(i,j)\notin\mathcal{N}} \Big[\lambda_{ij} - C_{ij}\log\lambda_{ij} + \log\Big(1 - e^{-\lambda_{ij}}\Big)\Big]\Big].$$

We first compute the derivatives of the loss w.r.t. D^2. Note that

$$\nabla_{D^2} e^{-\Lambda} = e^{-\Lambda} * \Lambda \text{ and } \nabla_\beta e^{-\Lambda} = -e^{-\Lambda} * \Lambda.$$

Denote $M = \frac{\Lambda}{1-e^{-\Lambda}}$, then (up to $\frac{1}{n^2}$ scale)

$$\nabla_{D^2} = (1-\Gamma) * \left(-\Lambda + C - \frac{e^{-\Lambda} * \Lambda}{1-e^{-\Lambda}}\right) = (1-\Gamma) * (C - M)$$

$$\nabla^2_{D^2} = (1-\Gamma) * \frac{\Lambda * (1-e^{-\Lambda}) - e^{-\Lambda} * \Lambda^2}{(1-e^{-\Lambda})^2} = (1-\Gamma) * M * (1 - e^{-\Lambda} * M)$$

Thus we can calculate the weights and the working response for the second-order approximation and apply *WPCMS* to update the reconstruction X.

Now we deal with the nuisance parameters. First, by analogy with the distance matrix one can compute the first and second order derivatives w.r.t. β (up to $\frac{1}{n^2}$ scale)

$$\nabla_\beta = \sum_{(i,j)\notin\mathcal{N}} (M_{ij} - C_{ij}) \text{ and } \nabla^2_\beta = \sum_{(i,j)\notin\mathcal{N}} M_{ij}(1 - e^{-\lambda_{ij}} M_{ij}).$$

We will use these derivatives to find optimal β when applying Newton's method. Next, we note that the optimal value for the Bernoulli parameter π can be found from the equation

$$\nabla_\pi = \frac{|\mathcal{N}|}{\pi} - \frac{n^2 - |\mathcal{N}|}{1-\pi} = 0.$$

This leads us to the explicit formula $\pi = \frac{|\mathcal{N}|}{n^2}$. Combining all the above steps leads us to the following *HPoisMS algorithm*.

Hurdle Poisson metric scaling (*HPoisMS*)

Set $\pi = \frac{|\mathcal{N}|}{n^2}$, then repeat until convergence:

1. For the current guess of X and β compute SOA:
   - evaluate $\Lambda = e^{-D^2(X)+\beta}$ and $M = \frac{\Lambda}{1-e^{-\Lambda}}$
   - calculate $\nabla_{D^2} = (1-\Gamma) * (C - M)$ and $\nabla^2_{D^2} = (1-\Gamma) * M * (1 - e^{-\Lambda} * M)$
   - compute $W = \nabla^2$ and $Z = D^2(X) - \frac{\nabla}{\nabla^2}$
2. Solve *WPCMS* problem with W and Z thereby updating X.

3. For fixed X run Newton's method to update β. Repeat until convergence:
   - evaluate $\Lambda = e^{-D^2(X)+\beta}$ and $M = \frac{\Lambda}{1-e^{-\Lambda}}$
   - calculate $\nabla_\beta = \sum_{(i,j)\notin\mathcal{N}}(M_{ij} - C_{ij})$ and $\nabla^2_\beta = \sum_{(i,j)\notin\mathcal{N}} M_{ij}(1 - e^{-\lambda_{ij}} M_{ij})$
   - update $\beta := \beta - \frac{\nabla}{\nabla^2}$.

F.2. Zero-inflated Poisson. In this section be build an optimization algorithm for the zero-inflated model. Let $a = \frac{\pi}{1-\pi}$ and rewrite the *ZIPoisMS* loss function as

$$\ell_{\mathit{ZIPoisMS}}(X, \beta, \pi; C) =$$

$$= \frac{1}{n^2}\Big[-\sum_{(i,j)\in\mathcal{N}} \log\Big(\pi + (1-\pi)e^{-\lambda_{ij}}\Big) - \sum_{(i,j)\notin\mathcal{N}} \log(1-\pi) + \sum_{(i,j)\notin\mathcal{N}} [\lambda_{ij} - C_{ij}\log\lambda_{ij}]\Big] =$$

$$= \frac{1}{n^2}\Big[-\sum_{(i,j)\in\mathcal{N}} \log\Big(a + e^{-\lambda_{ij}}\Big) - \sum_{1\le i,j\le n} \log(1-\pi) + \sum_{(i,j)\notin\mathcal{N}} [\lambda_{ij} - C_{ij}\log\lambda_{ij}]\Big]$$

We first calculate the derivatives of the loss w.r.t. D^2 (up to $\frac{1}{n^2}$ scale):

$$\nabla_{D^2} = -\Gamma * \frac{\Lambda * e^{-\Lambda}}{a + e^{-\Lambda}} + (1-\Gamma) * (-\Lambda + C) =$$

$$= -\Gamma * \frac{\Lambda}{ae^{\Lambda} + 1} + (1-\Gamma) * (C - \Lambda)$$

$$\nabla^2_{D^2} = -\Gamma * \frac{-\Lambda * (ae^{\Lambda} + 1) + ae^{\Lambda} * \Lambda^2}{(ae^{\Lambda}+1)^2} + (1-\Gamma) * \Lambda =$$

$$= \Gamma * \Lambda * \frac{ae^{\Lambda}(1-\Lambda) + 1}{(ae^{\Lambda}+1)^2} + (1-\Gamma) * \Lambda$$

Next, by analogy, we compute the derivatives w.r.t. the intercept β (up to $\frac{1}{n^2}$ scale):

$$\nabla_\beta = \sum_{(i,j)\in\mathcal{N}} \frac{\lambda_{ij}}{ae^{\lambda_{ij}} + 1} - \sum_{(i,j)\notin\mathcal{N}} (C_{ij} - \lambda_{ij})$$

$$\nabla^2_\beta = \sum_{(i,j)\in\mathcal{N}} \lambda_{ij} \frac{ae^{\lambda_{ij}}(1-\lambda_{ij}) + 1}{(ae^{\lambda_{ij}}+1)^2} + \sum_{(i,j)\notin\mathcal{N}} \lambda_{ij}$$

We will use these formulas to update the intercept via Newton's method. Now, we handle the nuisance parameter π. Unlike the Hurdle model, there is no explicit formula for optimal π; therefore, we also apply Newton's method to update it at each iteration. This requires us to compute the derivatives w.r.t. π (up to $\frac{1}{n^2}$ scale)

$$\nabla_\pi = -\sum_{(i,j)\in\mathcal{N}} \frac{1 - e^{-\lambda_{ij}}}{\pi + (1-\pi)e^{-\lambda_{ij}}} + \sum_{(i,j)\notin\mathcal{N}} \frac{1}{1-\pi} =$$

$$= -\sum_{(i,j)\in\mathcal{N}} \frac{e^{\lambda_{ij}} - 1}{\pi(e^{\lambda_{ij}} - 1) + 1} + \sum_{(i,j)\notin\mathcal{N}} \frac{1}{1-\pi}$$

$$\nabla^2_\pi = \sum_{(i,j)\in\mathcal{N}} \left(\frac{e^{\lambda_{ij}} - 1}{\pi(e^{\lambda_{ij}} - 1) + 1}\right)^2 + \sum_{(i,j)\notin\mathcal{N}} \frac{1}{(1-\pi)^2}$$

We combine all the steps in the *ZIPoisMS algorithm* stated below.

Zero-inflated Poisson metric scaling (*ZIPoisMS*)

Repeat until convergence:

1. For the current guess of X, β and π compute SOA:
   - evaluate $\Lambda = e^{-D^2(X)+\beta}$ and $a = \frac{\pi}{1-\pi}$
   - calculate
   $$\nabla_{D^2} = -\Gamma * \frac{\Lambda}{ae^{\Lambda}+1} + (1-\Gamma)*(C-\Lambda)$$
   $$\nabla^2_{D^2} = \Gamma * \Lambda * \frac{ae^{\Lambda}(1-\Lambda)+1}{(ae^{\Lambda}+1)^2} + (1-\Gamma)*\Lambda$$
   - compute $W = \nabla^2$ and $Z = D^2(X) - \frac{\nabla}{\nabla^2}$
2. Solve *WPCMS* problem with W and Z thereby updating X.
3. For fixed X, π compute $a = \frac{\pi}{1-\pi}$ and run Newton's method to update β.
   Repeat until convergence:
   - evaluate $\Lambda = e^{-D^2(X)+\beta}$
   - calculate
   $$\nabla_\beta = \sum_{(i,j)\in\mathcal{N}} \frac{\lambda_{ij}}{ae^{\lambda_{ij}}+1} - \sum_{(i,j)\notin\mathcal{N}} (C_{ij}-\lambda_{ij})$$
   $$\nabla^2_\beta = \sum_{(i,j)\in\mathcal{N}} \lambda_{ij}\frac{ae^{\lambda_{ij}}(1-\lambda_{ij})+1}{(ae^{\lambda_{ij}}+1)^2} + \sum_{(i,j)\notin\mathcal{N}} \lambda_{ij}$$
   - update $\beta := \beta - \frac{\nabla}{\nabla^2}$.
4. For fixed X, β compute $\Lambda = e^{-D^2(X)+\beta}$ and run Newton's method to update π.
   Repeat until convergence:
   - calculate
   $$\nabla_\pi = -\sum_{(i,j)\in\mathcal{N}} \frac{e^{\lambda_{ij}}-1}{\pi(e^{\lambda_{ij}}-1)+1} + \sum_{(i,j)\notin\mathcal{N}} \frac{1}{1-\pi}$$
   $$\nabla^2_\pi = \sum_{(i,j)\in\mathcal{N}} \left(\frac{e^{\lambda_{ij}}-1}{\pi(e^{\lambda_{ij}}-1)+1}\right)^2 + \sum_{(i,j)\notin\mathcal{N}} \frac{1}{(1-\pi)^2}$$
   - update $\pi := \pi - \frac{\nabla}{\nabla^2}$.

F.3. Negative Binomial. We extend the methodology to the negative binomial model. We start with taking the derivatives of the loss function

$$\ell_{NBMS}(X,\beta,r;C) = \frac{1}{n^2}\Big[\sum_{1\le i,j\le n} \log\Gamma(r) - \log\Gamma(C_{ij}+r) - r\log r + \\ + \sum_{1\le i,j\le n} (C_{ij}+r)\log(\lambda_{ij}+r) - C_{ij}\log\lambda_{ij}\Big]$$

w.r.t. D^2 leads us to the following equations (up to $\frac{1}{n^2}$ scale)

$$\nabla_{D^2} = -\frac{(C+r)*\Lambda}{\Lambda+r} + C = r\frac{C-\Lambda}{\Lambda+r}$$
$$\nabla^2_{D^2} = r\frac{\Lambda*(\Lambda+r)+\Lambda*(C-\Lambda)}{(\Lambda+r)^2} = r\frac{\Lambda*(C+r)}{(\Lambda+r)^2}$$

We will use these derivatives to calculate the *WPCMS* parameters and, subsequently, update the reconstruction X. Next, we calculate the derivatives w.r.t. β (up to $\frac{1}{n^2}$ scale)

$$\nabla_\beta = -r \sum_{1\leq i,j\leq n} \frac{C_{ij}-\lambda_{ij}}{\lambda_{ij}+r} \text{ and } \nabla^2_\beta = r \sum_{1\leq i,j\leq n} \frac{\lambda_{ij}(C_{ij}+r)}{(\lambda_{ij}+r)^2}$$

and use them to update β via Newton's method. Finally, to find optimal r we compute the derivatives w.r.t. this nuisance parameter (up to $\frac{1}{n^2}$ scale):

$$\nabla_r = \sum_{1\leq i,j\leq n} \psi_0(r) - \psi_0(C_{ij}+r) - \log r + \log(\lambda_{ij}+r) + \frac{C_{ij}-\lambda_{ij}}{\lambda_{ij}+r}$$

$$\nabla^2_r = \sum_{1\leq i,j\leq n} \psi_1(r) - \psi_1(C_{ij}+r) - \frac{1}{r} + \frac{1}{\lambda_{ij}+r} - \frac{C_{ij}-\lambda_{ij}}{(\lambda_{ij}+r)^2}$$

Here $\psi_0(\cdot)$ and $\psi_1(\cdot)$ correspond to the di- and tri-gamma function. We conclude this section with the *NBMS* algorithm.

Negative binomial metric scaling (*NBMS*)

Repeat until convergence:

1. For the current guess of X, β and r compute SOA:
   - evaluate $\Lambda = e^{-D^2(X)+\beta}$
   - calculate $\nabla_{D^2} = r\frac{C-\Lambda}{\Lambda+r}$ and $\nabla^2_{D^2} = r\frac{\Lambda*(C+r)}{(\Lambda+r)^2}$
   - compute $W = \nabla^2$ and $Z = D^2(X) - \frac{\nabla}{\nabla^2}$
2. Solve *WPCMS* problem with W and Z thereby updating X.
3. For fixed X, r run Newton's method to update β.
   Repeat until convergence:
   - evaluate $\Lambda = e^{-D^2(X)+\beta}$
   - calculate $\nabla_\beta = -r\sum_{1\leq i,j\leq n}\frac{C_{ij}-\lambda_{ij}}{\lambda_{ij}+r}$ and $\nabla^2_\beta = r\sum_{1\leq i,j\leq n}\frac{\lambda_{ij}(C_{ij}+r)}{(\lambda_{ij}+r)^2}$
   - update $\beta := \beta - \frac{\nabla}{\nabla^2}$.
4. For fixed X, β compute $\Lambda = e^{-D^2(X)+\beta}$ and run Newton's method to update π.
   Repeat until convergence:
   - calculate
   $$\nabla_r = \sum_{1\leq i,j\leq n} \psi_0(r) - \psi_0(C_{ij}+r) - \log r + \log(\lambda_{ij}+r) + \frac{C_{ij}-\lambda_{ij}}{\lambda_{ij}+r}$$
   $$\nabla^2_r = \sum_{1\leq i,j\leq n} \psi_1(r) - \psi_1(C_{ij}+r) - \frac{1}{r} + \frac{1}{\lambda_{ij}+r} - \frac{C_{ij}-\lambda_{ij}}{(\lambda_{ij}+r)^2}$$
   - update $r := r - \frac{\nabla}{\nabla^2}$.

F.4. Normal. As we mentioned in the main paper, if σ and β are known and fixed then conformation X can be found via optimizing $\ell_{\mathit{WPCMS}}(X; -\widetilde{C}+\beta, 1)$. At the same time, if X is fixed then optimal nuisance parameters have explicit form. These observations underlie the *NMS* algorithm.

Normal metric scaling (*NMS*)

Repeat until convergence:

1. For the current guess of β we solve the *WPCMS* problem with unit weights and working response $-\widetilde{C}+\beta$ thereby updating X.
2. For fixed X update $\beta := \frac{1}{n^2}\sum_{1\le i,j\le n}(\|x_i - x_j\|^2 + \widetilde{C}_{ij})$.
3. For fixed X, β update $\sigma^2 := \frac{1}{n^2}\sum_{1\le i,j\le n}(\widetilde{C}_{ij} + \|x_i - x_j\|^2 - \beta)^2$.

F.5. Exponential. As usual, we begin with taking the derivatives of the negative log-likelihood for the exponential model

$$\ell_{EMS}(X,\beta;C) = \frac{1}{n^2}\sum_{1\le i,j\le n} \widetilde{C}_{ij}\lambda_{ij} - \log(\lambda_{ij})$$

w.r.t. D^2. This leads us to the following equations (up to $\frac{1}{n^2}$ scale)

$$\nabla_{D^2} = \widetilde{C} * \Lambda - 1 \text{ and } \nabla^2_{D^2} = \widetilde{C} * \Lambda$$

Thus, the derivatives w.r.t. β (up to $\frac{1}{n^2}$ scale) are

$$\nabla_\beta = 1 - \sum_{1\le i,j\le n} \widetilde{C}_{ij}\lambda_{ij} \text{ and } \nabla^2_\beta = \sum_{1\le i,j\le n} \widetilde{C}_{ij}\lambda_{ij}$$

These derivations lead to the *EMS* algorithm.

Exponential metric scaling (*EMS*)

Repeat until convergence:

1. For the current guess of X, β compute SOA:
   - evaluate $\Lambda = D^2(X) - \beta$
   - calculate $\nabla_{D^2} = \widetilde{C} * \Lambda - 1$ and $\nabla^2_{D^2} = \widetilde{C} * \Lambda$
   - compute $W = \nabla^2$ and $Z = D^2(X) - \frac{\nabla}{\nabla^2}$
2. Solve *WPCMS* problem with W and Z thereby updating X.
3. For fixed X run Newton's method to update β.
   Repeat until convergence:
   - evaluate $\Lambda = D^2(X) - \beta$
   - calculate $\nabla_\beta = 1 - \sum_{1\le i,j\le n} \widetilde{C}_{ij}\lambda_{ij}$ and $\nabla^2_\beta = \sum_{1\le i,j\le n} \widetilde{C}_{ij}\lambda_{ij}$
   - update $\beta := \beta - \frac{\nabla}{\nabla^2}$.

APPENDIX G: SMOOTH DISTRIBUTION-BASED METRIC SCALING (*SDBMS*)

By analogy with the *SPoisMS* loss (9) one can combine the smoothing spline technique with general distribution-based metric scaling leading to the *SDBMS* loss

$$\ell_{SDBMS}(X,\Omega;C,K,\lambda) = \ell_{DBMS}(X,\Omega;C) + \lambda\,\mathrm{tr}(X^\top K X) \tag{5}$$

Following the outline from Section (12) one can build the optimization algorithm as follows.

First, we recall that $\mathrm{tr}(KS(X)) = -\frac{1}{2}\mathrm{tr}(KD^2(X))$, so the penalty term in the *SDBMS* loss is a linear function of $D^2(X)$. The updated first derivative involved in the second order approximation is therefore $\tilde{\nabla} = \nabla - \frac{\lambda}{2}K$ while the second derivative ∇^2 stays the same. The new working response matrix involved in the *WPCMS* is $\tilde{Z} = Z + \frac{\lambda}{2}\frac{K}{W}$. Finally, as $\Phi(K) = K$ the gradient step in PGD update is

$$\tilde{S} = XX^\top - \Phi(W * (Z + \frac{\lambda}{2}\frac{K}{W} - D^2(X))) = S + \frac{\lambda}{2}K$$

whereas the projection step minimizes the loss

$$\ell_{PCMS}(X;\tilde{S}) = \|S - XX^\top\|_F^2 + \lambda\,\mathrm{tr}(X^\top K X).$$

In other words, adding the smoothing penalty to the original *DBMS* problem is equivalent to replacing the *PCMS* projection step with its smoothing spline analog.

APPENDIX H: EXPECTED COUNTS FOR DBMS

In the main paper we employ the log-likelihood metric to assess the performance of our methods. Alternatively, one can evaluate the quality of the resulting reconstructions by computing the expected values for the count counts, which are λ_{ij} for *PoisMS*, *SPoisMS*, and *NBMS*; $(1-\pi)\lambda_{ij}$ for *ZIPoisMS*; and $(1-\pi)\frac{\lambda_{ij}}{1-e^{-\lambda_{ij}}}$ for *HPoisMS*, and quantifying the similarities between them and the observed counts C_{ij}. In Figure 4, we illustrate the dependence of the correlation between expected and observed counts on degrees-of-freedom for the five models proposed in this paper. We consider two correlation metrics: Pearson and stratified. We employ the method suggested in Section 7 to align the *SPoisMS* correlations with the ones obtained by *DBMS*. According to the Pearson correlation plot, *HPoisMS* continues to exhibit the highest training performance; however, the model with the lowest correlation, in this case, is *NBMS*.

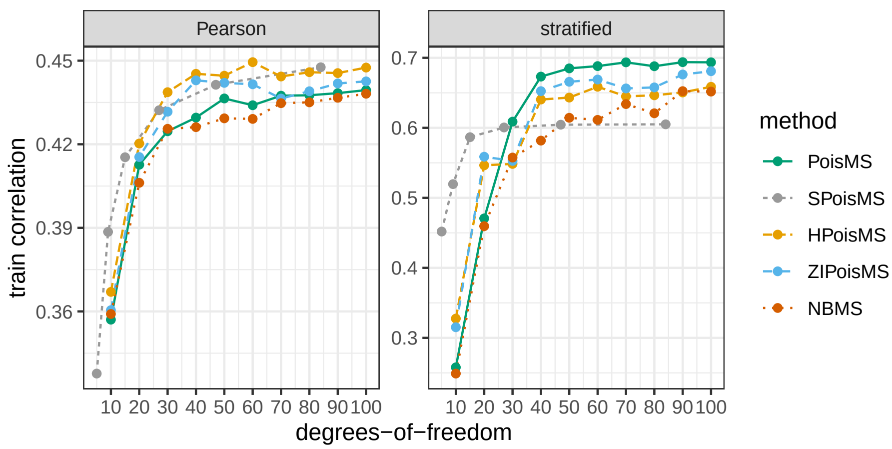


FIG 4. *Comparing five models via the bulk cell data. Each model was trained on the full contact matrix and the train correlation between expected and observed counts is reported. The best train Pearson correlation is achieved by the HPoisMS model, which is consistent with the findings in the left panel of Figure 8. However, NBMS is now the least favorable model in terms of the Pearson correlation metric.*

APPENDIX I: MFISH CONFORMATION

We provide a 3D image of three MFISH replicates. These replicates have numbers 1, 1000, and 1500 in the data supplementing the paper by Su et al. (2020). The substantial noise in the conformations is evident from the plot.

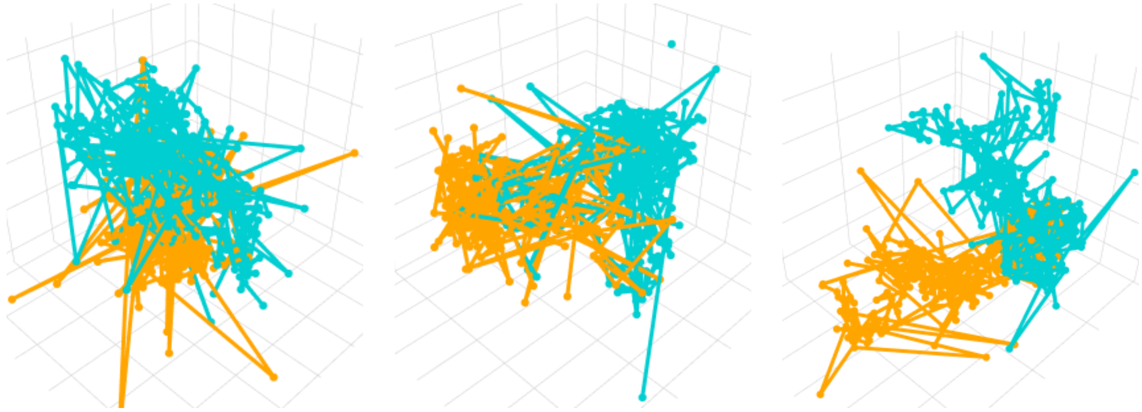


FIG 5. *MFISH replicates number 1, 1000 and 1500.*

APPENDIX J: DBMS REPRODUCIBILITY

We present the results on the reproducibility of the *DBMS* methodology using Hi-C data from biological replicates for resolutions 100kb, 50kb, 25kb, and 10kb. The primary Hi-C matrix C_{train} was used to compute the reconstruction X and the nuisance parameters. The stratified correlation between expected counts E and C_{test} are illustrated in the top panels of Figure 6. The bottom panels show the results when the roles of the replicate and primary matrices are exchanged.

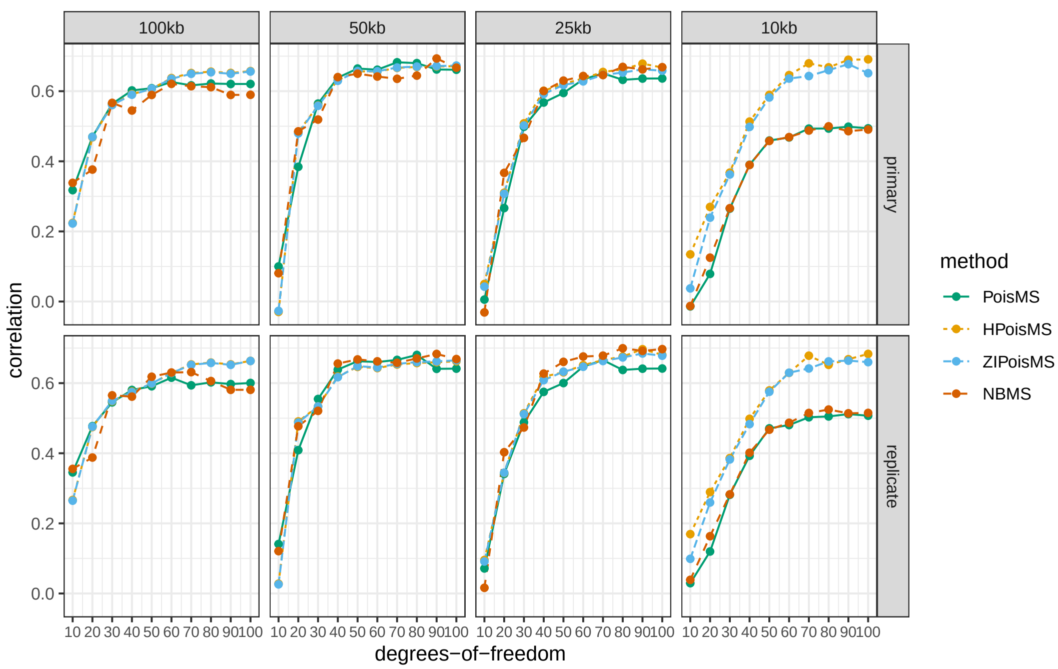


FIG 6. *Agreement between the observed and expected counts measured via stratified correlation coefficient and computed for resolutions 100kb, 50kb, 25kb, and 10kb. Each model was trained on C_{train}, which is the primary (top panels) or the replicate (bottom panels) contact matrix, and the correlation between expected and observed counts was measured on C_{test}, which is the replicate (top panels) or the primary (bottom panels) contact matrix.*

The plot shows high values of stratified correlation thereby validating the ability of the *DBMS* reconstructions to reproduce the replicate and primary contact matrices. In addition, we observe that the *ZIPoisMS* and *HPoisMS* models significantly improve the correlation value for high-resolution data, i.e. 10kb.

APPENDIX K: CONTINUOUS DBMS MODELS FOR NORMALIZED COUNTS

In this section, we examine two models for normalized contact counts: normal and exponential. The normalization of C was performed via Knight-Ruiz (KR) algorithm, and the heatmap of $\widetilde{C}$ is presented in Figure 8a. We begin by selecting the optimal degrees-of-freedom for both models using the elbow heuristic as shown in the left panel of Figure 7. The log-likelihood values in the plot suggest that the exponential distribution fits the observed normalized counts better that the normal distribution.

Next, we compute the expected count values as $\mu_{ij} = -\|x_i - x_j\|^2 + \beta$ for *NMS* and $\frac{1}{\lambda_{ij}} = e^{-\|x_i - x_j\|^2 + \beta}$ for *EMS*. We assess the performance of both methods by measuring the Pearson correlation between the expected and observed normalized counts. The dependence of the correlation values on the degrees-of-freedom is presented in the right panel of Figure 7. We note that the *EMS* model obtains higher correlation than the *NMS* model, which aligns with the conclusion drawn from the log-likelihood values.

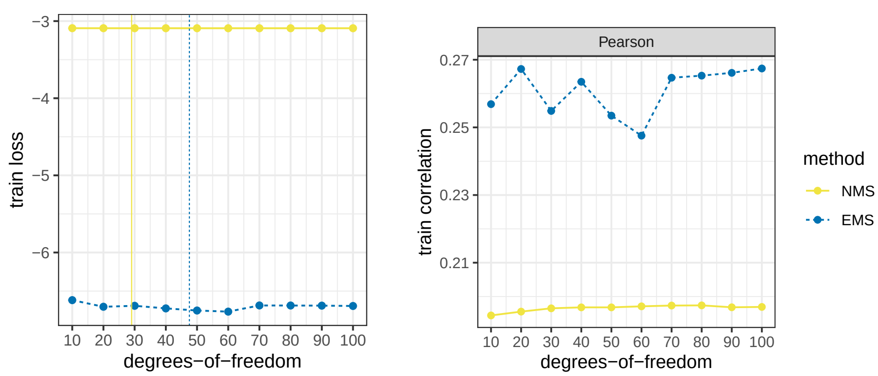
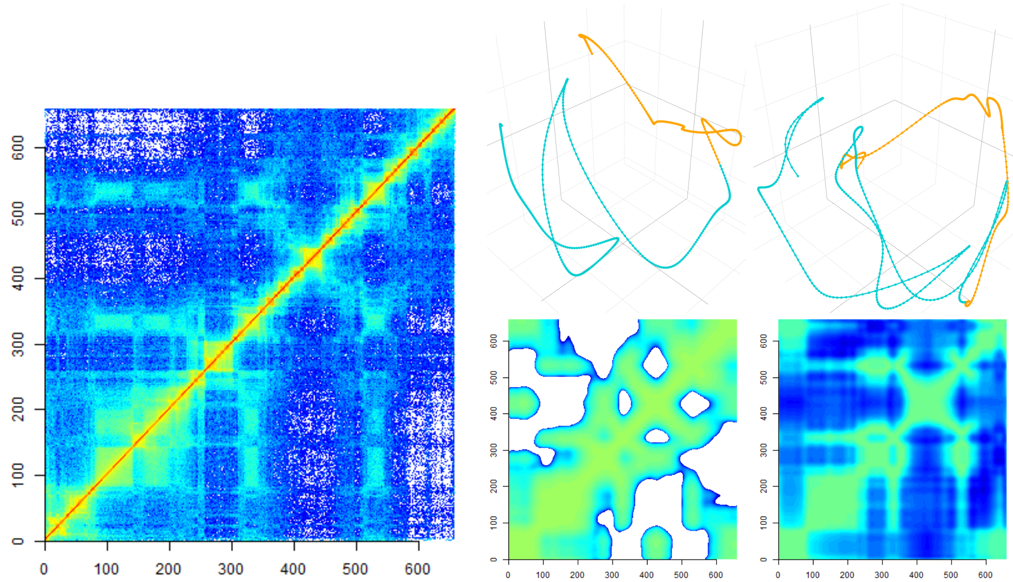


FIG 7. *Comparing NMS and EMS via the bulk cell data. Left panel: each model was trained on the full contact matrix and the train score is reported. The dashed lines represent the optimal degrees-of-freedom values detected for each model by the elbow method (df ≈ 30 for NMS, and df ≈ 50 for EMS). Right panel: each model was trained on the full contact matrix and the train correlation between expected and observed counts is reported.*

The heatmaps for expected counts as well as the resulting reconstructions produced by *NMS* and *EMS* with optimal degrees-of-freedom are presented in Figure 8. By analyzing the heatmaps, we infer that both models are capable of capturing such artifacts of $\widetilde{C}$ as block structure. However, unlike *EMS*, *NMS* results in a large portion of negative expected counts, which appears due to the full support of normal distribution. These negative values may serve as an explanation for the weaker performance of *NMS*.

Following the procedure from Section 15 we validate the *NMS* and *EMS* reconstructions by means of the MFISH dataset. In Figure 7, we present the “alternative” distributions that summarise the dissimilarity between the *NMS* and *EMS* reconstructions and MFISH replicates. Additionally, Figure 7 includes the “null” distribution that reflects the dissimilarities between the MFISH replicates. According to the plot, both continuous *DBMS* models result in reconstructions that, on average, demonstrate a higher degree of agreement to the MFISH dataset than an arbitrary chosen MFISH replicate.



(a) Normalized contact matrix (b) Normal (c) Exponential

FIG 8. *(a) Logarithm of the observed normalized contact counts represented as a heatmap; values* $\widetilde{C}_{ij} = 0$ *are presented in white. (b-c) Upper row : the projections of the resulting reconstructions obtained via two methods: NMS and EMS. The degrees-of-freedom is set to the value picked by the elbow method. (b-c) Bottom row: logarithm of the expected contact counts represented as a heatmap; to make the results compatible with (a), we use the same color range and show negative log-values in white.*

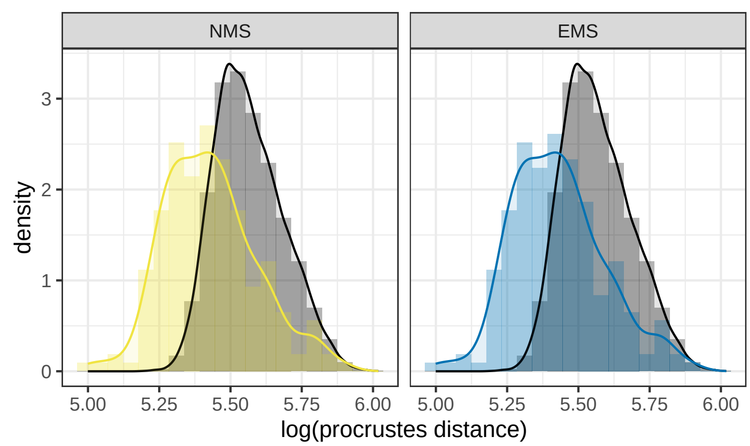


FIG 9. *Validation using the MFISH dataset. Each panel corresponds to a reconstruction method, either NMS or EMS. The black histogram illustrates the null distribution representing the variation in Procrustes distances between each pair of MFISH replicates. The colored histograms represent the variation in the dissimilarity between one of the evaluated reconstructions and the MFISH replicates.*

To compare the continuous and discrete *DBMS* models, we proceed as follows. We take the expected contact count matrices E produced by *PoisMS*, *HPoisMS*, *ZIPoisMS* and *NBMS* with the optimal degrees-of-freedom. We apply the Knight-Ruiz balancing algorithms to them, thus obtaining $\widetilde{E}$. We measure the Pearson correlation between $\widetilde{E}$ and normalized observed counts $\widetilde{C}$. Additionally, we compute the correlation between $\widetilde{C}$ and the expected counts obtained by the optimal *NMS* and *EMS* models. We present the resulting correlation values in Table 1. According to the table, even thought the continuous models were fitted directly to the normalized counts, they achieve significantly lower correlation values than the discrete models. This suggests that the proposed discrete models followed by the KR-normalization result in a more adequate fit to $\widetilde{C}$, thereby providing a solid foundation for modeling normalized contact matrices.

method	*PoisMS*	*HPoisMS*	*ZIPoisMS*	*NBMS*	*NMS*	*EMS*
correlation	0.453	0.458	0.45	0.46	0.196	0.25

TABLE 1
Comparison of the performance of the continuous and discrete DBMS models. The table represents the Pearson correlation between the expected and observed normalized counts.

To make sure that the high correlation of discrete methods in Table 1 is not a result of overfitting, we repeated the reproducibility assessment from Section J. Specifically, we fit the discrete *DBMS* methods on the primary C_{train}, compute expected contact matrices E, and normalize E via Knight-Ruiz. We then measure the correlation between $\widetilde{E}$ and the normalized replicate matrix $\widetilde{C}_{test}$. For continuous methods, we fit them directly on the normalized primary matrix $\widetilde{C}_{train}$, compute expected contact matrices E and correlate them with $\widetilde{C}_{test}$. The resulting correlation values are presented in Figure 10 and confirm the superior performance of the discrete *DBMS* models.

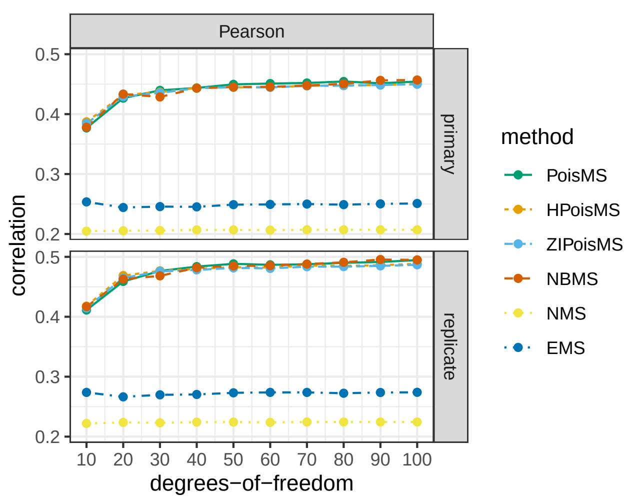


FIG 10. *Evaluating reproducibility of continuous and discrete DBMS models via the GM12878 data. Top panels: each model was trained on the primary contact matrix, and the correlation between normalized expected and normalized observed counts was measured on the replicate contact matrix. Bottom panels: each model was trained on the replicate contact matrix, and the correlation between normalized expected and normalized observed counts was measured on the primary contact matrix. The correlation values suggest that the discrete models achieve higher correlation than the continuous one.*

APPENDIX L: DBMS COMPLEXITY

Below, we illustrate the relationship between the number of epochs as well as the total number of iterations and the spline basis size for each of the four *DBMS* approaches. Recall that in the *DBMS* algorithm, one epoch corresponds to a single update of the second-order approximation matrices W and Z, which is equivalent to one step in Newton's method. An iteration, on the other hand, represents one step of the projected gradient descent in *WPCMS*. We sum the number of iterations across all the *DBMS* epochs.

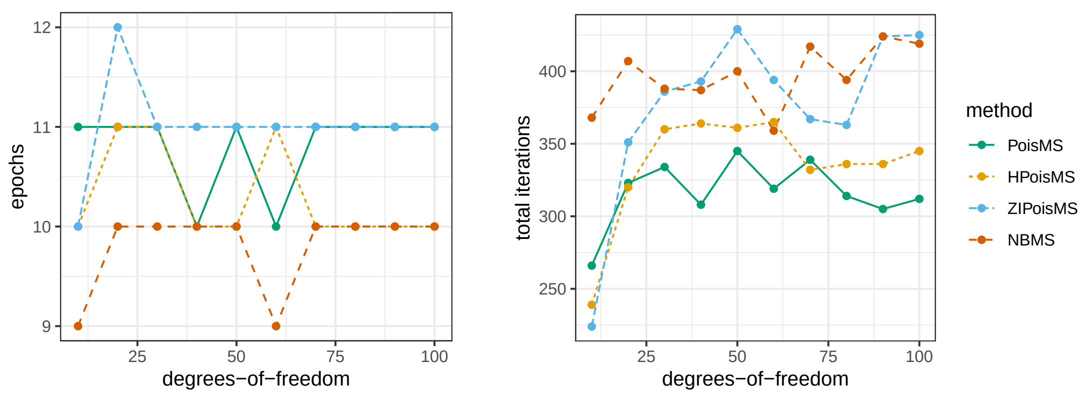


FIG 11. *Number of epochs (left) and the total number of iterations (right) required for the convergence of the PoisMS, HPoisMS, ZIPoisMS, and NBMS algorithms vs. degrees-of-freedom. Convergence is achieved when the relative change in the negative log-likelihood reaches a precision level of* $\epsilon = 10^{-5}$. *This criterion is met for each of the DBMS approaches within 9-12 epochs and 200-500 iterations.*

REFERENCES

GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall, United Kingdom.

SU, J. H., ZHENG, P., KINROT, S. S., BINTU, B. and ZHUANG, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182** 1641-1659.