

Approximate Bayesian Computation for Inferring Waddington Landscapes from Single Cell Data

Yujing Liu¹, Stephen Y. Zhang¹, Istvan T. Kleijn²
and Michael P.H. Stumpf^{1,3,4}

¹School of Mathematics and Statistics,
University of Melbourne, Australia

²Institute of Cancer Research, London, UK

³School of BioScience, University of Melbourne, Australia

⁴ARC Centre of Excellence for the
Mathematical and Statistical Analysis of Cellular Systems (MACSYS)
For correspondence: mstumpf@unimelb.edu.au

September 4, 2023

Abstract

Single cell technologies allow us to gain insights into cellular processes at unprecedented resolution. In stem cell and developmental biology snapshot data allows us to characterise how the transcriptional state of cells changes between successive cell types. Here we show how approximate Bayesian computation (ABC) can be employed to calibrate mathematical models against single cell data. In our simulation study we demonstrate the pivotal role of the adequate choice of distance measures appropriate for single cell data. We show that for good distance measures, notably optimal transport distances, we can infer parameters for mathematical models from simulated single cell data. We show that the ABC posteriors can be used to characterise parameter sensitivity and identify dependencies between different parameters, and to infer representations of the Waddington or epigenetic landscape, which forms a popular and interpretable representation of the developmental dynamics. In summary, these results pave the way for fitting mechanistic models of stem cell differentiation to single cell data.

1 Introduction

While developmental biology has largely progressed through observational studies, from the beginning of the 20th century this seemingly intricate, bewilderingly complex, yet robust, process has also fascinated mathematicians. Of particular importance has been C.H. Waddington's epigenetic landscape[24]. Originally intended as a metaphor for developmental processes, the idea of a landscape, has continued to guide thinking among biologists and mathematicians; the Fields medalist René Thom, for example, was interested in developmental biology as a potential manifestation of catastrophe theory.

Over the past decade there has been a resurgence in interest from mathematicians, developmental biologists, and bioengineers in the epigenetic or Waddington landscape. We can use it as a mathematical and computational tool to reason about and even predict cell fates[31, 52]. Recent studies [12, 40, 36, 16] have given us qualitative insights into the fundamental dynamics underlying differentiation at the cellular level. In a deterministic framework it was shown that even fundamental developmental dynamics can be understood in terms of elegant mathematical models of small gene regulation networks [36]. But to account for the (experimentally quantifiable) randomness prevailing among sub-cellular molecular processes we must extend such analyses to incorporate stochastic processes. This has become especially important because stochastic effects can change the dynamics not just quantitatively but qualitatively, profoundly reshaping the manifold on which dynamics occur. In specific, when external multiplicative noise are present, cell states represented by attractors in the landscape can be destroyed or created [23, 16].

While the mathematical analysis of dynamical systems has achieved a high level of maturity, our ability to challenge mathematical models of developmental systems with data is lagging behind both the mathematical theory and our capability to probe developmental systems in experiment. Technological advances in single cell biology provide us with exquisitely detailed snapshots of the transcriptomic states of single cells [30]; and before long we will also be able to collect single cell protein data of the required quality and quantity [37]. And while descriptive and statistical analyses of single cell data have progressed in lock-step with experimental technologies and new data, mechanistic modelling in light of data has progressed more slowly.

Landscapes for dynamical system models of developmental systems are typically formulated by a specific manner through thermodynamic approaches with detailed chemical assumptions about kinetics of gene transcription factors [43]. But the inference of parameters underpinning the landscapes remains an open problem. It is well known that variation of the parameters of a dynamical system can lead to changes in the cellular state or even the structure of the landscapes: for example, at some critical value of parameters, creation or destruction of attracting states may occur which leads to different landscape structures. Such changes in dynamical system due to the variation of parameters are known as bifurcations [25, 12] and are a central phenomenon of study in the dynamical systems literature. Therefore, to reconstruct the landscapes that agree experimental observations, it is necessary to have parameters adequately inferred.

Coupling single cell data to modelling has been challenging as it has been difficult to e.g. estimate reaction rate parameters for dynamical systems models. Here we explore the use of approximate Bayesian computation (ABC) [49] as a tool to estimate parameters for dynamical systems describing stem cell differentiation. In cases where conventional likelihood-based approaches fail or are difficult to apply because the likelihood is intractable, ABC methods can provide (approximate) answers.

Most current single cell data provide snapshots of the expression state of a system. Temporal information can typically only be gleaned indirectly, and a host of statistical approaches have been used to remedy this situation [39]. Each of these methods has, however, associated degrees of uncertainty. Here we take a complementary approach where data are assumed to have been generated by a stochastic dynamical system (at different times). We show below that

this approach can be used to infer model parameters and determine parameter sensitivities for models of stem cell differentiation. We show that the manner in which we summarise the data – or calculate distances between observed and simulated data – affects the efficiency and reliability of parameter inference profoundly.

2 Methods

2.1 Waddington Landscapes and Quasi-Potentials for Developmental Processes

Mathematically we can characterize the evolution of cell states, X_t , over time t by a stochastic differential equation (SDEs),

$$dX_t = \mu(X_t; \theta)dt + \sigma(X_t; \theta)dW_t \quad (1)$$

where $\mu(X_t; \theta)dt$ captures the deterministic dynamics of our system parameterized by θ and $\sigma(X_t; \theta)dW_t$ depicts the random variations or stochastic properties of this system parameterized by θ .

In many cases the focus has been on gradient systems, where we have a potential function, $U(X)$, such that

$$\frac{dX_t}{dt} = \mu = -\nabla U(X_t) \quad (2)$$

The potential function describes the deterministic dynamics of cell states at each time point and determines the attracting states. The choice of the stochastic dynamics can distort this potential profoundly.

While gradient dynamical systems are flexible and popular – they have also been argued to capture hall-marks of more general developmental systems inspired by developmental biology – they cannot describe all aspects, especially oscillatory and clock-like processes, including the cell cycle [10]. We nevertheless seek to capture many of the dynamics in terms of landscape descriptions, but we make the approximate nature explicit by referring to the mathematical description as the *quasi potential*, $\tilde{U}(X)$.

We generally cannot determine $U(X)$ or $\tilde{U}(t)$ analytically and instead rely on simulations. We can determine the approximate quasi-potential via the probability density function (p.d.f) $f(X_t)$, which contains the information about how probable state X_t is. The quasi-potential is then given by t [53],

$$\tilde{U}(X) = -\log p(X) \quad (3)$$

The quasi-potential is thus the logarithm of the probability that a cell is in state X [4].

2.2 ABC and ABC-SMC

In a Bayesian framework we estimate the posterior probability of a model's parameters via

$$P(\theta|D) \propto f(D|\theta)P(\theta)$$

where $f(D|\theta)$ is the likelihood of θ given the data D and where $P(\theta)$ is the prior for the model parameters, θ . In ABC the calculation of the likelihood is replaced by a comparison between the observed data and simulated data. ABC methods have the following generic form[2]:

Algorithm 1 ABC rejection algorithm

```

Require:  $\epsilon > 0$ ,  $N > 0$ 
 $i \leftarrow 0$ 
while  $i < N$  do
     $\theta^* \sim P(\theta)$ .            $\triangleright$  Sample a candidate parameter  $\theta^*$  from proposal
    distribution  $P(\theta)$ 
     $X \sim f(D|\theta^*)$   $\triangleright$  Simulate a data set  $X$  from the model with parameter  $\theta^*$ 
    if  $d(D, X) \leq \epsilon$  then
         $\theta_i \leftarrow \theta^*$             $\triangleright$  accept  $\theta^*$ 
         $i \leftarrow i + 1$ 
    end if
end while
```

where d is a distance function and the tolerance $\epsilon \geq 0$ defines the intended degree of alignment between D and X .

The result of any ABC inference is a set of parameter samples from the approximate posterior distribution $P(\theta|d(D, X) \leq \epsilon)$. When ϵ is sufficiently small then the distribution $P(\theta|d(D, X) \leq \epsilon)$ can be regarded as a reliable approximation for the “actual” posterior distribution $P(\theta|D)$ [2].

The basic algorithm outlined above, known as ABC rejection, is typically too slow for any problem of interest with more than a few parameter. A range of improvements have been developed in the literature, including Markov chain Monte Carlo, and Sequential Monte Carlo (SMC) approaches. Here we use the SMC approach of Toni et al. [49] adapted to the case of single cell data obtained from simulations of dynamical systems as the data collected from true developmental systems. In this case a key insight is to find a good distance appropriate for single cell data. We will discuss this in the next section.

2.3 Developmental model and simulation procedure

Our analysis will focus on a model that has been used to model embryonic stem cell differentiation processes [15], see Figure 1, where each line represents non-linear inhibition or promotion among these factors. In specific, this model is characterised by the temporal dynamics of four transcription factors: Nanog(N), Oct4-Sox2(O), Fgf4(F) and Gata6(G) where Nanog, Oct4 and Sox2 are complex for maintaining pluripotency and Gata6 is a standard bio-marker for cell differentiation. The detailed representation of their relationship is shown in Figure 1 where each line represents non-linear inhibition or promotion among these factors.

We will model the stochastic feature of this processes by implementing a Wiener process on each factor [9].

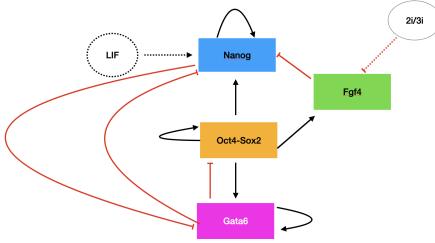


Figure 1: The transcription factor interaction relationship circuit with external factors influences where red line stands for inhibition and black line stands for promotion

This system gives rise to two distinct cell states: (i) the pluripotent state is characterised by high values of Nanog and low levels of Gata6; (ii) the differentiated state is characterised by the opposite behaviours for both genes (see Figure 3). The deterministic part of the system is given by [15, 9]

$$\begin{aligned} \frac{d[N]}{dt} &= \frac{k_0[O](c_0 + c_1[N]^2 + k_0[O] + c_2LIF)}{1 + (k_0[O](c_1[N]^2 + k_0[O] + c_2LIF + c_3[F]^2)) + c_4[O][G]^2} - \gamma[N] \\ \frac{d[O]}{dt} &= \alpha + \frac{e_0 + e_1[O]}{1 + e_1[O] + e_2[G]^2} - \gamma[O] \\ \frac{d[F]}{dt} &= \frac{a_0 + a_1[O]}{1 + a_1[O]} - \gamma[F] \\ \frac{d[G]}{dt} &= \frac{b_0 + b_1[G]^2 + b_3[O]}{1 + b_1[G]^2 + b_2[N]^2 + b_3[O]} - \gamma[G] \end{aligned} \quad (4)$$

where the value of parameters we will use in this model, according to Chickarmane [15], are:

$$\begin{aligned} a_0 &= 0.01, a_1 = 1 \\ b_0 &= 0.005, b_1 = 0.005, b_2 = 1, b_3 = 1 \\ c_0 &= 0.01, c_1 = 0.4, c_2 = 1, c_3 = 0.1, c_4 = 0.00135 \\ e_0 &= 0.01, e_1 = 1, e_2 = 1 \\ k_0 &= 0.005, \gamma = 0.01 \end{aligned} \quad (5)$$

Other than the four transcription factors, there are also some external signals known to influence this system and the propensity functions given above. Leukemia inhibiting factors (LIF) is a cytokine known to inhibit cell differentiation [22] by having a positive effect on the maintenance of Nanog levels. When the level of LIF decreases, the cells will differentiate (Figure 2). Similarly, i2/i3 which are represented by I_3 in our propensity functions, are two different sets of small molecule inhibitors found to be useful in maintain pluripotency in vivo[54]. In our model, increase of I_3 will lead to the suppression of Fgf4 which in turn reduces the effect of suppression on Nanog and therefore sustain the pluripotency. Furthermore, this model also includes the possibility of reprogramming by introducing signal α in the second propensity function. It can be referred as the reprogramming rate which is in charge of the loop of positive feedback and provides the possible backward transition from the differentiated state back into

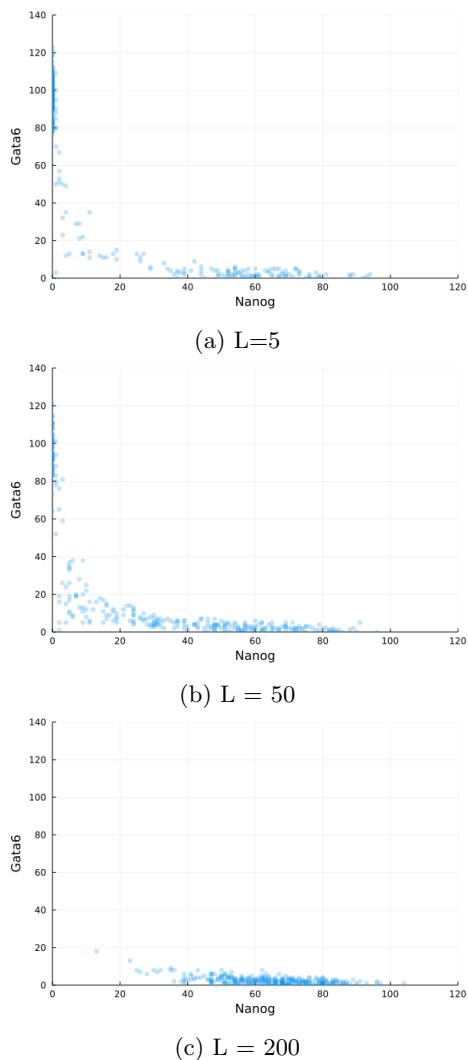


Figure 2: The scatter plots of the relationships between Nanog and Gata6 at the end time for different values of LIF

the pluripotent state: When α reaches the critical value, Oct4-Sox2 will induce Nanog and keep raising it towards sufficiently high value and in consequence it will reduce Gata6 level by antagonism[15]. For simplicity, here we only consider the effect of LIF, which is more significant than others, and I3 & α are assumed to be zero throughout this analysis. The level of LIF was chosen as fifty for both reference data and simulations in ABC-SMC.

2.3.1 Simulation procedure

In practice it may be impossible to obtain suitable time course experimental observations. Instead, experimental observations are given as snapshots covering cells at different stages of development or differentiation. To generate our test data we sample cells at ten equally spaced time points. This leads to a dataset that mimics aspects of real experiments [48, 14]. Our ABC-SMC procedure is adapted to this sampling scheme (our reference data and all the simulation data in ABC-SMC were $4 \times 10 \times 300$ (4 factors; 10 time points and 300 trajectories)). The distance is obtained by summing over all ten time points (i.e. $d = \sum_{i=1}^{10} d(D_i^*, D_i)$, where D_i and D_i^* are 4×300 matrices).

For the stochastic simulations (using the Gillespie algorithm[47]) we use the Julia packages *Catalyst.jl* and *DifferentialEquations.jl*. For the ABC-SMC algorithm we used *ApproxBayes.jl*.

We choose the sequence of tolerance threshold ϵ_t based on the 30% quantiles of the population of particles in the last iteration until it reaches the target tolerance threshold ϵ_T or exceeds the set maximum number of simulations (10^6). That is $\epsilon_t = q_{\alpha=0.3}(\theta_{t-1})$, $t \in [1, T]$. We seek to accept 1000 particles of parameters θ_t in each population t of the ABC-SMC algorithm. For Kernel Density Estimation with multivariate normal distributions we use *KDEstimation.jl* and we set the bandwidth according to Silverman's rule. And the computation of Optimal Transport is performed in *OptimalTransport.jl* in *Julia*.

For all parameters, uniform prior distributions, with lower bound given by 1/10 of the true parameters in Equation 5 and upper bounded given by 10 times the true parameters, are used. Since all the data are acquired by simulations rather than from real experimental results, we can choose the final tolerance threshold ϵ_T by simulating data with exactly the same parameters and comparing the distance between them. We run 1000 simulations for the true parameters, 100 with parameters drawn from the priors for each of the distance metrics. All the comparisons are summarized in the density plot, Figure 4. We choose specific ϵ_T values for each distance metric based on these results to correspond to the peak of the density obtained for the true parameters. For example, we choose $\epsilon_T = 0.46$ for the Optimal Transport distance. In practice choice of the ϵ_T schedule is difficult and can affect convergence and speed of convergence considerably [44]; it also depends crucially on the choice of distance metric, and below we discuss a range of suitable distance metrics for single cell data and their relevance for inferring the parameters of models of cell differentiation.

3 Results

3.1 Distances for Single Cell Data

In Eqn. (3) we determine the quasi-potential $\tilde{U}(X)$ from the probability distribution over cell states, X . For our ABC SMC procedure we therefore need suitable distances between probability distributions. In this section we examine and compare some measures that quantify the discrepancies between probability distributions, as well as their underlying geometric properties before adopting them in our ABC-SMC inference scheme.

3.1.1 KL-Divergence

We start with the KL-Divergence which is one of the simplest and most commonly used measure in comparing probability distributions. For data $D \sim Q$ and model distribution P_θ , the KL-divergence is defined as,

$$d_{KL}(P_\theta, Q) = \int q(x) \log \left(\frac{q(x)}{p_\theta(x)} \right) \mu(dx) \quad (6)$$

Here q and p_θ are the corresponding probability density functions with respect to the same probability measure μ . The KL-divergence is closely related to maximum likelihood theory [51]:

$$\theta_{minKL} = \operatorname{argmax}_\theta \frac{1}{n} \sum_{i=1}^n \log P_\theta(x_i) = \operatorname{argmax}_\theta \log p_\theta(x) = \operatorname{argmax}_\theta p_\theta(x) = \theta_{MLE} \quad (7)$$

However, if there exist x_i such that $q(x_i) = 0$, the KL-Divergence diverges to infinity which can become problematic in numerical computations. Adding a small amount of random noise may help, but at the risk of adding errors. Because of the numerical issues we prefer other measures of the distances between probability distribution.

3.1.2 Bhattacharya Distance

The Bhattacharyya distance is used in signal selection and pattern recognition[41] and defined by,

$$B(P, Q) = -\log \left(\int \sqrt{q(x)p(x)} dx \right) = -\log(\rho) \quad (8)$$

for distributions P, Q with probability density functions $p(x)$ and $q(x)$; $\rho(P, Q) = \int \sqrt{p(x)q(x)} dx$ is also known as the Bhattacharyya coefficient. Unlike the KL divergence, the Bhattacharyya distance is symmetric and unbounded ($0 \leq B \leq \infty$).

An advantage of the Bhattacharya distance is that it has been proved that for any two sets of parameters θ_1, θ_2 if we have $B(\theta_1) > B(\theta_2)$ then there exist $\pi = (\pi_1, \pi_2)$ prior distributions that satisfy $P_e(\theta_1, \pi) < P_e(\theta_2, \pi)$ where $P_e(\theta, \pi)$ is the error probability with parameter set θ and prior probability π [26]. This property makes it particularly meaningful for parameter estimation in an ABC framework while considering the optimal Bayes error probabilities. Furthermore, this metric gives good results for Gaussian noise problems, which is useful for SDE problems modelled that use standard Brownian noise terms [41].

Even though this criterion works for most common distribution functions, there are two drawbacks in our setting,

1. Our model is implicit (i.e. the data points are generated by a set of equations describing the biophysical processes), but the corresponding distributions cannot be represented by a density function without using kernel density estimation (KDE) or a different density estimation methods in high-dimensional space. However, due to the curse of dimensionality, the number of data points required to have accurate estimates increases dramatically with increasing dimensionality[33], distance metrics which require KDE would be problematic not only in terms of accuracy but also efficiency.
2. This metric only measures the difference of probability density in a pointwise way which ignores the underlying geometry name of the probability space[21].

Therefore, we next introduce distance metrics that do not require KDE and which take the geometry of feature space into account.

3.1.3 Optimal Transport

Next we consider an optimal transport distance metric that is free of density estimation by comparing samples directly and maintains the underlying geometry in feature space. This metric has been widely used in the field of machine learning and image recognition for its efficiency [17]. Intuitively, the key to Optimal Transport is to figure out the minimal cost that is required to transform one image into another, or in our case, one probability distribution into another.

Suppose we have two probability distributions P, Q on the probability space χ, Γ with probability measure μ, ν respectively. Let $T_\# \mu$ denote the forward of μ by T . That is

$$T_\# \mu(B) = \mu(x : T(x) \in B) = \mu(T^{-1}(B)), B \in \sigma(\Gamma).$$

The original version of optimal transport is defined by [45]

$$\inf_{T_\# \mu = \nu} \int c(x, T(x)) \mu(dx) \quad (9)$$

where $c : \chi \times \Gamma \mapsto [0, \infty)$ is a measurable function on $\chi \times \Gamma$ known as the cost function.

However, this technique is clearly not applicable when there does not exist any forward equation T such that $T_\# \mu = \nu$. To overcome this, we adopt Kantorovich's version of Optimal Transport or the p -Wasserstein distance $W_p(P, Q)$, defined as [45],

$$W_p(P, Q) = \left(\inf_{\pi \in \Pi(P, Q)} \int c(x, y) d\pi(x, y) \right)^{1/p}, p \geq 1 \quad (10)$$

where $\Pi(P, Q)$ represents all the joint probability measures on $\chi \times \Gamma$ with marginal measures, μ and ν , respectively. This version is known to be well defined in general scenarios[45].

If in addition we have $f_P : \chi \mapsto \mathbb{R}$ and $f_Q : \Gamma \mapsto \mathbb{R}$ bounded and measurable functions with

$$f_P(x) - f_Q(y) \leq c(x, y) \quad (11)$$

it is easy formulate the duality form of Optimal Transport [1]

$$W_p(P, Q)^p = \sup_{f_P, f_Q} \left(\int f_P(x) \mu(dx) - \int f_Q(x) \nu(dy) \right). \quad (12)$$

Since we are working on the same probability space for both P and Q , say $\chi = \Gamma$, we can have the generalized form under 1-Lipschitz (Lip1) continuous functions on χ [8],

$$W_p(P, Q)^p = \sup_{f \in Lip1} \left(\int f(x) P(dx) - \int f(x) Q(dx) \right). \quad (13)$$

If we have two different probability distributions, say $P, Q \in P_\chi$, then the mixture distribution

$$P_t = (1-t)P + tQ, \forall t \in [0, 1] \quad (14)$$

can be considered as a curve connecting P, Q under the whole space of probability distributions, P_χ . It can be shown that for $p = 2$, the Wasserstein distance will give a minimal geodesic under such a probabilistic space of distributions[8]. In other words, whenever Optimal Transport proposes a probability distribution that transports from P to Q under χ , we will have the shortest path or distributions between P and Q . This result gives an extremely meaningful insight in our ABC-SMC framework to compare samples and simulations in geometry. It is worth noting that there may exist many minimal geodesics given by the Optimal transport which satisfy Equation (14).

3.1.4 Maximum Mean Discrepancy

We also consider the Maximum Mean Discrepancy (MMD) measure which has been used for of goodness of fit tests [27], and in an ABC framework [3]. The principle here is to find a kernel function that will give different expected values from two samples when they are not drawn from the same distribution. The underlying rationale is that the evaluations of such function at sample points, which are drawn from different probability distributions, can provide enough information about their difference[7].

The initial definition of MMD is as follows: Let \mathcal{F} be a class of functions $f : \chi \mapsto \mathbb{R}$ and P, Q be distributions of samples from X, Y which are independent and identically distributed with respect to P and Q ; then MMD can be defined as

$$MMD(\mathcal{F}, P, Q) = \sup_{f \in \mathcal{F}} (E_P[f(x)] - E_Q[f(y)]) \quad (15)$$

It is clear that if and only if X, Y are drawn from the same distribution (i.e. $P = Q$) then the MMD is zero for any functions $f \in \mathcal{F}$; but if the class of functions \mathcal{F} is too "large", the statistic would be much greater than 0 for most of the finite samples X, Y , which "strengthen" the differences in distribution.

In order to avoid this scenario while simultaneously allowing reasonable discrepancies between $P & Q$ to be detected, there must be a suitable restriction on

the class of function \mathcal{F} . The trade-off is achieved by restricting \mathcal{F} to be the unit ball in reproducing kernel Hilbert space (RKHS)[7]. The MMD can be calculated from this space by calculating the expected values of this kernel function under distributions P, Q . Let \mathcal{H} be a RKHS on feature space χ induced by a feature map $\phi : \chi \mapsto \mathbb{R}$. The kernel is given by the inner product between the feature maps under RKHS,

$$k(x, y) := \langle \phi(x), \phi(y) \rangle_{\mathcal{H}} \quad (16)$$

It is guaranteed that MMD can detect the difference between distributions P, Q of samples from X, Y . When \mathcal{F} restricted as the unit ball in the reproducing kernel Hilbert space (i.e. $\mathcal{F} := f : \chi \mapsto \mathbb{R} \mid \|f\|_{\mathcal{H}} \leq 1$), then $MMD(\mathcal{F}, P, Q) = 0$ if and only if $P = Q$ [7]. With characteristic kernels (i.e. kernel mapping is injective), MMD is proven to be a probabilistic metric which takes values in distribution functions [46]. A standard choice of this is the Gaussian kernel,

$$k(x, y) = e^{\frac{(x-y)^2}{2\sigma^2}}. \quad (17)$$

We will use the median heuristic to set the bandwidth of Gaussian kernels with $\sigma = \frac{\text{median}(X, Y)}{2}$.

In practice, given samples from X with sample size m , and from Y with sample size n , it is straightforward to determine the corresponding MMD,

$$\begin{aligned} MMD^2(\mathcal{F}, X, Y) &= \frac{1}{m(m-1)} \sum_{i \neq j}^m k(x_i, x_j) \\ &\quad + \frac{1}{n(n-1)} \sum_{i \neq j}^n k(y_i, y_j) - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m k(x_i, y_j) \end{aligned} \quad (18)$$

Since MMD can be easily computed by empirical observations, adopting it with ABC-SMC simulators would be very efficient.

MMD is defined similarly to Optimal Transport 13 under different sets of functions \mathcal{F} . Such metric problems are known as Integral Probability Metrics(IPMs)[32]. It has been shown that all such IPMs preserve the minimal geodesic property and it has further been proved that there exists a unique minimal geodesic under MMD in comparison to Optimal Transport (For details see Theorem 5.3 in [8])

3.2 ABC inference depends on distances

The obtained approximate posterior distribution of all the parameters for each distance metric are summarized in Figure 8,9,10. Our results of posterior distributions suggest a good agreement in estimating parameters using ABC-SMC with different metrics. With regard to the identifiability of parameters, we figure out that there is a well structured shape of probability distributions for most of the parameters ($c_1, c_2, c_3, e_1, b_1, b_2, b_3$) but other parameters are more likely to have “flat” distributions which are not well inferable.

Here we test the accuracy of the estimated posterior distributions by drawing particles from such posterior distributions and see if we can reproduce the dynamic trajectories of our system. The simulation result using drawn particles

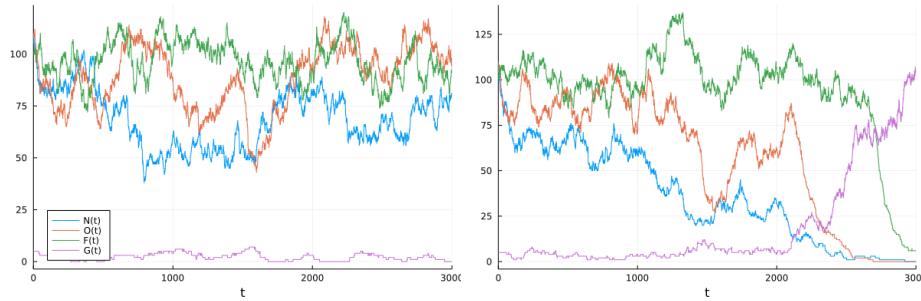


Figure 3: The sample trajectories of the four key transcription factors over time. The left one is Nanog stays pluripotent and the right one is Nanog differentiated into Gata6.

from ABC-SMC results with different distance metrics is given in Figure 5 and it is much akin the pattern of dynamics as in reference data (Figure 2b). Our result suggest both, the applicability of ABC-SMC in parameter estimations from single cell data, and the importance of choosing an adequate distance measure.

We choose the sequence of tolerance thresholds for MMD as the 30% percentile, which here gives rise to a schedule of,

$$\epsilon \in \{2.88, 0.79, 0.28, 0.15, 0.10, 0.008, 0.06\}. \quad (19)$$

which decreases slowly , see Figure 4. For comparison, the sequence of tolerance thresholds for Optimal Transport is

$$\epsilon \in \{4.61, 2.72, 1.15, 0.75, 0.59, 0.5, 0.46\} \quad (20)$$

which decreases constantly even after reaching the the boundary support of density for true parameters in Figure 4. Our result indicate that it in an ABC-SMC context it might be easier for Optimal Transport to take an efficient path to parameter inference than MMD. Theoretical reason for that would be worth studying in the future.

The ABC-SMC posterior distribution helps to analyze the relationships among all the parameters in our system. We can determine underlying correlations between pairs of parameters in the pair-wise plots (Figure 8). We find that c_1 and c_3 are strongly correlated with a clear linear trend and some slight linear relationships are also shown between e_1 and e_2 and b_1 and b_2 . With regards to the identifiability of relationship between b_1 and b_2 , we find that the performance of the Bhattacharya distance lags behind those of other distances (Figure 10).

Many such dynamical, including biological models, are know to exhibit “sloppy” behaviour: the dynamical behaviors of such systems are usually controlled by a small number of parameters [50, 19]. In addition to that, varying correlated parameters will still reproduce the same exhibiting behaviors of such systems. For example, in our result, it would be easy to determine the value of c_3 or c_1 if the other one is known while reproducing the same trajectories of transcription factors due to the correlation between them. We will show that our approach can help resolve this problem by identifying such set of correlated

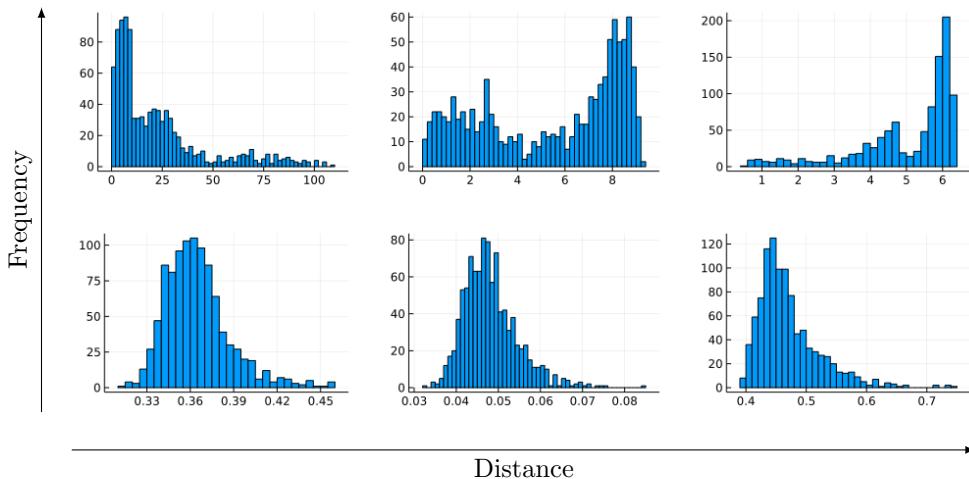


Figure 4: The distribution of distances of 1000 comparisons for different distance metrics. Top first: Bhattacharyya distances with parameters drawn from the priors. Top second: MMD with parameters drawn from the priors. Top third: Optimal Transport with parameters drawn from the priors. Bottom first: Bhattacharyya distances with identical parameters. Bottom second: MMD with identical parameters. Bottom third: Optimal Transport with identical parameters.

parameters from joint distribution and the detail of this “sloppy” analysis will be discussed in the following section.

With regard to computational cost, we found that MMD is the fastest one and Optimal Transport follows as reported in the theoretical time complexity[35]. And with Bhattacharya distance, it takes more than three times amount of time as in MMD where most of them are consumed in probability density estimation by KDE. Therefore, our results suggest the priority of density free methods in our ABC-SMC framework in single cell model.

3.3 Parameter Sensitivity

Visualizing and analyzing the sensitivity of parameters that are in high dimensions is typically hard and so does in our joint posterior distributions. To counteract this, the most simplest and convenient way is directly focusing on the variance of posterior distributions and so forth adopting Principle Component Analysis (PCA) to measure the variability of parameters [5, 42].

Starting with the posterior distributions of our parameters obtained from ABC-SMC, we can use the traditional Method of Moments to estimate the sample covariance matrix Σ :

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \quad (21)$$

And we can therefore estimate the variance of each parameters after applying

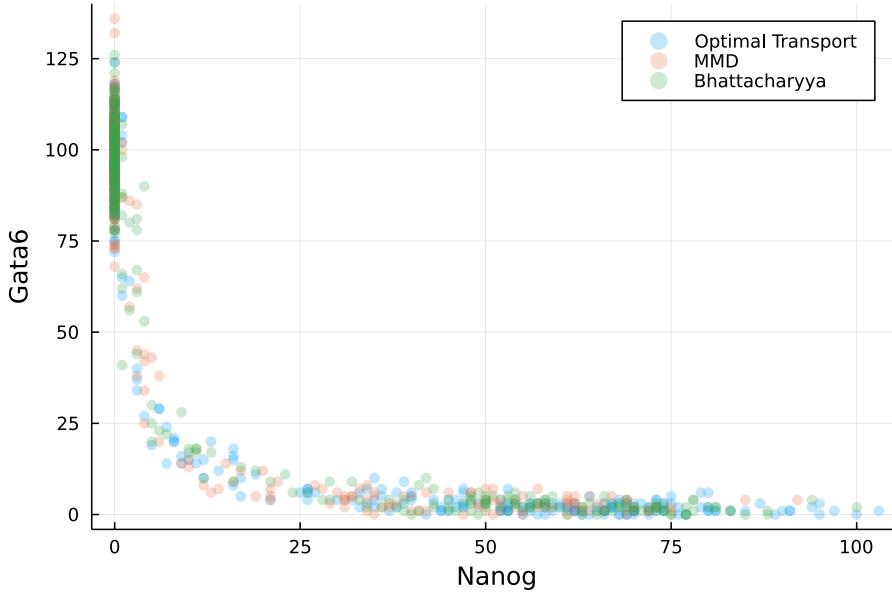


Figure 5: The plot of Nanog versus Gata6 at final time point using sampled particles from the last population; LIF was set as 50

the spectral decomposition to Σ :

$$Q^T \Sigma Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p) \quad (22)$$

where $Q = (q_1, \dots, q_p)$ is the eigen-parameter (vector). This eigenvector forms a orthogonal matrix and $\lambda_1, \lambda_2, \dots, \lambda_p$ are the corresponding variances of each parameter which are known as the principal components of Σ . Therefore we can easily measure the variability of data explained by each parameter by diagonalizing the covariance matrix and perform the sensitivity analysis.

If $q_{j,k}$ is the element of direction associated to the k^{th} parameters then we will have the projection of θ_k through the inner product,

$$c_{j,k} = q_{j,k} \cdot \theta_k \quad (23)$$

where $c = (c_1, \dots, c_p)$ are such principal components. We then have

$$\begin{aligned} \text{cov}(\theta_k, c_j) &= \text{cov}(q_{j,k} \cdot \theta_k, c_j) \\ &= q_{j,k} \lambda_k \end{aligned} \quad (24)$$

which gives the correlations between principal components c_j and eigen-parameters θ_k . Thus, we can use the standardized square correlations:

$$\rho_{\theta_k, c_j}^2 = \frac{q_{j,k}^2 \lambda_k}{\sum_{i=1}^p q_{i,k}^2 \lambda_i} \quad (25)$$

to interpret the proportion of variance eigen-parameter θ_k explained by principal component c_j . From this we can use the total proportion of variability explained

by parameter θ_k to quantify the sensitivity of parameter θ_k by,

$$s_k = \sum_{i=1}^p \rho_{\theta_k, c_i}^2 \quad (26)$$

We note that using PCA to analyze sensitivity is the same as eigenvalue analysis of the Hessian matrix around the MLE or the equation of least square errors [42].

We can define the log posterior density by

$$h(\theta) = \log(P(\theta|D)) = \log(P(D|\theta)) + \log(\pi(\theta)) \quad (27)$$

And using Taylor expansion of logarithm along with the log posterior distributions, we obtain [42],

$$h(\theta) = h(\hat{\theta})(\theta + \hat{\theta})^T h''(\hat{\theta})(\theta - \hat{\theta}) + O(\theta^3), \quad (28)$$

where $h''(\hat{\theta})$ is the Hessian matrix of log posterior distribution as in Equation ??.

When assuming the parameters are drawn from the multivariate normal distribution we can easily deduce the asymptotic covariance matrix of θ by the Fisher information matrix:

$$\text{Var}[\theta] = [-h''(\hat{\theta})]^{-1} \quad (29)$$

From this, we can conclude that the eigenvalues of the Variance-covariance matrix λ_k in Equation 22 and the eigenvalues of the corresponding Hessian matrix v_k , which results in asymptotic variance by assuming Multivariate Normality, are inversely related[28]:

$$\lambda_k = \frac{1}{v_k} \quad (30)$$

which legitimates the statistical property of our PCA result.

We summarize the result of variability/sensitivity of each parameter in Figure 6 following by the steps above. It is known that for change of value in sloppy parameters will alter the system behaviors slightly and change of that in stiff parameters will lead to dramatic change in dynamics. In terms of parameter variance, if we characterize parameters that explain more than 1% of the variance for the posteriors as sloppy and otherwise as stiff we can conclude that parameters $a1, b2, b3, c1, c2, e1, e2$ are the sloppy parameters with higher uncertainties in our system.

Our result also suggests that all the regeneration rates(c_0, e_0, a_0, b_0) as well as the constant degradation rate γ are more stiff than others. This is not quite surprising since these parameters actively decide the extent of change in each factor throughout the whole time. For gene transcription factors Oct4-Sox2 and Fgf4, all other parameters are sloppy except their corresponding regeneration rates and the constant degradation rate γ . This result has an agreement with that higher variance in Oct4-Sox2 and Fgf4 along with the time are expected in contrast to the others as shown in [10]. In view of the scaling parameters k_0, c_4 for Oct4-Sox2 and Gata6 which are stiff in factor Nanog, we found that the effect of Oct4-Sox2 and Gata6 on Nanog and hence on the fate decision making

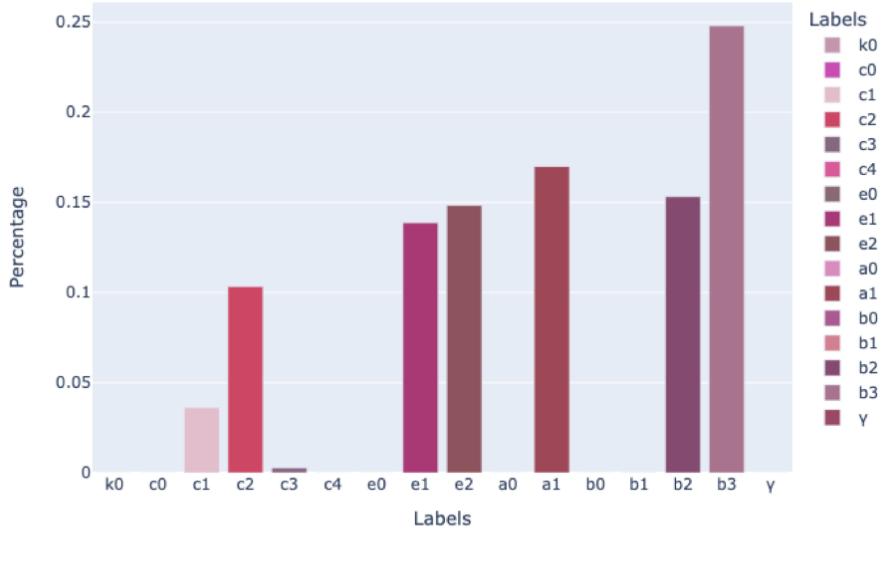


Figure 6: Variability of the model explained by different parameters

in our system is very distinct. In contrast, the effect of Nanog on Gata6 is less based on the sloppy scaling parameter b_2 . As a compensate, self-regeneration or degradation rate b_1 is much more significant on Gata6 which justifies its role as bio-marker of differentiation.

4 Discussion

The inference of parameters from experimental single cell data for complex systems is of significant practical interest. While there are many previous approaches based upon different Bayesian framework or ABC algorithms[38, 13, 6, 38, 11, 28], summary statistics coupled with some simple distance metrics like the Euclidean distance of L^p norm are commonly chosen[18]. However, recent studies, inspired by Waddington’s landscape, has revealed that the existence of potential provides a meaningful description of the system;s dynamics and is quantitatively associated with the probability distribution over states X_i [53, 4, 9]. Therefore, we argue that employing such basic distance metrics will result in information loss due to non-sufficient summary statistics.

In this study, we investigate parameter inference through the ABC framework using different distance metrics for a specific biological system. We explore the effectiveness of three metrics: Maximum Mean Discrepancy (MMD), Bhattacharya distance, and Optimal Transport. Our results have shown that the inference of parameters based on these metrics can successfully capture the dynamics of this biological system.

To see the significance of the quasi-potential landscape as a significant factor in determining model parameters within the ABC paradigm, we conducted a comparison by employing our ABC-SMC algorithm with the simple Euclidean distance metric. Specifically, we considered two types of summary statistics: the first and second moments ($E(X), E(X^2)$) and the logarithm-transformed probability of cells in pluripotent states ($\log(\frac{p}{1-p})$), where p was determined using K-means with 2 classes.

We summarized the results of the posterior predictive distribution check using $S(X) = \log(\frac{p}{1-p})$ as the test statistic (Figure 7). We outlined the extreme quantity p ($Pr(S(X) > S(D))$) and Mean Squared value (MSE) and observed that the posterior distributions obtained using distance metrics that compare probability distributions were more consistent with the simulated experimental data, outperforming those obtained with the Euclidean distance. Therefore, we concluded that accessing information from the landscape is crucial for determining the parameters of dynamical systems. To achieve this, it is necessary to adopt distance metrics capable of quantifying the differences between intrinsic probability distributions. The result of this comparison highlights the importance of considering the landscape structure and employing suitable distance metrics for accurate parameter inference within the ABC framework.

Our findings suggest that among the evaluated metrics for comparing distributions, MMD yields the highest accuracy. It is known that Optimal Transport and MMD belong to the same category of metrics known as Integral Probability Metrics (IPMs) [32]. Moreover, when they both use the same kernel functions, Optimal Transport can be considered as an interpolation of MMD [35]. Common knowledge suggest that Optimal Transport would perform better in determining the target estimate, especially when employing gradient flows such as in the gradient descent approach [21]. However, our results show a different outcome, with MMD exhibiting better performance in our framework. Therefore, our result suggest that MMD is more advantageous for applications in non-gradient approaches, like in the ABC (Approximate Bayesian Computation) scheme.

With regard to computational cost, both MMD and Optimal Transport metrics are quite similar, and they outperform Bhattacharya distance significantly. The key reason for the computational difference with Bhattacharya distance lies in the estimation of the probability density function through the KDE method. This drawback would become more significant as the dimensionality increases.

Considering the computational difference, we strongly recommend using MMD as the distance metric where samples are directly compared without density estimation and implicitly quantifying the differences between distributions. This property makes MMD computationally efficient and well-suited for dealing with high-dimensional data and it offers a more practical and efficient option for parameter inference with ABC algorithm in dynamical systems.

Although the potential landscape offers valuable insights into the dynamics of biological systems, it is not the sole factor governing these processes. Conventionally, the quasi-potential is defined by taking the negative logarithm of probability distributions, following Boltzmann's rule, assuming a static and time-homogeneous landscape [34]. However, recent research has delved deeper into this field, revealing that cell systems are more like time-inhomogeneous, and highlighting the existence of transient landscapes [20, 9].

To accurately capture the time evolution of stem cells, it is essential to con-

sider the curl of probability flux, which interacts with the potential landscape to govern the ongoing trajectories of cell fates [29]. In our approach, although we attempted to quantify the variational structure of the landscape by collecting data points at different times, we still lost the information related to the probability flux.

Since we have demonstrated the significance of landscape information in inferring parameters from single-cell data, we believe that our framework can be further enhanced and better understood by incorporating and measuring the probability flux under appropriate distance metrics. This extension would allow us to gain a more comprehensive understanding of the underlying dynamics and improve the accuracy of parameter inference in complex systems.

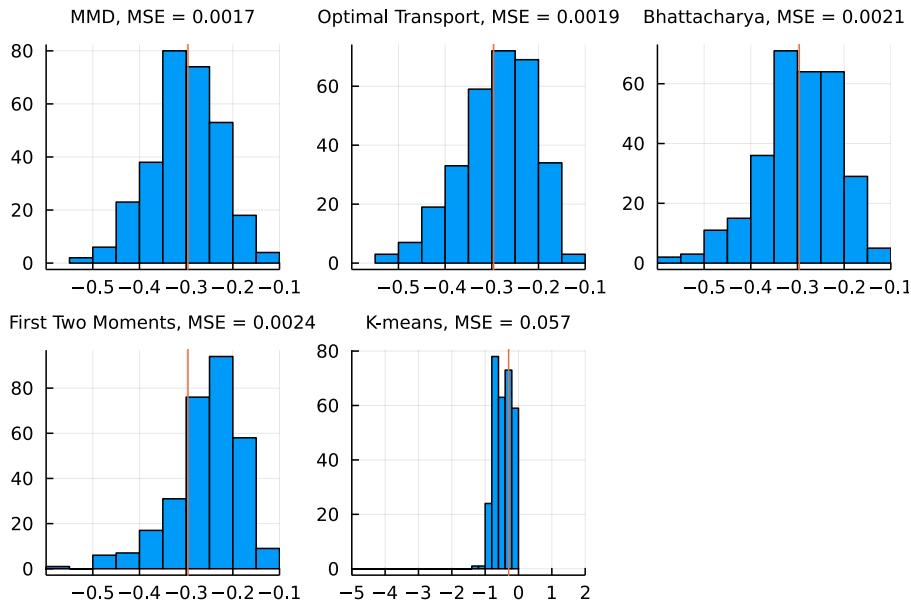


Figure 7: Realized vs. posterior predictive distributions for ABC-SMC results obtained with the use of different distance metrics: $S(X) = \log(\frac{p}{1-p})$ (where p is the probability stem cell ends in pluripotent state), compared to simulations from the posterior predictive distribution for each distance metric. The vertical line presents the test statistics from simulation of true parameters and the corresponding Mean Squared Error (MSE) for different metrics are presented.

Acknowledgment

YL and MPHS gratefully acknowledge funding from the ARC through an ARC Laureate Fellowship to MPHS (FL220100005); SYZ acknowledges funding from the Australian Government Research Training Program and an Elizabeth and Vernon Puzey Scholarship. ITK was funded through a Wellcome Trust PhD studentship.

References

- [1] Michael Arbel, Anna Korba, Adil Salim, and Arthur Gretton. Maximum mean discrepancy gradient flow, 2019.
- [2] Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate bayesian computation in population genetics. *Genetics*, 162:2025–2035, 12 2002.
- [3] Espen Bernton, Pierre Jacob, Mathieu Gerber, and Christian P Robert. Approximate bayesian computation with the wasserstein distance. *Journal of The Royal Statistical Society Series B-statistical Methodology*, 81:235–269, 04 2019.
- [4] Sudin Bhattacharya, Qiang Zhang, and Melvin E Andersen. A deterministic map of waddington’s epigenetic landscape for cell fate specification. *BMC Systems Biology*, 5:85, 2011.
- [5] Christopher M. Bishop. Bayesian pca. In *Proceedings of the 11th International Conference on Neural Information Processing Systems*, NIPS’98, pages 382–388, Cambridge, MA, USA, 1998. MIT Press.
- [6] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112:859–877, 02 2017.
- [7] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Scholkopf, and A. J. Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006.
- [8] Leon Bottou, Martin Arjovsky, David Lopez-Paz, and Maxime Oquab. Geometrical insights for implicit generative modeling, 2017.
- [9] Rowan D Brackston, Eszter Lakatos, and Michael . Transition state characteristics during cell differentiation. *PLOS Computational Biology*, 14:e1006405–e1006405, 09 2018.
- [10] Rowan D. Brackston, Eszter Lakatos, and Michael P. H. Stumpf. Transition state characteristics during cell differentiation. *PLOS Computational Biology*, 14(9):e1006405, 2018.
- [11] Joshua Burton, Cerys S. Manning, Magnus Rattray, Nancy Papalopulu, and Jochen Kursawe. Inferring kinetic parameters of oscillatory gene regulation from single cell time-series data. *Journal of The Royal Society Interface*, 18, 09 2021.
- [12] Elena Camacho-Aguilar, Aryeh Warmflash, and David A. Rand. Quantifying cell transitions in *C. elegans* with data-fitted landscape models. *PLOS Computational Biology*, 17(6):e1009034, 2021.
- [13] Kieran R. Campbell, Adi Steif, Emma Laks, Hans Zahn, Daniel Lai, Andrew McPherson, Hossein Farahani, Farhia Kabeer, Ciara O’Flanagan, Justina Biele, Jazmine Brimhall, Beixi Wang, Pascale Walters, IMAXT Consortium, Alexandre Bouchard-Côté, Samuel Aparicio, and Sohrab P.

- Shah. clonealign: statistical integration of independent single-cell rna and dna sequencing data from human cancers. *Genome Biology*, 20, 03 2019.
- [14] Thalia E Chan, Michael P H Stumpf, and Ann C Babtie. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Systems*, 5(3):251–267.e3, 2017.
 - [15] Vijay Chickarmane, Victor Olariu, and Carsten Peterson. Probing the role of stochasticity in a model of the embryonic stem cell – heterogeneous gene expression and reprogramming efficiency. *BMC Systems Biology*, 6(1), 2012.
 - [16] Megan A. Coomer, Lucy Ham, and Michael P.H. Stumpf. Noise distorts the epigenetic landscape and shapes cell-fate decisions. *Cell Systems*, 10 2021.
 - [17] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation distances. 2013.
 - [18] Christopher C. Drovandi, Anthony N. Pettitt, and Malcolm J. Faddy. Approximate bayesian computation using indirect inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(3):317–337, 2011.
 - [19] Kamil Erguler, Michael P H Stumpf, and Michael P H Stumpf. Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Molecular Biosystems*, 7(5):1593–1602, 2011.
 - [20] Xiaona Fang and Jin Wang. Nonequilibrium thermodynamics in cell biology: Extending equilibrium formalism to cover living systems. *Annual Review of Biophysics*, 49:227–246, 05 2020.
 - [21] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences, 2018.
 - [22] J.M. Hanson, J.A. Mol, and B.P. Meij. Expression of leukemia inhibitory factor and leukemia inhibitory factor receptor in the canine pituitary gland and corticotrope adenomas. *Domestic Animal Endocrinology*, 38:260–271, 05 2010.
 - [23] Kondepudi D. Horsthemke, W., editor. *Noise-Induced Transitions in Physics, Chemistry, and Biology*, pages 164–200. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.
 - [24] Sui Huang. The molecular and mathematical basis of waddington’s epigenetic landscape: A framework for post-darwinian biology? *BioEssays*, 34(2):149–157, 2011.
 - [25] Sui Huang, Yan-Ping Guo, Gillian May, and Tariq Enver. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, 305:695–713, 05 2007.

- [26] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communications*, 15(1):52–60, 1967.
- [27] Oscar Key, Tamara Fernandez, Arthur Gretton, and François-Xavier Briol. Composite goodness-of-fit tests with kernels, 2021.
- [28] Oleg Lenive, Paul D. W. Kirk, and Michael P. H. Stumpf. Inferring extrinsic noise from single-cell gene expression data using approximate bayesian computation. *BMC Systems Biology*, 10, 08 2016.
- [29] Qiong Liu and Jin Wang. Quantifying the flux as the driving force for nonequilibrium dynamics and thermodynamics in non-michaelis–menten enzyme kinetics. *Proceedings of the National Academy of Sciences*, 117:923–930, 12 2019.
- [30] Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R. Bialas, Nolan Kamitaki, Emily M Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.
- [31] Naomi Moris, Cristina Pina, and Alfonso Martinez Arias. Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*, 17:693–703, 11 2016.
- [32] Alfred Muller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443, 1997.
- [33] Thomas Nagler and Claudia Czado. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. 2015.
- [34] Hong Qian. Open-system nonequilibrium steady state: Statistical thermodynamics, fluctuations, and chemical oscillations. *The Journal of Physical Chemistry B*, 110:15063–15074, 07 2006.
- [35] Aaditya Ramdas, Nicolas Garcia, and Marco Cuturi. On wasserstein two sample testing and related families of nonparametric tests, 2015.
- [36] David G Rand, Archishman Raju, Meritxell Sáez, Francis Corson, and Eric D Siggia. Geometry of gene regulatory dynamics. 118, 09 2021.
- [37] C REDIT, S CHA, and N AI. Single-cell proteomics: challenges and prospects. *nature methods*, 20:317–318, 2023.
- [38] Robert J. H. Ross, R. E. Baker, Andrew Parker, M. J. Ford, R. L. Mort, and C. A. Yates. Using approximate bayesian computation to quantify cell–cell adhesion parameters in a cell migratory process. *npj Systems Biology and Applications*, 3, 03 2017.
- [39] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A comparison of single-cell trajectory inference methods. *Nature biotechnology*, 37(5):547–554, 2019.

- [40] M. Sáez, R. Blassberg, E. Camacho-Aguilar, E. D. Siggia, D. Rand, and J. Briscoe. A quantitative landscape of cell fate transitions identifies principles of cellular decision-making. *bioRxiv*, 2021.
- [41] Fred C. Schweppe. State space evaluation of the bhattacharyya distance between two gaussian processes. *Information and Control*, 11:352–372, 09 1967.
- [42] Maria Secrier, Tina Toni, and Michael P. H. Stumpf. The abc of reverse engineering biological signalling systems. *Molecular BioSystems*, 5(12):1925, 2009.
- [43] Madeline A. Shea and Gary K. Ackers. The or control system of bacteriophage lambda. *Journal of Molecular Biology*, 181:211–230, 01 1985.
- [44] Daniel Silk, S Filippi, and Michael P H Stumpf. Optimizing threshold-schedules for sequential approximate bayesian computation: applications to molecular systems. *Statistical applications in genetics and molecular biology*, 12(5):603–618, 2013.
- [45] Michael Snow and Jan van Lent. The monge-kantorovich optimal transport distance for image comparison, 2018.
- [46] Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R. G. Lanckriet. Hilbert space embeddings and metrics on probability measures, 2009.
- [47] Gautier Stoll, Eric Viara, Emmanuel Barillot, and Laurence Calzone. Continuous time boolean modeling for biological signaling: application of gillespie algorithm. *BMC Systems Biology*, 6:116, 2012.
- [48] Patrick S Stumpf, Rosanna C G Smith, Michael Lenz, Andreas Schuppert, Franz-Josef Müller, Ann C Babtie, Thalia E Chan, Michael P H Stumpf, Colin P Please, Sam D Howison, Fumio Arai, and Ben D Macarthur. Stem cell differentiation as a non-markov stochastic process. *Cell Systems*, 5(3):268–282.e7, 2017.
- [49] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P.H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, 2008.
- [50] Mark K. Transtrum, Benjamin B. Machta, Kevin S. Brown, Bryan C. Daniels, Christopher R. Myers, and James P. Sethna. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*, 143:010901, 07 2015.
- [51] Wim van Dam, Peter Grunwald, and Richard Gill. The statistical strength of nonlocality proofs. 2003.
- [52] Sean T Vittadello and Michael P H Stumpf. Open problems in mathematical biology. *Math Biosci*, 354:108926, Dec 2022.

- [53] Jin Wang, Li Xu, Erkang Wang, and Sui Huang. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophysical Journal*, 99:29–39, 07 2010.
- [54] Qi-Long Ying, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. The ground state of embryonic stem cell self-renewal. *Nature*, 453:519–23, 2008.

5 Appendix

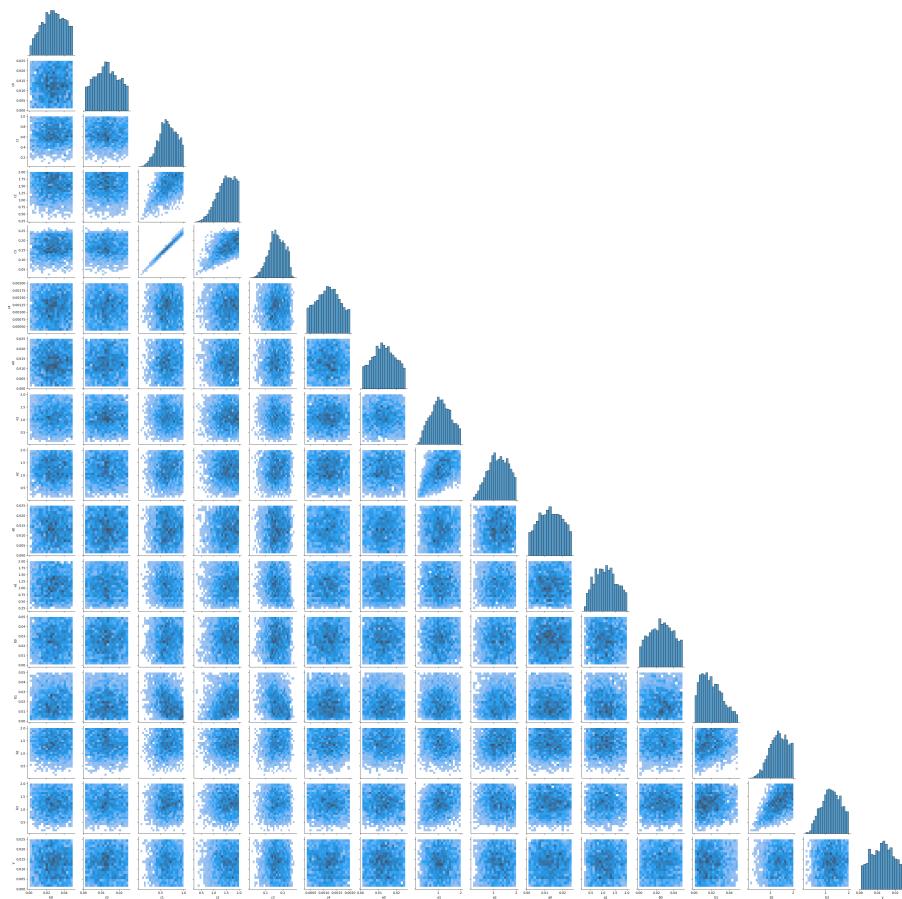


Figure 8: Results of ABC-SMC in parameter estimations for the stem cell model using Optimal Transport. Diagonal: Histogram of each parameter. Lower triangle: 2D pairwise scatter plot of the joint posterior distributions between two parameters.

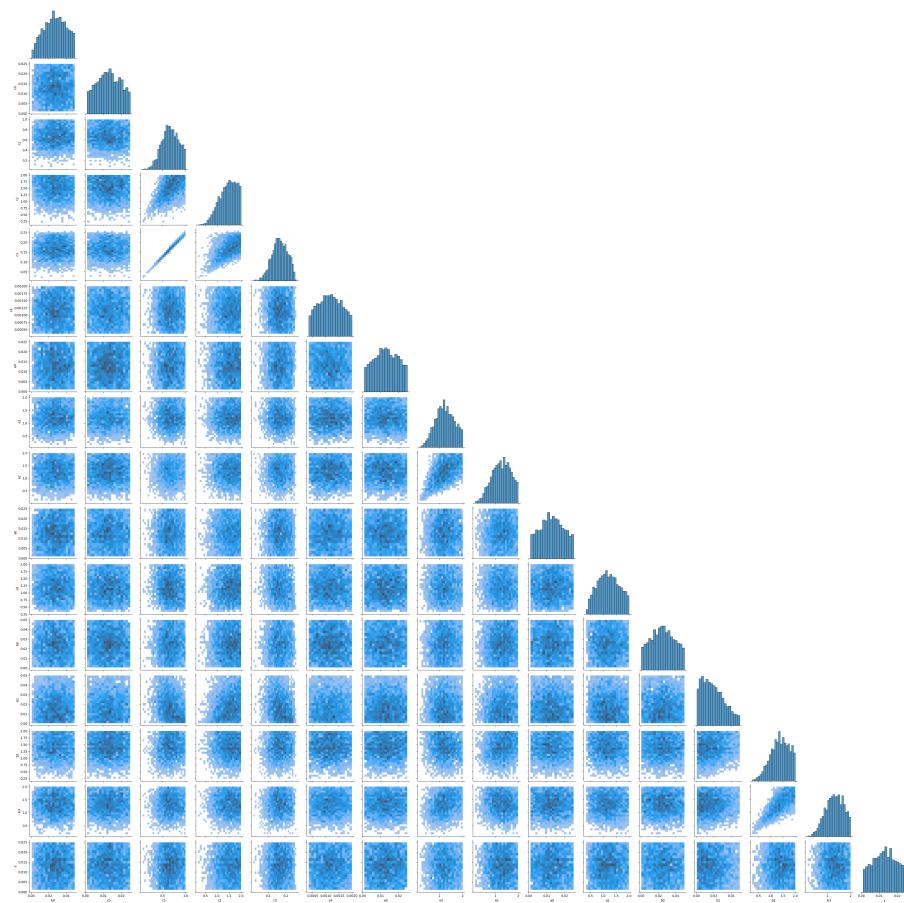


Figure 9: Results of ABC-SMC in parameter estimations for the stem cell model using MMD, similar as Figure 8

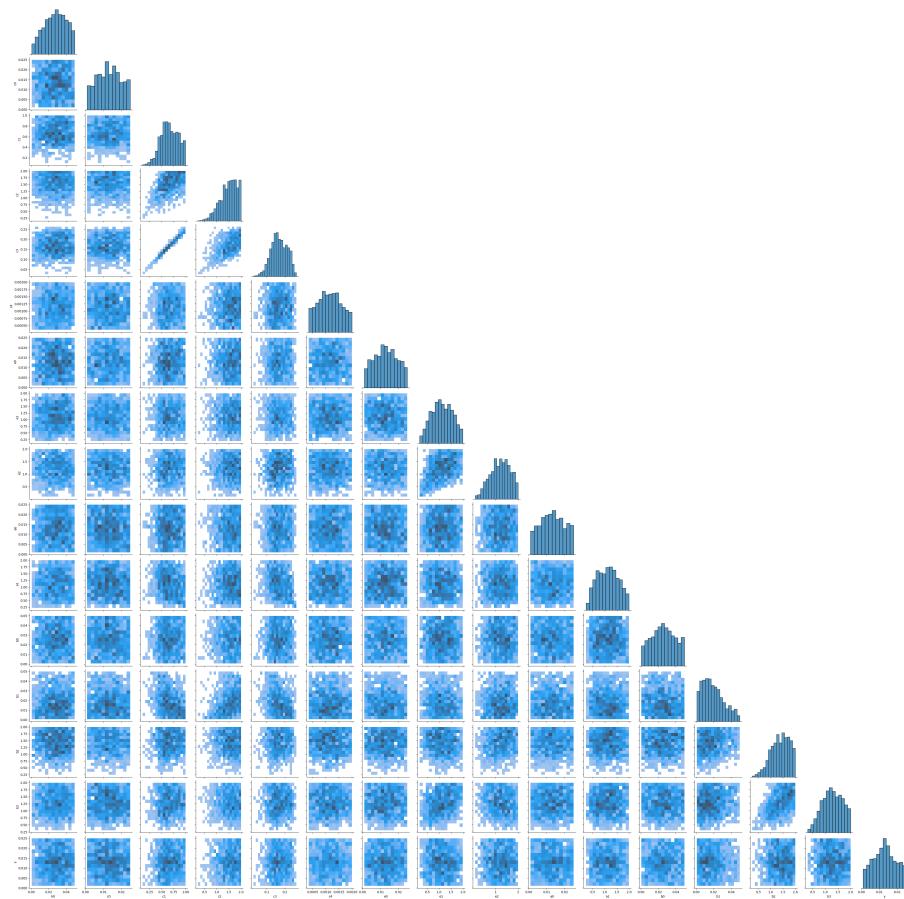


Figure 10: Results of ABC-SMC in parameter estimations for the stem cell model using Bhattacharya distance, similar as Figure 8