



Bayesian Inference Using Sequential Monte-Carlo Algorithm for Dynamic System Models

Chaolin Han

August 16, 2020

MSc in High Performance Computing

The University of Edinburgh

Year of Presentation: 2020

Abstract

This is the bit where you summarise what is in your thesis.

Contents

1	Introduction	1
2	Background	3
2.1	Zebrafish spinal cord regeneration	3
2.2	Mathematical modelling	4
2.3	Bayesian inference	6
2.4	Software tools	9
3	Mathematical modelling	10
3.1	Observed data	10
3.2	Hypothesis and models	11
3.2.1	Basic model	11
3.2.2	Alternative models	12
3.2.3	Limitations	15
4	Parameter estimation and model comparison	16
4.1	Code and implementations	16
4.2	Hyperparameters and implementation options experiments	17
4.2.1	Perturbation kernels	19
4.2.2	Adaptive functions and factors	22
4.2.3	Data size, prior distribution and population	25
4.3	Parameter estimation and model comparison	27
4.3.1	Model 1, 2 and 3	27
4.3.2	Model 4 and 5	31
4.3.3	Sensitivity of parameters	36
5	Performance experiments	39
5.1	Scaling-up	39
5.2	Discussions	41
6	Conclusion and future works	46
A	System and software environment	48
A.1	ABC SMC implementation	48
A.1.1	Local machine	48

A.1.2	Remote machine	49
B	Data and supplementary figures	50
B.1	Infer-back experiments	50
B.1.1	Parameter values used to generate synthetic data	50
B.1.2	Kernel experiment: median epsilon schedule	51
B.2	Parameter inference and model comparison results	52

List of Tables

3.1	Parameters introduced in the basic model (model 1) [REMOVE unit]	13
3.2	Parameters introduced in model 4 and 5	15
4.1	Estimated parameter values of model 3	28
4.2	Inferred parameter values of model 5	35
5.1	Performance data	41
A.1	Environment on local machine	48
A.2	Site packages on local machine	48
A.3	Hardwares on local machine	49
A.4	Environment on local machine	49
B.1	Parameter values used to generate synthetic data for model 1	50

List of Figures

2.1	Working model of the influence of the innate immune system on axonal regrowth	4
2.2	ABC SMC sampling process	8
3.1	Mean of the observed data	11
3.2	Interactions modelled in the basic model	12
4.1	Synthetic data generated with known parameter values	18
4.2	Total required samples of different kernels	21
4.3	Acceptance rates of different kernels	21
4.4	Factors and adaptive distance experiments	24
4.5	Simulated trajectories	24
4.6	Data size and prior distribution range experiment	25
4.7	Simulated trajectory from the 20th population of model 1, 2 and 3	29
4.8	Epsilon trends and acceptance rates of model 1, 2 and 3	29
4.9	Estimated posterior distribution of parameters in model 3	30
4.10	ABC SMC based model comparison of model 1, 2 and 3	30
4.11	Simulated data from the last population of model 3, 4 and 5. Error bar indicates SEM (for N and Φ the SEM is relatively small)	32
4.12	Simulated trajectory from the 30th population of model 5, with 25:75 factor applied	33
4.13	Model probabilities of model 3, 4 and 5 in ABC SMC based model comparison	34
4.14	Inferred posterior distribution of parameters in model 5	35
4.15	Explained total variance in original parameters by each PC	36
4.16	TODO	37
5.1	Speedup of the program	40
5.2	Average samples in a second, under different number of cores	41
5.3	Local optimum and abnormal pattern	44
5.4	Parameter density distribution before and after the concentration movement in generation 11	45
B.1	Total sampling size of different kernels, using median epsilon strategy . .	51
B.2	Acceptance rates of different kernels, using median epsilon strategy . .	51

B.3	Simulated data from the last population of model 1, 2 and 3, using uniform prior. Error bar indicates SEM (for N and Φ the SEM is relatively small)	52
B.4	Joint density distribution of each parameter pair	53
B.5	Credible intervals of parameters in model 5	54

Chapter 1

Introduction

Zebrafish can regenerate its spinal cord after injury; tissue regeneration observed at the injury site is a complex biological process with interacting parts and dynamical feedback between different types of cells and molecules. These cells and molecules are involved in the inflammation and repair dynamically, and a complex interacting network can be drawn according to it.

Mathematical models and techniques can help us to resolve the complexity in this dynamic system; however, how to derive the parameter estimates of the models remains a challenging work. In this dissertation, our interest lies in the modelling of the regeneration process that happens around the lesion site, especially in the parameter estimates and comparison of possible models. According to available researches [1], a mathematical model was first proposed, but the addressed interactions remain to be precisely defined, and quantitatively examined in experiments. We want to find the best parameter sets of the model using the published data, and ranking models under specific metrics and possibly find the best-fit model. In these steps, models are expected to retain mathematical and biological significance regarding the complex interactions, and differential equations are suitable for this.

For the parameter estimation tasks, optimisation-based methods usually have trouble in determining global optima from local optimal, and noise in the observed data may lead to over-fitting. Compared to optimisation-based methods, Bayesian inference techniques can help to relieve these two concerns[2]. Moreover, approximate inference methods (Approximate Bayesian Computation, ABC) make likelihood-free inference possible, which is suitable for our task as writing down likelihood as a function expression for the dynamic system can be hard.

Such methods introduce computation-intensive sampling-and-test patterns, where a massive number of samples are drawn and followed by complex computations. As the sampling is independent of each other in a certain period, the parallel implementation of the algorithm becomes feasible and is expected to accelerate the parameter inference on a large scale when using HPC resources.

This dissertation mainly focuses on the implementations and experiments of parameter estimation using ABC SMC (sequential Monte-Carlo) method, and discussions on the results. Additionally, performance experiments illustrated the strong-scaling performance of the algorithm. Chapter 2 talks about the background information and theories, Chapter 3 shows how the mathematical models are derived from hypothesis, and Chapter 4 describes the ABC SMC implementations and experiments on the models in detail. A brief scaling-up performance experiment is included in Chapter 5, and Chapter 6 summarises the conclusions and states the future works on this topic.

Chapter 2

Background

2.1 Zebrafish spinal cord regeneration

Compared to mammals, zebrafish can regenerate the spinal cord after injury, and possibly recover swimming function. They are widely used in regenerative ability researches due to this ability. Experiments discovered that although being cut off, the axonal of nerve cells can regenerate and bridge across the lesion site, which critically determined whether zebrafish can regain the swimming function [3]. Tsarouchas et al. [1] studied the complex interactions that happen at the lesion site. The inflammation affects the regeneration: if it is promoted, it accelerates the axonal regeneration; if it is inhibited, it reduces the regeneration[1]. Further experiments showed that two kinds of cytokines, $il-1\beta$ and $tnf-\alpha$, are related the inflammation, and they are dynamically controlled by immune cells, i.e. immune cells produce and control the cytokines-mediated inflammation [1]. Cytokines are small proteins, work as messengers to pass signal across different cells, which are similar to hormones and neurotransmitter.

An interaction map (Figure 2.1) can be drawn according to the evidences found in experiments.[1]. After the injury, a massive immune response was observed, with the increasing numbers of several kinds of cells. This dynamical system consists of complex interactions and dynamical feedback between cells and cytokines. For example, $il-1\beta$ promotes macrophages' increase. As a primary source, more macrophages leads to more expression of $tnf-\alpha$ mRNA.

Consequently, $tnf-\alpha$ and $il-1\beta$ plays determinative roles in the axonal regeneration. $il-1\beta$ promotes the regeneration at early stages, but later has an inhibition effect; $tnf-\alpha$ promotes the regeneration all the time. The presence of the cytokines also affects the number of immune cells, while as sources, more immune cells produce more cytokines. Interactions between different types of cells are also considered in the system (e.g., the number of neutrophils is affected by macrophages).

This dynamic system is suitable for mathematical modelling, where a set of equations can be written to represent the dynamics of different objects, e.g. each type of cells and

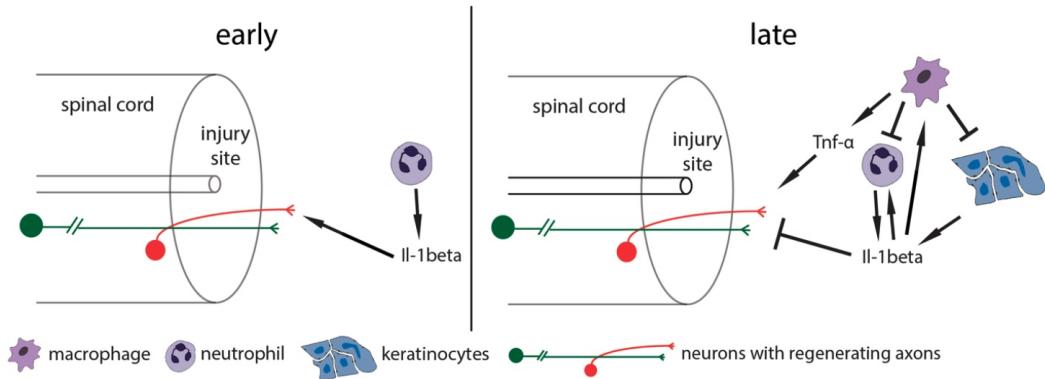


Figure 2.1: Working model of the influence of the innate immune system on axonal regrowth, from Tsarouchas et al. [1]. Lines ended with arrows represent promoting effect, lines ended with T-connectors represent inhibition

cytokines. Our interest lies in the modelling of the regeneration period of the dynamic system, and we wish to find that if we can quantitatively study the interactions inside the system, based on the proposed hypothesis [1], furthermore based our hypothesis which may include different interactions.

The biological background of works in this dissertation is mostly related to the experimental evidence, quantitative data and proposed hypothesis in [1]. The published data on the quantitative measurement of cells and cytokines make the modelling and further analysis possible. Some other interactions proposed in our modelling stages are enlightened from Anderson et al. [4], where intracellular cytokine signalling networks were established and computationally modelled.

2.2 Mathematical modelling

Mathematical models can express real-life problems in mathematical forms, and provide qualitative or quantitative inside-look of the question through various mathematical tools and techniques. They can be used for prediction, classification, testing our understanding or hypothesis and many other tasks; thus, they are widely used to solve practical problems. However, for a model that describes the zebrafish spinal cord regeneration process, we were less interested in the model's predictive functionality; the main focus is on the exploration of uncertainties in parameter estimates and comparison with alternative models. In this case, we wish to use modelling techniques and implement inference methods, test our hypothesis and hopefully find and compare well-fitted models.

According to our understanding of the biological mechanism in the regeneration, hypotheses should be proposed in a way that can represent real mechanism as much as possible. Usually, a real-life problem, e.g. the spinal cord regeneration, consists of

complex dynamics and interactions with many other environmental factors to be considered. An elaborate description of the real cases using an overly complicated mathematical model sometimes can be impossible; assumptions and hypotheses are often proposed to make necessary abstract and simplification for a final reasonably detailed model. Some qualitative models are illustrative but straightforward for the observed patterns, e.g. Lotka-Volterra (L-V) equation for predator and prey models. They are usually analytically solvable, and their properties are used to make predictions (e.g. steady states in the ecosystem). If a more precise model is desired, usually more influences or interactions are considered and added as terms of equations to a re-constructed model, e.g. considering environmental factors, a changing environmental capacity and predation from other species in L-V model.

Models like the L-V equation are mostly analytically solvable when analysing properties of the model. However, analytical solutions for many other general models are not feasible because of more complicated dynamics, or the model itself is not a deterministic model. In the areas where solutions are not analytically feasible, computational techniques are usually applied for approximate solutions. Solvers for different models had been proposed to simulate the system or find roots for some special points, such as Explicit Runge-Kutta [5] and LSODA.

For the regeneration process that we wanted to model, differential equations are our preference. Non-linear ordinary differential equations are used in this dissertation, where the only independent variable is time t , denoting the post-lesion time (hours). Further models can also consider other models in other forms, e.g. stochastic models and partial differential equations. For ordinary differential equations (ODE) system, there are several available computational solvers using different algorithms, implemented in MATLAB¹, Python² and many other languages. These solvers help to simulate the trajectories of the dependent variables, given initial values and a set of fixed parameters. The decency of fit can be measured by compared the simulated data and observed real data.

Often further researches on the modelled problems require the model to be deterministic on its parameters. Parameters in dynamic system models usually have their practical significance, e.g. describing reaction speed, environmental capacity or the birth rate. There are many approaches to derive these parameters from data or experience. For a popular topic and widely used models, parameters' values can be read from previous literature. However, in many other cases, we need to find values explicitly for our model, where numerical values for parameters are mostly preferred as they are suitable for practical analysis and comparison.

To find the best parameters' values given observed data, either optimisation-based approaches (e.g. least-squares fit (LS), maximal likelihood estimation (MLE)) or inference-based approaches can be applied. Optimisation-based approaches may face two major

¹<https://uk.mathworks.com/help/matlab/math/choose-an-ode-solver.html>

²https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html

drawbacks: over-fitting and local optima [2]. Bayesian inference approaches gain popularity in recent years, as it can relieve the concerns in over-fitting and provide a measurement for the uncertainty in the inferred values. Local optima can still be a concern, and many efforts have been proposed for a more thorough exploration of the parameter space to avoid local modes.

2.3 Bayesian inference

Bayesian inference approaches for parameter estimation put Bayes Rule in the centre:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \quad (2.1)$$

where p is the probability density function, θ denotes the parameter set, D is the given observed data.

Alternatively, as $p(D)$ is fixed once the observed data is provided, the rule can be rewritten as

$$p(\theta|D) \propto l(D|\theta)p(\theta) \quad (2.2)$$

where l represents the likelihood function. In other words, Eqn. 2.2 can be described as ‘posterior is proportional to the product of likelihood and prior’, and here posterior of parameters is our target in the parameter estimation task. Prior $p(\theta)$ should express our initial believes or estimates on the parameter values before observing the data and conducting experiments; likelihood function $l(D|\theta)$ is used to describe how well the parameter set θ explains the data D , in probability values. Once the data is given, θ is the only parameter in the likelihood function. Some optimisation-based methods attempt to find estimates of parameter values θ s.t. maximise the likelihood. However, such methods will struggle when facing a problem where it is hard to specify an analytical expression for likelihood function, let alone calculating and comparing values of it.

Based on Eqn. 2.2, to find the posterior of the parameter set, prior can be set according to the problem context and our understanding and cover a reasonable range. Computational methods can be used to approximate likelihood function, e.g. Markov Chain Monte Carlo (MCMC) [6]. As an alternative solution, Approximate Bayesian computation (ABC) approaches do not focus on the evaluation of likelihood function but focus on simulating the data from the given model and observed data, then use the simulated data to approximate the true posterior. ABC-rejection [7] is a simple example of the ABC process, and further algorithms e.g. sequential Monte-Carlo (ABC SMC)[8] were proposed to improve the efficiency.

General steps for ABC algorithm includes the following (ABC-rejection):

1. Draw a sample θ^* from parameters’ prior distribution. θ^* is a vector that consists of values for each parameter

2. Plug θ^* into the model, and use the model to generate simulated data D^* , where D^* is of the same form as the observed data D , e.g. includes same summary statistics
3. The discrepancy between the simulated data and observed data is measured by a distance function $d(D^*, D)$. For a pre-fixed threshold value ϵ , if $d(D^*, D) < \epsilon$ then the sample θ^* is accepted
4. Repeat 1, 2 and 3 until enough numbers of samples are accepted. The posterior distribution then can be approximated by these samples

In our study of the regeneration model, full trajectories of the dependent variables are used as the data to be compared, as our goal is to find a model with a parameter set that can intuitively recover the real process as much as possible. The mean values at each time point for each dependent variable are calculated to form D . For each sample, the ODE is accumulatively solved at the same time points and generate the simulated trajectories D^* . Using other statistics to summarise the raw data are also possible, but care must be taken as the choice of summary statistics can largely affect the resultant posterior [9, 10].

Instead of fixing the threshold value, the algorithm can be more adaptive by using a sequence of decreasing threshold values ϵ_t , and derive new populations (called generation) of samples from the accepted samples (called particles) in the previous population. An example is ABC SMC [8], which can significantly reduce the required samples compared to ABC-rejection [11]:

1. Set a threshold schedule $\{\epsilon_1, \epsilon_2, \dots, \epsilon_T\}$ where T is the total number of populations and $\{\epsilon_1, \epsilon_2, \dots, \epsilon_T\}$ is a descending sequence; set the number of particles in each population as N , set the population index $t = 1$
2. if $t = 1$, draw samples from prior distribution until N particles are accepted under threshold ϵ_1 , i.e. each of N particles generates data D^* s.t. $d(D, D^*) < \epsilon_1$
If $t > 1$, draw samples from previous population $\{\theta_{t-1}^*\}$ with weights $\{w_{t-1}\}$, perturb the particles using a perturbation kernel $K(\theta^*|\theta)$ and obtain $\theta^{**} \sim K(\theta^*|\theta)$
3. Repeat step 2 until N particles are accept under threshold ϵ_t
4. Calculate the weights of the accepted particles. For particle i , the weight is

$$w_t^i = \begin{cases} 1 & \text{if } t=1 \\ \frac{p(\theta_t^i)}{\sum_{j=1}^N w_{t-1}^j K(\theta_t^j|\theta_{t-1}^j)} & \text{if } t \neq 1 \end{cases} \quad (2.3)$$

5. Normalise $\{w_t^i\}$, increase population index t by 1
6. Repeat step 2, 3 and 4 until $t > N$
7. Output the accepted particles $\{\theta_t^*\}$ and their weights $\{w_t^i\}$ in each generation t

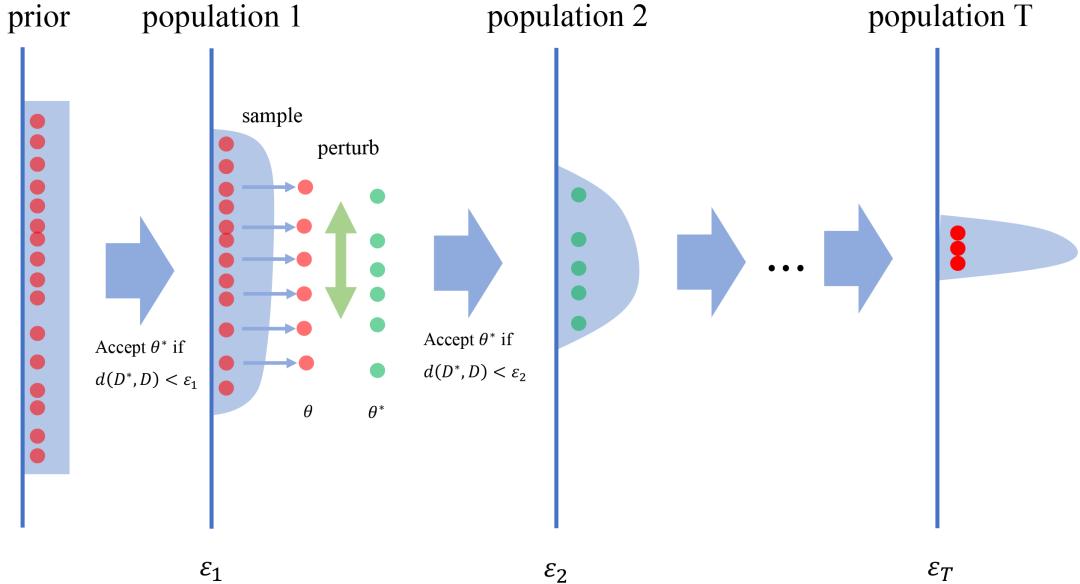


Figure 2.2: ABC SMC sampling process. θ^* denotes a sample drawn from previous population, which is used to generate simulated data D^* . d is a distance measurement, $d(D^*, D)$ represents the discrepancy between observed data and simulated data. for each population t , ϵ_t is a threshold criterion to determine whether to accept that drawn sample

An illustrative graph of the ABC SMC procedures can be found in Figure 2.2. In implementations, many factors can affect the efficiency and goodness of fit; some of them are discussed in Chapter 4. Rather than a pre-fixed threshold schedule, median or quantile based epsilon schedule can save the total sampling size and thus improve acceptance rates of particles [12]. Some other improvements include using adaptive population size for each generation [13], using adaptive distance function [14] and switching kernels [15], etc. In our implementations, these options were tried and compared when applicable.

In addition, ABC SMC is also capable of model selection tasks [16]. Instead of draw samples from prior of a single model, we can draw samples from multiple models m_i and approximate $p(m_i, \theta | D)$, which is the joint probability distribution of model m_i and its accepted particles θ . In each generation, the marginal distribution $p(m_i | D)$ is calculated as the accepted θ is known; thus models can be ranked according to their probabilities.

The feature that ABC algorithm can perform likelihood-free inference makes it widely used in many dynamic modelling problems, especially in the area of biology [2, 11, 17]. However, ABC is not restricted to these areas, as it can be generally used in inference tasks to give parameter estimates or model rankings. E.g., it has been successfully applied in cosmology studies [18].

2.4 Software tools

As we decided to apply ABC SMC on our regeneration models for its high-dimensional parameter estimation and comparison of different models, a code development and experiments workflow was expected to be built. The ABC SMC algorithm needs a whole set of mathematical environments (e.g. class of distributions, models and particles) and closely linked sampling and fitting algorithms (e.g. KDE fitting and Gaussian sampling), thus suitable platforms with related software packages are preferred. Regarding this, several packages in R and Python were studied.

Software packages that integrally implement ABC SMC are preferred, and more options in the settings of the algorithm, more customisable features make it better. Moreover, as the ABC SMC itself is a computationally intensive task that contains massive parallel potentials, an ideal software package is expected to support at least one kind of parallel options, e.g. multi-core, GPU accelerating and clusters. Otherwise, a modification or rewrite of the software package is expected to produce a reliable input-output framework for our parameter estimation and model selection experiments, where much more work would be introduced.

Preliminarily pyABC packages [19] in Python was chosen for implementations. pyABC supports multiple types of parallelisation, e.g., multi-core processing and distributed cluster, and supports resume stored runs, which made our experiments easier. It is well documented and flexible in customisations of the algorithm.

Some other ABC SMC related packages were also examined and considered as backups, include ELFI¹, ABC-SysBio [2], EasyABC², GpABC³, etc. These packages may have their advantages (e.g. ABC-SysBio supports CUDA acceleration) and considered as alternative implementation methods, or can be used to compare performance difference in future works.

Additionally, some other software tools were also needed in ODE solving, result analysis and visualisations. The software and hardware environments used in this dissertation are listed in Appendix A.

¹<https://elfi.readthedocs.io/en/latest/>

²<https://cran.r-project.org/web/packages/EasyABC/index.html>

³<https://github.com/tanhevg/GpABC.jl>

Chapter 3

Mathematical modelling

Mathematical models can describe different kinds of dynamic systems, and thus can be used in prediction, analysis and other tasks. An ideal model in our case, can represent reasonable interactions or effects between immune cells and cytokines and reproduce features and characteristics that were observed in experimental data.

Proposed models for the regeneration process are in the form of non-linear ordinary differential equations (ODE). The only independent variable is post-lesion time t , and the equations consists of derivatives of dependent variables (number of cells, mRNA expression of cytokines) with respect to t . The equations were mostly an interpretation of the proposed interaction maps, with terms correspond to interaction paths.

3.1 Observed data

Our models were built on the basis of the existing experimental findings from Tsarouchas et al.[1]. The available measurement data included the numbers of three kinds of cells (neutrophil N , macrophage Φ and microglia) and the relative mRNA expression of four kinds of cytokines (il-1 β , tnf- α , tgf- β 1a and tgf- β 3). As proposed by Tsarouchas et al., neutrophil and macrophage play important roles in the spinal cord regeneration with il-1 β and tnf- α being the media. According to this, our initial models focused on the changes of four dependent variables N , Φ , β (for il-1 β) and α (for tnf- α) that are most important and determinative in the regeneration. N and Φ is of the unit ‘number of cells’, β and α was recorded as ‘relative mRNA expression’ (compared to the value at the initial time point i.e. 0 hpl) and considered unitless.

It was noted that the variance of the measured data is relatively high. The summary statistic used for the parameter estimations was mean of measurement data at each time point, assuming that measured data was Gaussian-like distributed. To validate this, the distribution of the measured data points was plotted and examined to see if the mean value can represent the distribution. The result was that at most time points the measurement values are Gaussian-distributed, although some distributions were skewed. One

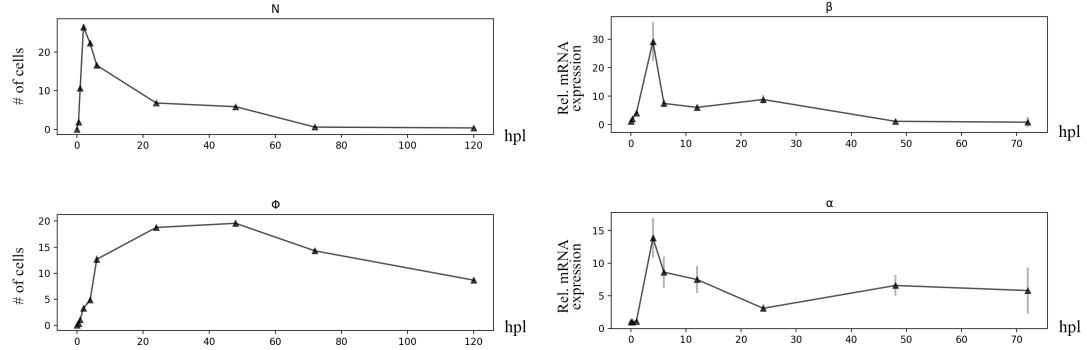


Figure 3.1: Mean of the observed data for neutrophil (N), macrophage (Φ), $\text{il-1}\beta$ and $\text{tnf-}\alpha$, from experiment results of [1]. Error bars indicate standard error of mean

abnormal distribution was observed at time point 120 hpl for macrophage where there are two concentrations. Despite this, mean value could summarise most data and thus was still used as the target observed data. A plot of the mean of the four variables is shown in Figure 3.1, which is our target trajectories for the models to fit.

3.2 Hypothesis and models

5 models in total are proposed according to different hypothesis. At first our tests and implementations of ABC SMC for parameter estimations use only the basic model for developing propose. After the parameter estimation framework is built and tested, more models are proposed, in order to calibrate and adjust the basic model such that it can represent the observed regeneration process better or can be used to test our hypothesis.

All these models assume the involved interactions is within two kinds of cells (neutrophil and macrophage) and two kinds of cytokines ($\text{il-1}\beta$ and $\text{tnf-}\alpha$) and use the data presented in Figure 3.1 for the inference task. Interactions or effects from other cells or cytokines is not considered as there might not be corresponding data available.

3.2.1 Basic model

A preliminary model (Eqn. 3.1) was proposed according to [1] and then mainly used to build and test the code of inference framework. An interaction map of the model is shown in Figure 3.2. This model is a simplification a the process described in [1] with some minor interactions ignored. To describe the parameters' units, we denote the unit of N and Φ i.e. number of cells as 'cell', and denote β and α 's relative mRNA expression as unitless for simplicity. This model included 12 parameters, all of which

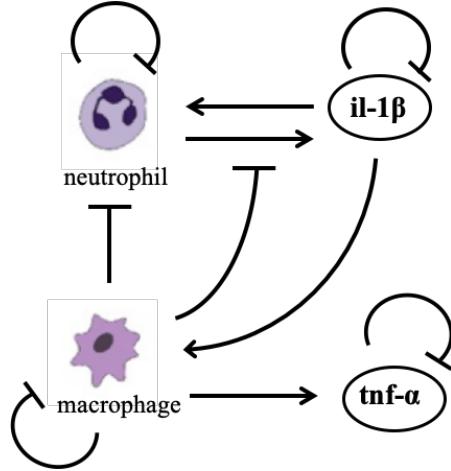


Figure 3.2: Interactions modelled in the basic model (model 1) based on Tsarouchas et al.[1]. Lines ended with arrow represent promoting effect, lines ended with T-connectors represent inhibition

should be positive real number. A description of these parameters can be found in Table 3.1.

$$\begin{aligned}
 \frac{dN}{dt} &= \lambda_N + \kappa_{N\beta}\beta - \mu_N N - \nu_{N\Phi}N\Phi \\
 \frac{d\Phi}{dt} &= \lambda_\Phi + \kappa_{\Phi\beta}\beta - \mu_\Phi\Phi \\
 \frac{d\beta}{dt} &= \frac{s_{\beta N}N}{1 + i_{\beta\Phi}\Phi} - \mu_\beta\beta \\
 \frac{d\alpha}{dt} &= s_{\alpha\Phi}\Phi - \mu_\alpha\alpha
 \end{aligned} \tag{3.1}$$

It was assumed that there are negative feedbacks the cells and cytokines, i.e. overcrowding inhibits the increasing rates. As sources, macrophage produces TNF- α and neutrophil produces IL-1 β , however, macrophage was assumed to have negative effect on the IL-1 β , term $1 + i_{\beta\Phi}\Phi$ in Eqn. 3.1 was used to represent this. Despite of self-driven increase of immune cells, IL-1 β was assumed to promote the increase of both cells. Neutrophil was also assumed to be inhibited by the total immune cells presented in the lesion site, term $\nu_{N\Phi}N\Phi$ was used to denote this.

3.2.2 Alternative models

Model 2 and model 3 As the observed data indicated, this dynamic system has a steady state where after a long time (≥ 120 hpl), the inflammation is resolved and

Parameter	Definition	Units
λ_N	Self-increase rate of neutrophil	cell/h
$\kappa_{N\beta}$	Promoting effect coefficient by il-1 β	cell/h
μ_N	Coefficient of negative feedback of N	h^{-1}
$\nu_{N\Phi}$	Coefficient of inhibition of both N and Φ	cell $^{-1} \cdot h^{-1}$
λ_Φ	Self-increase rate of macrophage	cell/h
$\kappa_{\Phi\beta}$	Promoting effect coefficient by il-1 β	cell/h
μ_Φ	Coefficient of negative feedback of Φ	h^{-1}
$s_{\beta N}$	Production rate from N	cell $^{-1} \cdot h^{-1}$
$i_{\beta\Phi}$	Coefficient of inhibition to the production	cell $^{-1}$
μ_β	Coefficient of negative feedback of β	h^{-1}
$s_{\alpha\Phi}$	Production rate from Φ	cell $^{-1} \cdot h^{-1}$
μ_α	Coefficient of negative feedback of α	h^{-1}

Table 3.1: Parameters introduced in the basic model (model 1) [REMOVE unit]

immune cells should not be present at the injury site any more. Regarding this, the parameter for self-increase rate for the macrophage and neutrophil i.e. λ_Φ and λ_N cannot be constant. An exponentially decaying λ term is introduced in the new model (model 2, Eqn. 3.2). Also, the inhibition of il-1 β production, i.e. term $i_{\beta\Phi}\Phi$ is considered to be ignored for a more simple model, in which case the relative expression of il-1 β is only affected by the number of neutrophil and the negative feedback from itself. This case corresponds to model 3, written as Eqn. 3.3.

Model 2 and model 3 introduced one extra parameter a and removed λ_Φ , which is a coefficient in the exponentially decaying λ_N , determining the decay speed, with the unit h^{-1} .

$$\begin{aligned}
 \frac{dN}{dt} &= \lambda_N e^{-at} + \kappa_{N\beta}\beta - \mu_N N - \nu_{N\Phi}N\Phi \\
 \frac{d\Phi}{dt} &= \kappa_{\Phi\beta}\beta - \mu_\Phi\Phi \\
 \frac{d\beta}{dt} &= \frac{s_{\beta N}N}{1 + i_{\beta\Phi}\Phi} - \mu_\beta\beta \\
 \frac{d\alpha}{dt} &= s_{\alpha\Phi}\Phi - \mu_\alpha\alpha
 \end{aligned} \tag{3.2}$$

$$\begin{aligned}
\frac{dN}{dt} &= \lambda_N e^{-at} + \kappa_{N\beta}\beta - \mu_N N - \nu_{N\Phi} N\Phi \\
\frac{d\Phi}{dt} &= \kappa_{\Phi\beta}\beta - \mu_\Phi\Phi \\
\frac{d\beta}{dt} &= s_{\beta N} N - \mu_\beta\beta \\
\frac{d\alpha}{dt} &= s_{\alpha\Phi}\Phi - \mu_\alpha\alpha
\end{aligned} \tag{3.3}$$

Model 3 can be regarded as a simplification of model 2, as it can be treated as model 2 with parameter $i_{\beta\Phi} = 0$. Among the proposed three models, model 1 is a naive one that was proposed at very first time and used as an ODE ‘template’ to build and test our parameter inference framework. As the implementation was successful, model 2 and 3 was proposed, as we were trying to calibrate some terms and find better models. After fitting, model 2 is supposed to be better than model 1 as it corrects the problem that appears at the final time points (which were related to the steady states of immune cells). Model 3 made a small simplification based on model 2 and thus it was theoretically less ‘general’ than model 2.

Model 4 and model 5 After the first model selection experiment among the exiting 3 models, it was found that some significant features presented in the observed data were not recovered by any of the model. To resolve this, attempts were tried to introduce additional reasonable interactions within the dynamic system, considering the biological and mathematical context. Extra promoting effect to the expression of tnf- α was considered, by either adding a phenomenological term $d_{\beta\alpha}\beta$ (which means the same effect as directly promoting but the underlying mechanism is unclean), or adding a term $f_{\beta\alpha}$ that represents a promoting effect to the production process of tnf- α , namely model 4 (Eqn. 3.4) and model 5 (Eqn. 3.5).

$$\begin{aligned}
\frac{dN}{dt} &= \lambda_N e^{-at} + \kappa_{N\beta}\beta - \mu_N N - \nu_{N\Phi} N\Phi \\
\frac{d\Phi}{dt} &= \kappa_{\Phi\beta}\beta - \mu_\Phi\Phi \\
\frac{d\beta}{dt} &= s_{\beta N} N - \mu_\beta\beta \\
\frac{d\alpha}{dt} &= s_{\alpha\Phi}\Phi - \mu_\alpha\alpha + d_{\beta\alpha}\beta
\end{aligned} \tag{3.4}$$

$$\begin{aligned}
\frac{dN}{dt} &= \lambda_N e^{-at} + \kappa_{N\beta}\beta - \mu_N N - \nu_{N\Phi} N\Phi \\
\frac{d\Phi}{dt} &= \kappa_{\Phi\beta}\beta - \mu_\Phi\Phi \\
\frac{d\beta}{dt} &= s_{\beta N}N - \mu_\beta\beta \\
\frac{d\alpha}{dt} &= (s_{\alpha\Phi} + f_{\beta\alpha}\beta)\Phi - \mu_\alpha\alpha
\end{aligned} \tag{3.5}$$

Parameter	Definition	Units
$d_{\beta\alpha}$	Coefficient of promoting effect from β	h^{-1}
$f_{\beta\alpha}$	Coefficient of promoting the production of α by β	$cell^{-1} \cdot h^{-1}$

Table 3.2: Parameters introduced in model 4 and 5

3.2.3 Limitations

Some limitations were included in Section 3.1, where the observed data have a relative high standard deviation at certain measurement points; more measurement repeats or more measurement points could help to represent more accurate trajectories of the modelled variables and thus more possible models can be proposed. On the other hand, more accurate observed are always beneficial to mathematical modelling, especially to complex dynamic systems.

As stated, our models are largely based on the experimental evidence from Tsarouchas et al.[1]. Some hypothesised interactions may differ from real case, and there may exist undiscovered mechanisms. These could lead to model misspecification, where the model could fit the data but with wrong interaction paths, or the model can hardly fit the data and recover some local features observed in the data.

Chapter 4

Parameter estimation and model comparison

For the proposed models, exploration of the uncertainties in the parameter estimations is our main task in this chapter, which involved the application of ABC SMC and several experiments. After that, the parameter inference framework was also used for model selection.

The build and test of the code were under Python environment with `pyABC`[19]. Some other code, e.g. shell scripts and R code were also used to perform the experiments and result analysis. Software and system environment used in our implementation can be found in Appendix A. The build and test were performed on local computers and HPC facilities available within EPCC.

4.1 Code and implementations

According to our preliminary studies, a ABC SMC framework implemented using `pyABC` in Python was adopted in this project. `pyABC` is a popular SMC software packages[19] which has been used form many ABC SMC inference applications[13]. `pyABC` provides an open-source framework for likelihood-free inference, which is a NumPy implementation of ABC SMC algorithm and a useful tool-box for our own inference and analysis tasks. Besides, it supports multi-core execution implemented using multiprocessing built-in package, and it made our further studies in the scaling-up performance possible.

The inference framework code for the project was developed and tested in local environment (macOS 10.15.6, see Table A.1) at first using a small input, then the functional version was deployed on compute node of ARCHER and Cirrus for parameter inference experiments and model selection tasks. The scaling-up experiment was performed on Cirrus. The profiling and its analysis were performed on local machine. Some nec-

essary development and debug were performed on the remote machines to ensure that experiments run correctly on compute node.

The ODE model-related code and utilities were firstly developed to enable reusable functions, variables and data calls. It included the code for the ODE equations, ODE solver, data structures and data format transform functions, etc. By doing this in separate code files, we could change implementation options and activate ABC SMC easily in another ‘trigger’ file, without creating definition and duplicated code for the models, observed data and parameter priors.

Some missing features were found when using `pyABC`, therefore some modifications were made to the source code of `pyABC`.

For example, the laboratorial measured data were not considered ‘complete’ or ‘tidy’ for modeling and analysis purpose: the available time points for cells and cytokines were different, e.g. cell number was not measured at 0.25 hpl and cytokines’ expression was not measured at time point 120 hpl. These missing values were treated as ‘ignored’ when calculating the distance between observed data and simulated data, and represented by NaNs. However, in some distance calculations NaN was not acceptable; to cope with this, built-in distance function was modified. Other modified code include some visualisation and result representation enhancements

Analysis of the program output was also wrapped up into functions which can be easily called after an ABC SMC run is finished, which includes multiple visualisations and summary statistic of the results. For different tasks there are also separated analysis code file. Other analysis tools used in the project include built-in database visualisation tool, R code and Microsoft Excel.

The project code was managed via `git` version-control and available as a GitHub repository¹; it also made the cross-hardware synchronisation of code versions convenient.

After several development iterations, code files for ABC SMC implementation now mainly had two types: code files for infer-back experiments (using synthetic data), and code files for parameter inference (using experimental data). Model selection can be easily done by adding more model to the implementation code file.

4.2 Hyperparameters and implementation options experiments

As the key focus of this project was on the parameter estimations of dynamical system, the ability of inference on the proposed models was firstly examined before actual inference using the experimental observed data.

¹https://github.com/chaolinhan/MSc_Project

In the first part, synthetic data is used with known true parameter values. The algorithm implementation is evaluated by the efficiency and goodness of resultant model under different implementation options and ABC SMC hyperparameters.

Then according to findings in the first part, several options with good performance are tried in the inference using experimental observed data (Figure 3.1). To obtain a more accurate and general results, these experiments are repeated for 3 or more times.

The ABC SMC inference can have largely different performance when using different hyperparameters and implementation options. To firstly test the ability of inference and observe the results of different options, synthetic data with known true parameter values are used as target data. by doing this, we could directly compare the inferred parameters to the true value, and compare the simulated trajectories to the true trajectories and then observe the result to see whether the inference is successful and how well the model fit the data. Also, by setting the same final threshold value, the efficiency under different options can be compared and help us to choose proper options when applying ABC SMC on the observed data. The true value of parameters that were used to generate synthetic data is listed in Table B.1.

The true values were obtained by fitting model 1 onto the experimental data (Figure 4.1) using least-square optimisation method. Available least-square fit tools in `SciPy` were used. The fit algorithm was Trust Region Reflective². Target data was then generated using these parameter values via ODE integrate solver³.

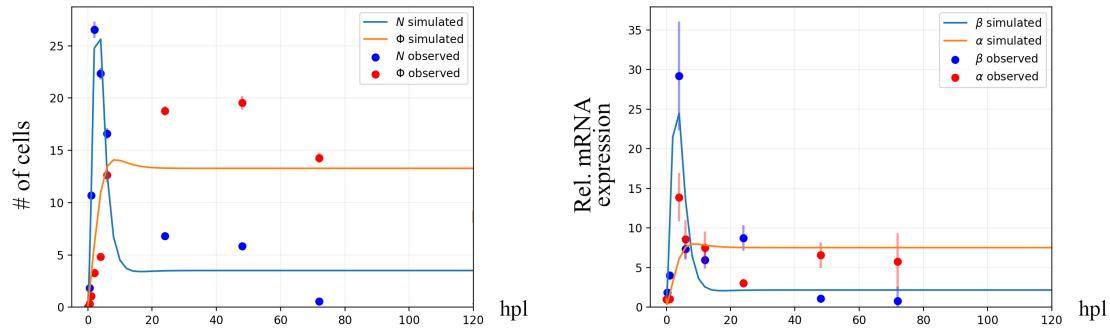


Figure 4.1: Synthetic data generated with known parameter values (line plot), compared to experimental data (scatter plot). Known parameter values are obtained from a least square fitting of the observed data

The following topics were studied using the synthetic data and ABC SMC inference framework.

²https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html

³<https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.odeint.html>

4.2.1 Perturbation kernels

Perturbation kernels work in the sampling process. In each generation t , samples are firstly taken from the previous population $\{\theta_{t-1}\}$ with weights $\{w_{t-1}\}$, then perturbed using the perturbation kernel $K(\theta|\theta^*)$. We keep sampling until N particles are accepted, where N is the pre-set population size. After that the new weights are calculated and normalised. Figure 2.2 illustrates the process.

The perturbation kernel is called in the sampling of every particles and determines how the new perturbed particle is chosen, thus it is influential to both computation complexity and the resultant posterior distribution. Generally, a local perturbation kernel may face the risk of being stuck in local modes (e.g., local optimal), but it may need less computational operations, or could generate a population with a higher acceptance rate if the successive epsilon values are close; A kernel with wide variance, or spreading out in a large space could help in resolving the local optimal problem by a more thoroughly exploration of the parameter space, however it can be more computation-intensive and result in a lower acceptance rate. A desired optimal kernel should balance the trade-offs; their property and criteria is discussed in [15].

There are several common choice of perturbation kernels. Among them multivariate normal kernel and local M-nearest neighbour model is preferred to be applied on our models. A covariance matrix Σ_t of accepted particles is calculated form previous generation and used in multivariate normal kernel: $K(\theta|\theta_{t-1}) \sim \mathcal{N}(\theta_{t-1}, \Sigma_t)$. It is illustrated to be more efficient than uniform kernel and component-wise normal kernel in relecting the true posterior structure [15]. It has been proved to perform well in several dynamic system models [2, 17, 11] for the parameter estimation and model selection tasks. The *scaling* $\in (0, 1]$ parameter in pyABC will be multiplied to the covariance to produce a ‘narrower distributed’ perturbation result.

Local M-NN kernel provided by pyABC is also tried to provide a comparison. A local kernel density estimation (KDE) fit is used with M nearest neighbours considered.

The kernel experiments is designed to explore the efficiency of SMC on our dynamical systems. Given the same fixed threshold schedule, kernels that need less total samples and have higher acceptance rate would be our preference. The experiments compared multivariate normal kernel with local MNN kernels with different parameters, using synthetic data to infer back the parameters. The acceptance rate among each time point and total required samples are compared after the experiment to give suggestions on the kernel selection in the real data inference. As the threshold schedule is fixed, the final population should have similar discrepancy to the target data thus here the goodness of fit i.e. recover of target data trends/features and errors of the inferred parameters compared to true values is not discussed.

Kernel experiment results

To compare kernels, a fixed schedule schedule is used with minimal epsilon i.e. ϵ_t set to 10.0. From the total required samples graph Figure 4.2, the efficiency can be compared. Among the tested kernels, the local M-NN with $M=50$ has the best performance: it requires the least number of samples and has higher acceptance rates among almost all generations. Local M-NN with $M=750$ is the slowest kernel.

For local M-NN kernels, generally a greater M will lead to lower acceptance rate and more required particles in each generation. As shown in Figure 4.3, local M-NN with $M=750$ is the lowest curve. Consequently, if greater M is test, then it will have even lower acceptance rates; the maximum $M=2000$ (whole population is considered) will have the lowest acceptance rates.

Multivariate normal kernel has a performance between local M-NN $M=250$ and $M=100$. It proves that rather than a trivial normal kernel, multivariate normal kernels is more efficient facing the concentrations of joint distributions among multiple parameters [15]. ‘Scaling’ option can narrower the distribution calculated from the last population, thus making the kernel more ‘local’ by sampling under a smaller variance. Similar to small M in local M-NN, small ‘scaling’ is also more efficient in our experiment of the model 1. Besides the fixed threshold schedule, a median threshold schedule with the same final threshold $\epsilon_{20} = 10.0$ is also tested and similar results are observed (Figure B.1 and B.2, Appendix B2).

A more ‘local’ kernel was more likely to give better performance by efficiently sampling around concentrations in parameters’ distribution and approximate the posterior distribution. This holds in most cases under a given target final threshold value ϵ_t , however may not produce the proper result want: a more local kernel, e.g. local M-NN with small M and multivariate normal with ‘scaling’ is more likely to be stuck in local optimality, as the kernel is less likely to sample particles that are far from local concentrations, although these local optimal are still accepted under given threshold. In other words, using local kernels the epsilon may quickly converge to to local optimal; if a even smaller epsilon is desired, the local kernel can hardly generate enough particles to find another matched local modes. Also, the shapes of posterior distribution will affect the performance of local kernels [15]. One obvious example is presented in Figure 4.2, where local M-NN with $M=50$ suffered from a local optimal: the last generation takes much more samples to meet the required threshold.

Conclusion

Local kernels can be fast bur unstable in some case; for a general parameter inference of our models, multivariate kernels are preferred, as a good fit is our prior target; some local kernels are also worth trying after the multivariate normal kernel, e.g. local M-NN with $M \leq 250$ (for population size of 2000, i.e. 12.5% of the population) and multivariate normal with scaling. Also, multivariate M-NN is worth trying, although it

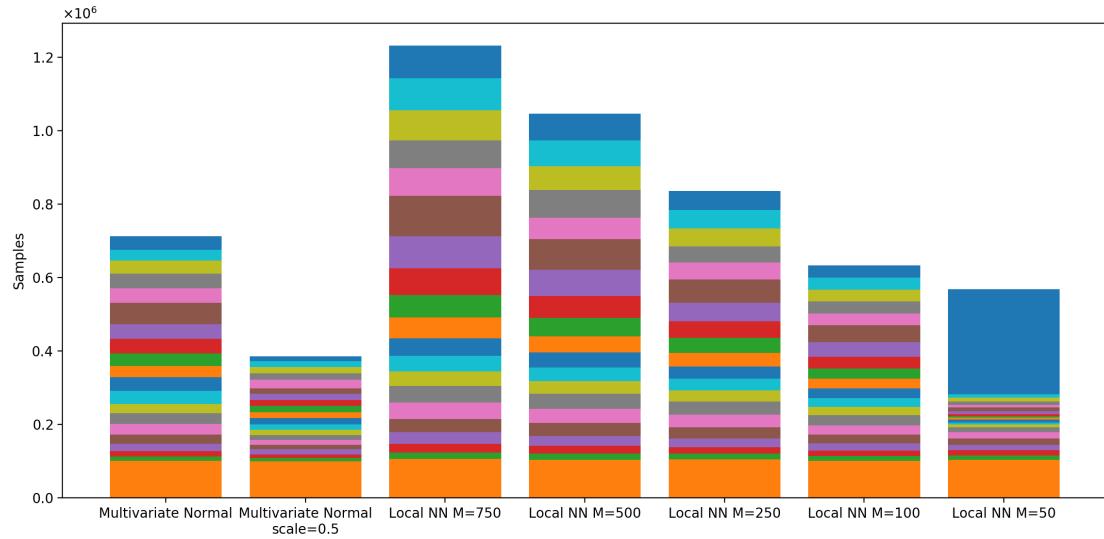


Figure 4.2: Total required samples of different kernels after 20 populations (2000 particles in each population). Different color represents different generations (bottom to top: population 1 to population 20)

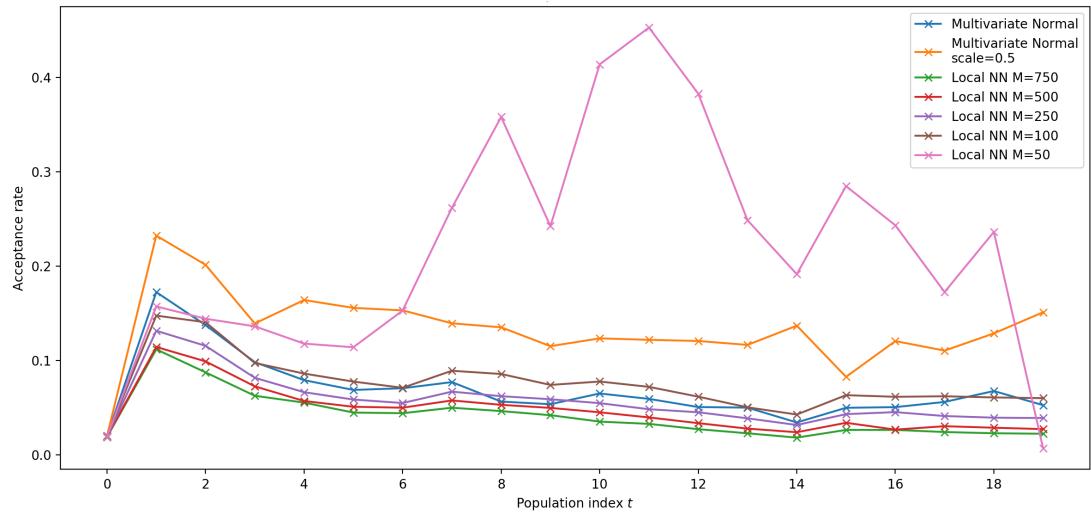


Figure 4.3: Acceptance rates of different kernels in 20 populations. Each population has 2000 particles

was not built in pyABC and requires additional implementations.

4.2.2 Adaptive functions and factors

SMC can be more adaptive by introducing adaptive distance function and adaptive population; besides, factors can be applied to manually ‘normalise’ the data.

The trivial distance function is set to Euclidean distance (2-norm distance)

$$D = \sqrt{\sum_i \Delta x_i^2} \quad (4.1)$$

where i is the index of data points and Δx_i is the discrepancy between observed data and simulated data at data point i , i.e. $\Delta x_i = x_{i,\text{simulated}} - x_{i,\text{observed}}$. In this case data points are 12 time points of four variables i.e. 48 data points in total.

A weighted 2-norm distance can be written as

$$D = \sqrt{\sum_i w_i \Delta x_i^2} \quad (4.2)$$

where a weight w_i is assigned to every data point. If data point i have a higher weight, then the data value is regeared to be more informative, i.e. giving more help in inferring the true posteriors; weights can be either pre-set according to prior knowledge of the problem, or using adaptive method to be dynamically calculated according to , known as adaptive distance.

Adaptive distance is changing the weights of all data points after each generation iteration, trying to assign informative data points higher weights. Additionally, implementation of adaptive in pyABC also introduces additional factors f_i multiplied to weights [14]

$$D = \sqrt{\sum_i f_i w_i \Delta x_i^2} \quad (4.3)$$

Factor is helpful as an option for efficiency. When some data points are equally informative, the manually pre-defined can make the distance more focused on certain data point, or behave like a normalisation that can balance the scales of different summarised statistics (here is the 48 mean values). There are two appliance case studied: (1) factors used as an normalisation option and (2) factors used to give more focus on some

features of the observed data. For (2), as we noticed that the main features (e.g. rapid increase, peaks, fluctuations) are mostly observed in the first half of the time points i.e. 0 - 30 hpl, factors are tried in the later parameter estimation of real data where a more accurate fit of the curve is desired.

Adaptive population [13] can adapt the population size of each generation according to the mean correlation of variation error of the previous population. In our early test the maximal population size that is allowed is set to 5,000 and the adaptive strategy always select the upper bound, as a results of wide parameter space or the wide posterior approximation shape.

Experiment results

Two types of factor and adaptive distance function were tested by running the same number of generations (20 generation, 2000 particles per generation). The first type of factor is 25:75 factor, where first half of the trajectory data was assigned higher factor (75), and second half of the data was assigned lower factor (25); this could make the first half of the data more ‘important’ in the inference. The second factor was range factor, which tried to normalise the trajectory of each dependent variable according to its data range.

From Figure 4.4, different types of factor gave different performance. 25:75 factor required much more samples in the first generation (orange bar in the left of Figure 4.4) and also did not improve much in later generations, thus it is the slowest one. The first half of observed data contained more features and information, assigning it with higher factor would make the algorithm think that it is more important to fit the first half and thus focus on the hard part, and consequently more particles were needed in the inference process. The acceptance rates of 25:75 factor among different generations were also generally slightly lower than the standard. By contrast, range factor required less total samples and gave generally higher acceptance rates than standard, which made it considerably faster. Adaptive distance gave the best efficiency, as it significantly improved the acceptance rates and resulted in even less required samples (right, Figure 4.4).

However, the result of factors adaptive distance could not be directly compared to standard run, because it changed the distance measurement by adapting weights or manually setting factors. Factors were normalise to make it more close the standard distance measurement. As a result, the same final threshold value did not guarantee that adaptive distance can reach the same level of fit as others. After 20 populations, adaptive distance had a poor fit (Figure 4.5), while the standard run and two types of factors had similar tight fits of the data, which all converge to the true posterior. Additional in our implementations, variance factor (which tries to normalise the data according to the dependent variables’ variance) was also tries and gave very similar performance as range factor.

To conclude, because the factors and weights are directly multiplied to Δx and thus

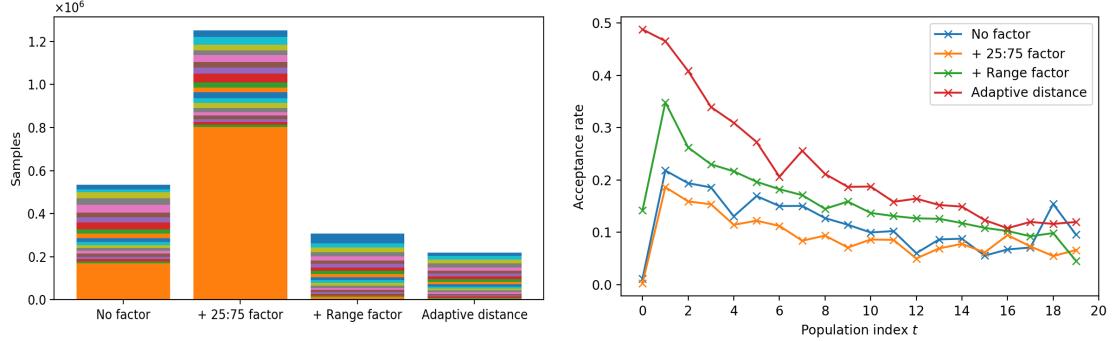


Figure 4.4: Factors and adaptive distance experiments. Left: total required samples. Different color represents different generations (bottom to top: population 1 to population 20). Right: acceptance rates in each generation.

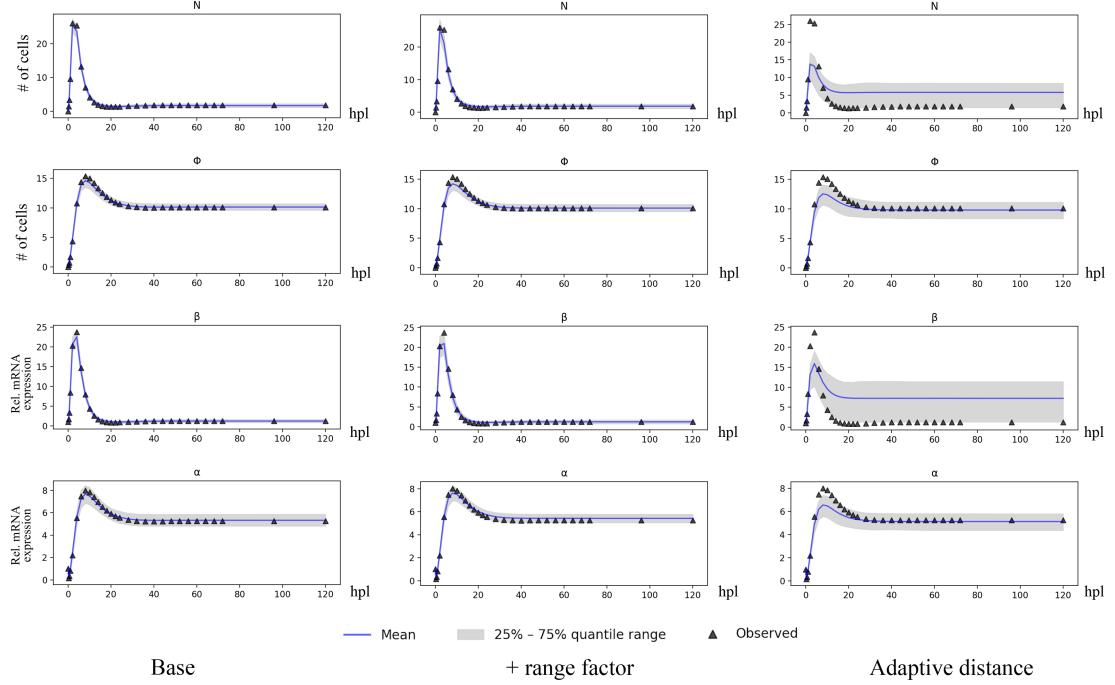


Figure 4.5: Simulated trajectories from the last population in the factor and adaptive distance experiment. 500 randomly chosen particles from last population are used to generate a sequence of trajectories where mean and inter-quartile range are calculated from. 25:75 factor showed very similar curve as the range factor.

the metrics for distance are changed, the direct compare of the efficiency under the same threshold schedule was unfeasible; we could only compare the results after the same number of generations. Factors gave nearly the same inferred results but required different number of samples, hence cares must be taken when choosing factor schemes. Adaptive distance did not result in an accurate result as others. As a result, our further

implementations used standard distance function, and factors were tried for possible improvements.

4.2.3 Data size, prior distribution and population

Some other hyperparameters, e.g. feed-in data size, prior distribution, number of populations and number of particles in each population were also of our interest. Experiments were designed to explore how these options affect the goodness of fit and efficiency of the implementation.

Regarding the existing dynamic systems model, SMC was applied with two data size options, three prior distribution (uniform distribution, wider uniform distribution and log-uniform distribution). The population options i.e. population size and number of populations were also tried with different values. These experiments intended to give suggestions on the later SMC implementation on experimental data.

Experiments results

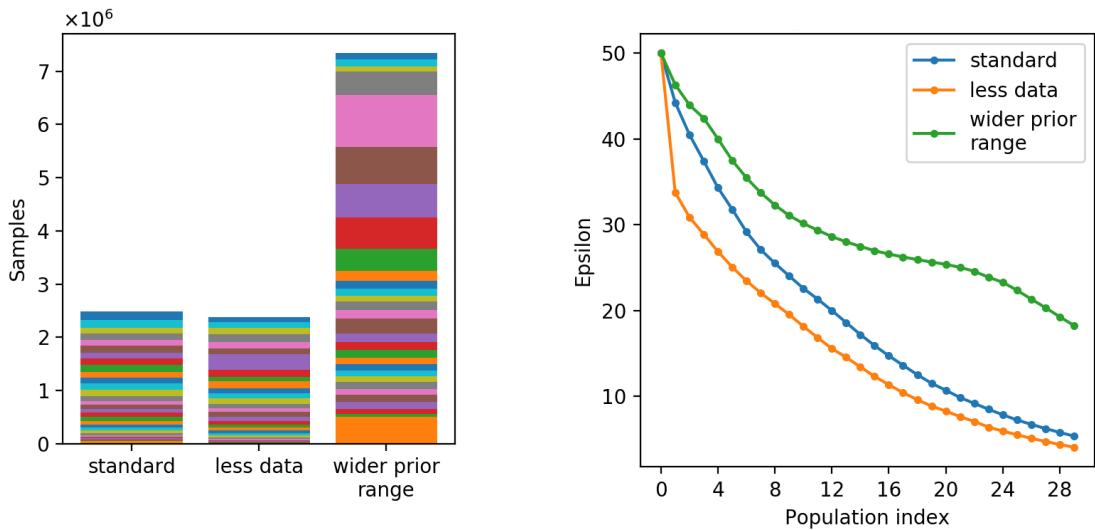


Figure 4.6: Data size and prior distribution range experiment. Left: total required samples. Right: epsilon values under median epsilon schedule

The results from data size and prior distribution range is shown in Figure 4.6. The standard implementation is fed with 120 data points (30 data points for each of the four variables), the ‘less data’ is fed with 48 data points, which is the case of the real experimental measurement. Although much less (60%) data was used, the total required sampling numbers was only 4.47% less. This indicated that the data size does not affect the sampling process much and the inferred model are all considered acceptable when

compared to the synthetic data, as they finally reached a similar epsilon threshold ((b), Figure 4.6), but care should still be taken when using less data points, or using other more summative statistics (e.g. mean, standard deviation, peak values, starting and ending values etc.) as the target data: as shown in [11], less data can lead to a model with more variance, under-fitting or missing some local features such as local peaks.

The prior range and distribution can largely affect the inference progress (Figure 4.6. (a)). Samples were drawn from a high dimensional parameter space in each sampling process, a wider parameter prior interval in only one dimension can cause the parameter space to be much bigger, and consequently the exploration and sampling of parameter vectors would take much more efforts. Figure 4.6. (b) also indicates that the convergency is slower for wider prior ranges. As a result, a proper prior interval based on our experience and belief with seeing the data should be set as narrow as possible, as inference are largely affected by the prior interval and its length; improper prior could even make the algorithm miss the true posterior [2].

How to specify prior can be tricky from problem to problem. For our task in the parameter inference of regeneration model, biologically motivated priors were used, which were general reflection of biological constraints. Also the units of the parameters help us in identifying the possible scale of the parameters.

Choice of options

This subsection aims to preliminarily explore options in ABC SMC implementations. From the above results, we adopted multivariate normal kernel as the kernel for our real inference on observed data; Local M-NN kernels were also planned to be tested. In the real inference, 25:75 factor were also tried to see if it could help in recovering more features and resulting a better fit. because the improvement by using less data was inapparent, we did not try to change the target summary data size. Prior intervals were explored explicitly and gradually shrunk for our models.

Some other hyperparameter, e.g. number of generations, population size (i.e. number of particles per generation) were not addressed in this subsection. In the real inference, population size was set to be greater than 2000 to avoid insufficient exploration of the previous population; in some cases than high variance or uncertainties in the results were observed, population size as large as 20,000 were used. If the epsilon did not show a convergency trend, the run could be extended to more generations until a stable result was observed.

4.3 Parameter estimation and model comparison

4.3.1 Model 1, 2 and 3

Using the suggested options in the previous section, the target data (Figure 3.1) is prepared as input for the SMC inference framework. Regarding the result, the population size and number of generations are adjusted several times to obtain an general informative posterior which should be stable across repeated runs and avoid local optimal as much as possible.

Parameter estimation of model 1, 2 and 3 was tried with different prior distribution setting: distribution range [0, 25] and [0, 75], and distribution type uniform and log-uniform. All these ABC SMC runs uses 2,000 particles per population and 30 generations.

Log-uniform-shaped prior seems to give a better fit of the model, and wider prior range [0, 75] does not give a more accurate result. Then distribution range [0, 25] was tried, in case some parameter have a true posterior laying between [25, 50].

The results of prior distribution log-uniform [0, 50] is shown in Figure 4.7. For model 3, the acceptance rates and epsilon path are presented in Figure 4.8. For model 3, the estimated parameters' posterior is shown in Figure 4.9. Figure 4.7 plots the estimated curve of models, using the mean value of each parameter's approximated posterior; also 1000 particles (each particle is a parameter set) were sampled and used to generate simulated trajectories of the four variable, and the mean (blue) and 25th to 75th percentile range (grey) are calculated from the 1000 simulated data.

Before applying model comparison methods, some features of the three model can still be compared. From simulated curves, It can be seen that all the inter-quartile ranges are tight around the mean value curve, which indicates the approximated posterior are concentrated to some degree; the peak-shaped posterior distributions (Figure 4.9) agrees with this. All epsilon values converges to a steady level, and model 3 has a lower final epsilon value and consequently the simulated trajectory is more close to the observed data (model 3, Figure 4.7).

The acceptance rates is fluctuating and have a gradually increase trend for all the three models, as the acceptance rate becomes higher when the true posteriors are gradually approximated. Compared to model 1, model 2 gives a better fit of the decreasing trend at the second half of the observed data, by using a exponentially decaying self-increase rate λ_N instead of a constant. The resultant simulated data from model 2 and 3 are in the similar trends and the most obvious difference is that for model 3, the simulated data of Φ and α are more 'flat'. We expect model 3 to be the best model as it reaches a threshold value and intuitively gives a better fit, although the fit could not be considered generally a good representation of the biological process that we want to model, as the simulated data are significantly biased from the observed data for tnf- α . A sharp peak at 4 hpl is observed from the measurement (Figure 3.1) but no similar features are represented by

Parameter	Estimated mean value (uniform)	Estimated mean value (log-uniform)
λ_N	14.1	13.3
a	9.67	0.0170
$\kappa_{N\beta}$	18.0	0.259
μ_N	13.2	0.0180
$\nu_{N\Phi}$	0.742	0.131
$\kappa_{\Phi\beta}$	0.239	0.156
μ_Φ	0.154	0.0350
$s_{\beta N}$	6.75	1.21
μ_β	6.38	1.32
$s_{\alpha\Phi}$	17.0	5.83
μ_α	11.9	2.43

Table 4.1: Estimated parameter values of model 3, using uniform prior distribution and log-uniform prior distribution respectively

any of the three models.

[model comparison among model 1, 2, 3]

Next a model selection experiment with different prior was conducted, using the same ABC-SMC framework. As in the parameter inference, two distributions (uniform and log-uniform) are tested with three different interval ([0, 25], [0, 50] and [0, 75]). The results show that model 3 wins with nearly 100% model probability at most runs after 30 generations, except log-uniform [0, 75]. All the model probability results are of great discrepancy (e.g. 100% model 3 and 0% for model 1 and 2), so Bayesian factor is not calculated.

Figure 4.10 shows the model selection process, under log-uniform [0, 50] prior distribution. Model 3 gains total advantage over model 1 and 2 after 7th population. Together with conclusions above, model 3 is considered to be the the best model to represent the target data so far.

Compared to uniform distributed prior, we found that log-uniform distributed prior is more suitable in the inference. Log-uniform results in an obviously better fit and narrower inter-quantile interval (Figure 4.7 and B.3), and the final epsilon value is also smaller. Log-uniform tends to assign values that are close to the left boundary (i.e. zero in our case) higher density, i.e. they are more likely to be sampled in the first population. As a result, the estimated parameter posteriors are expected to have smaller mean and/or skewed to left of the x-axis. The estimated values proved this (see 4.1; also observed in other models and prior interval experiments). It indicates that some of the true parameter tend to have small values that are close to zero, and using log-uniform distribution could results in a faster and more satisfying inference. Regarding this, further experiments all use log-uniform distribution only.

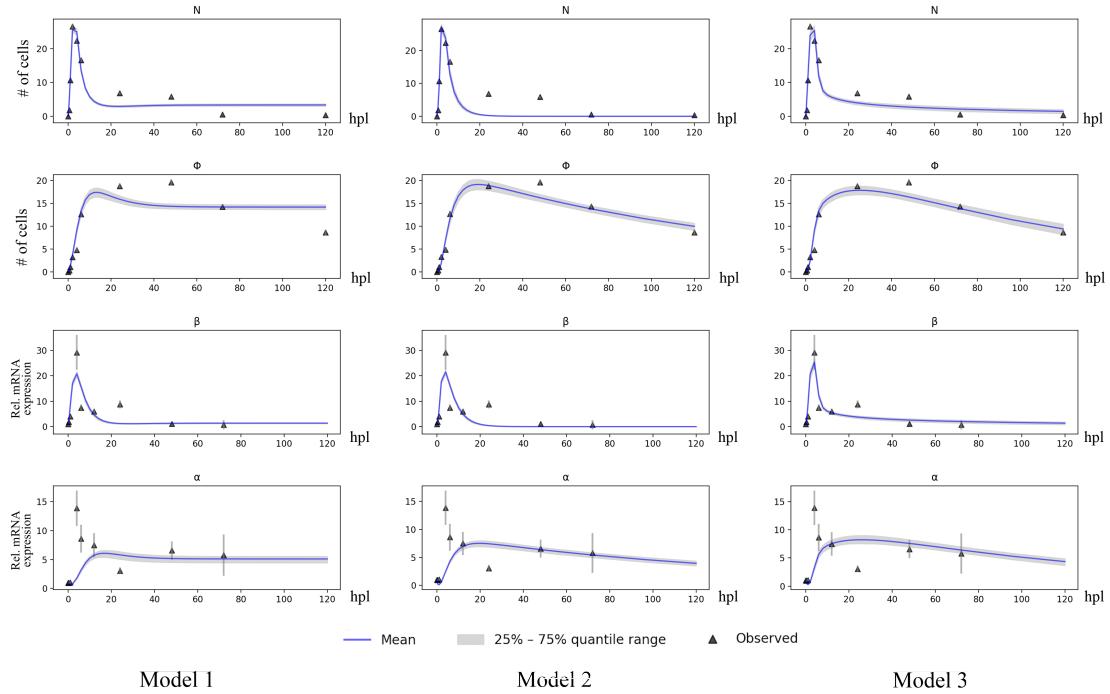


Figure 4.7: Simulated trajectory from the last population of model 1, 2 and 3. Error bar indicates SEM. 500 randomly chosen particles from last population are used to generate a sequence of trajectories where mean and inter-quartile range are calculated from

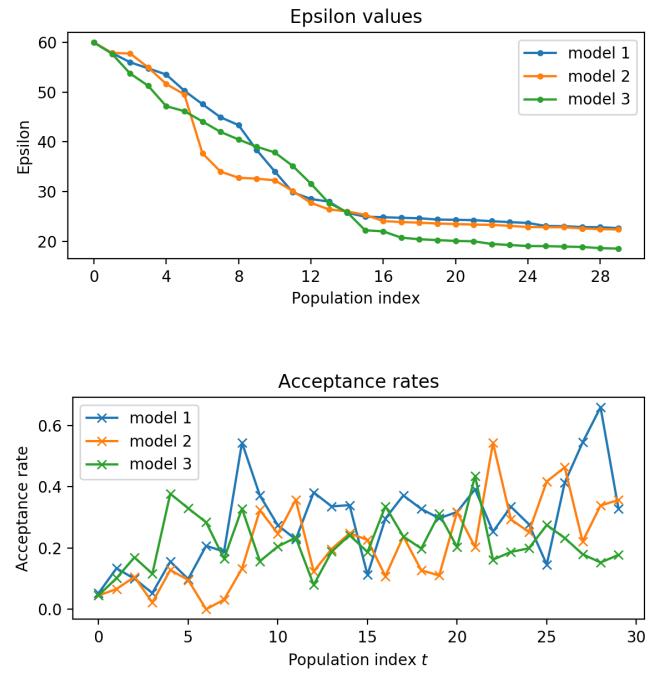


Figure 4.8: Epsilon trends and acceptance rates of model 1, 2 and 3

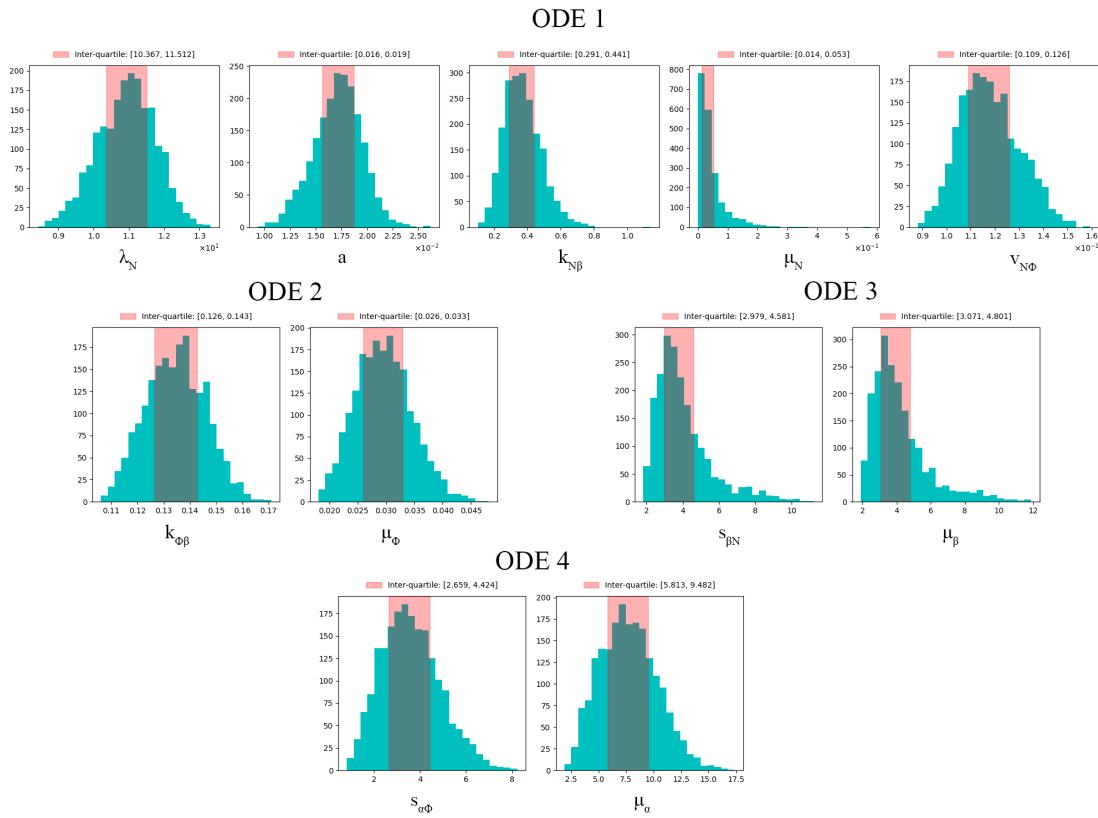


Figure 4.9: Estimated posterior distribution of parameters in model 3. Shaded range indicates 25%–75% quantile of the population

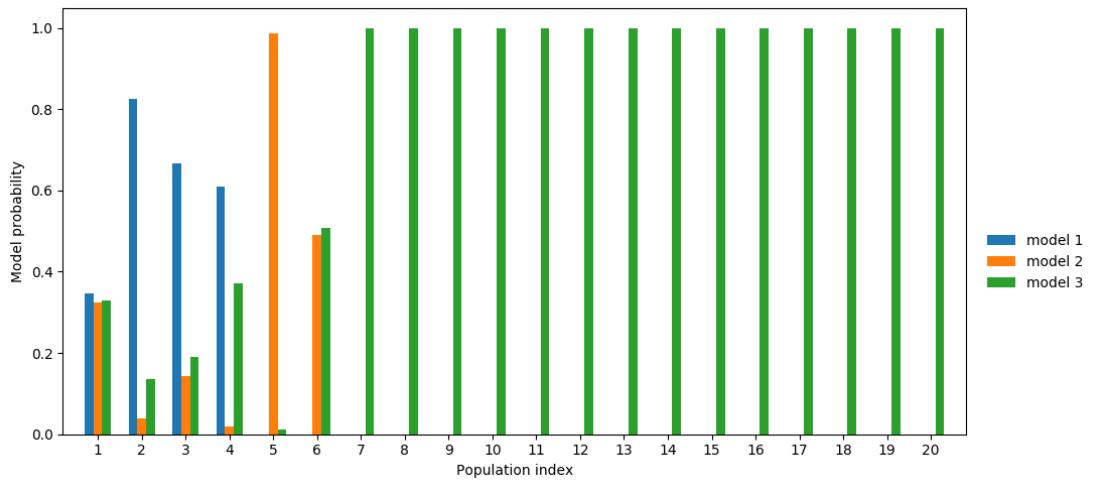


Figure 4.10: ABC SMC based model comparison of model 1, 2 and 3. From population 8 to population 30, model3 always wins

4.3.2 Model 4 and 5

[features that derive model 4 and 5]

We observed that the existing models cannot well represent the trajectory of relative expression of tnf- α : from experimental data, a rapid increase of tnf- α expression is observed in 0–4 hpl, followed by a rapid drop; all the models cannot well recover this feature and most test results see a less-rapid increase followed by a gradually decree and the observed peak value is not reached (e.g. Figure 4.7). Also in some other case e.g. Figure B.3, tnf- α trajectories is well fitted at the cost of under-fitting Φ . Hence efforts had been spent to propose more alternative models that can possibly solve this problem.

The two proposed alternative models is based on model 3 and try to add extra term in the tnf- α equation $d\alpha/dt$. As macrophage is the main source of tnf- α production, and a similar sharp increase-and-drop trend is observed in the il-1 β trajectory, two hypothesis and corresponding terms are propose as follows.

Model 4 It is assumed that the expression of il-1 β would have a promoting effect, representing by a phenomenological term $d_{\beta\alpha}\beta$ and $d_{\beta\alpha}$ is a new model parameter (positive constant). The promoting effect here is regarded to have the equivalent effect as directly promoting by il-1 β , but the underlying mechanism is unclear so far. In mathematical view, this additional term can accelerate the expression of tnf- α in the first few time points and help to recover the peak-shaped trajectory. The proposed model is written as Equation 3.4.

Model 5 Alternatively, we considered promotions to the tnf- α production, i.e. $s_{\alpha\Phi}\Phi$. It is assumed that il- β could accelerate the production process. An additional term $f_{\beta\alpha}\beta$ is introduced to production rate, with $f_{\beta\alpha}$ being a new model parameter (positive constant). This model meets the biological context considering the source of tnf- α . The proposed model is written as Equation 3.5.

[further comparison of model 3, 4 and 5]

Model 4 and 5 are base on model 3, so in this phase experiments of these three models are conducted for separately parameter inference and overall model comparison. Based on our experience in conducted experiments, we used the following setting for parameter inference:

- population size: 2000 and 5000
- number of populations: 30 generations
- prior: [0, 50] log-uniform
- perturbation kernel: multivariate normal kernel

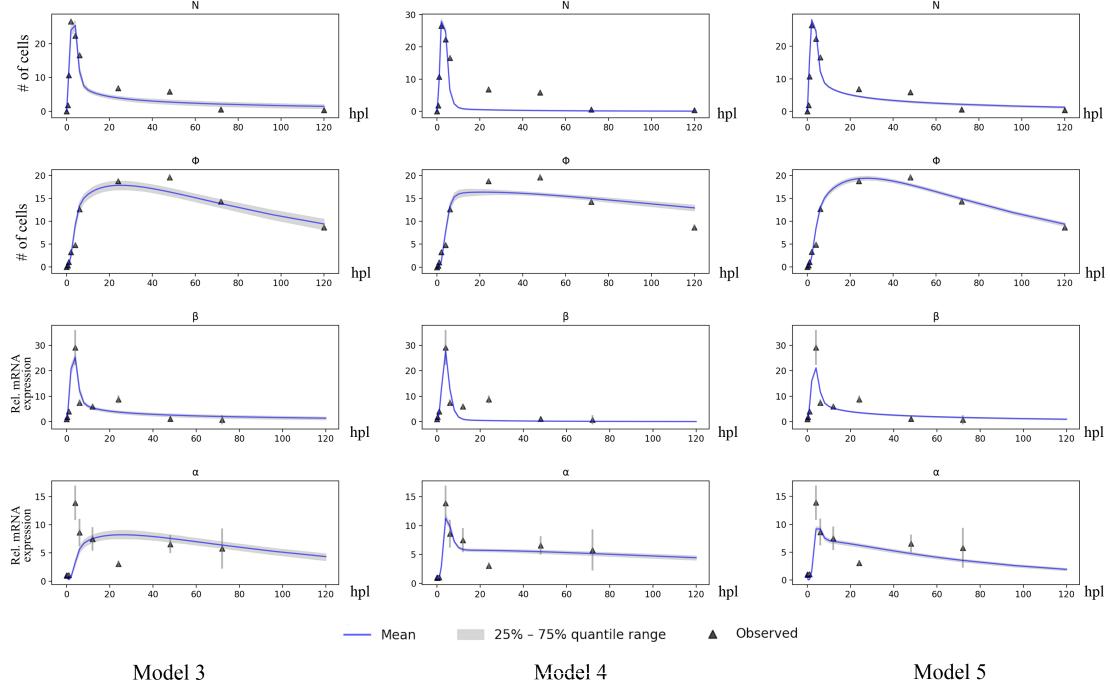


Figure 4.11: Simulated data from the last population of model 3, 4 and 5. Error bar indicates SEM (for N and Φ the SEM is relatively small)

Factors (Equation 4.3) are also tried to find if we can force the inference framework to give more importance on the first few data points. In this experiment, for each trajectory first 8 data points are assigned higher factors (0.75), and the rest 4 data points are assigned with lower factors (0.25). All these runs are conducted on remote machines and the output database files are retrieved from analysis.

Figure 4.11 shows the simulated data from the inferred models, with population size 2000. It can be seen that the addressed problem in the trajectory of $\text{tnf-}\alpha$ is partially relieved in model 4 and 5: compared to model 3, a peak appears around 4 hpl; it can represent more features in the observed data and consequently the final reached epsilon (after 30 populations) is smaller than that of model 3, which proves that our proposed modifications in model 4 and 5 are helpful in recovering more features in observed data.

By intuition model 4 and 5 would be the best models so far that can recover most of observed data features. Model 1, 2 and 3 can also converge to a low final epsilon value (20), but a key feature observed in target data – a peak at around 4 hpl – is not fitted. Some features, e.g. fluctuations in $\text{il-1}\beta$ and tnf-alpha trajectories at around 20 hpl are still not ideally fitted; all existing models give a steady or gradually decrease trends for all the four variables after 20 hpl, and fluctuations in the observed data are fitted by smooth and steady curves.

After applying factors, the results were not largely different. As expected, applying

factors made the data points at first half (where exists most of the fluctuations) more ‘focused’, however the inferred models are close to one without factors, which could indicate that for model 4 and 5, applying factors will not largely affect the final convergency and the resultant posterior.

When we are trying more generations with factors, an ‘over-fitting’ behaviour was observed at the trajectory of macrophage (Figure 4.12). In this case, the discrepancy of simulated data and observed data mainly comes from the last 2 data point (72 hpl and 120 hpl) of macrophage. As additional importance was assigned the the first few data points of each trajectory, the inference algorithm tended to preferentially fit first data points and thus relative large discrepancy at the tails were acceptable.

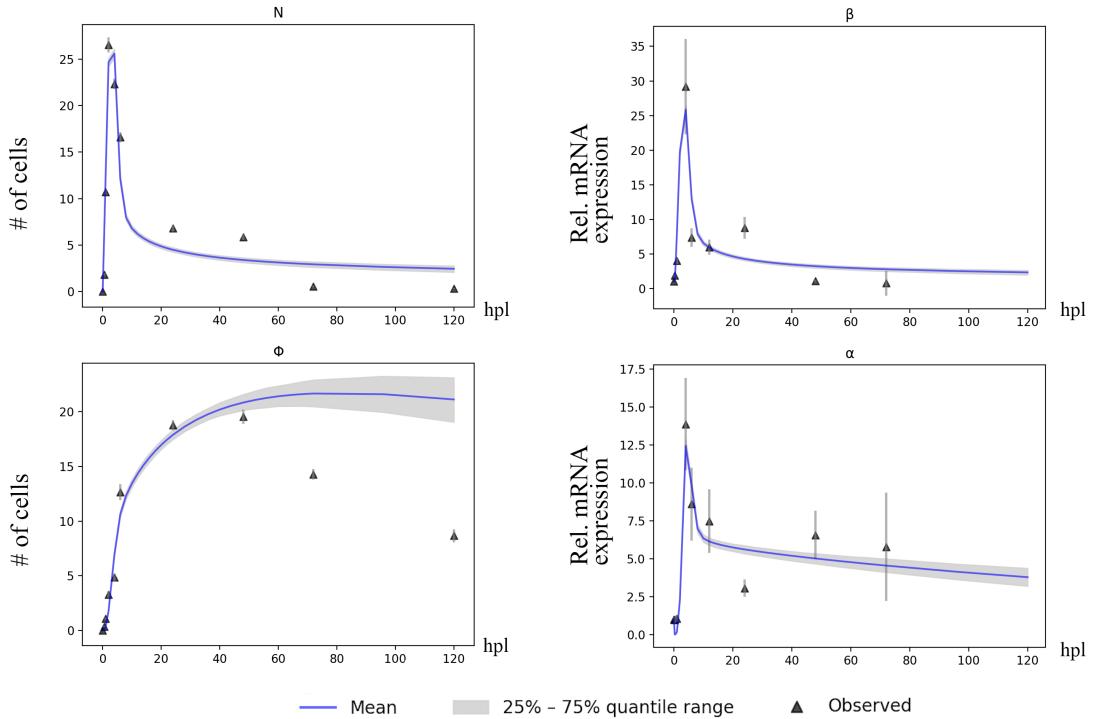


Figure 4.12: Simulated trajectory from the 30th population of model 5, with 25:75 factor applied. Error bar indicates SEM (for N and Φ the SEM is relatively small).

Next, we performed a model comparison on model 3, 4 and 5, using the same ABC SMC inference framework. Figure 4.13 shows the model probabilities at different generations.

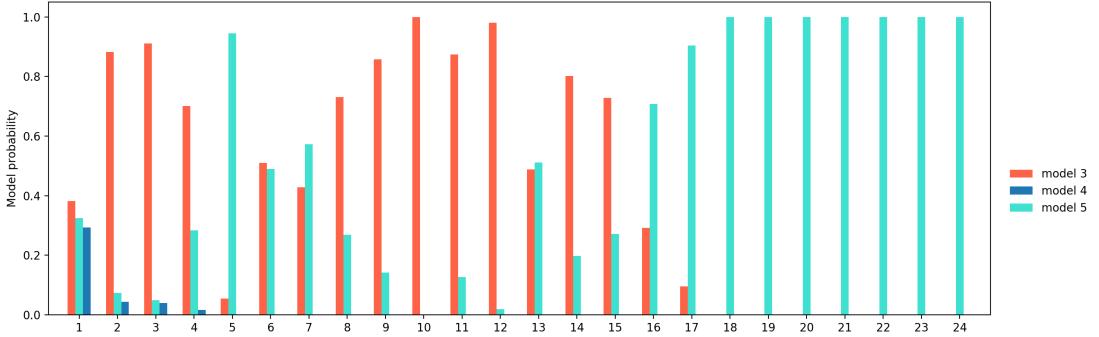


Figure 4.13: Model probabilities of model 3, 4 and 5 in ABC SMC based model comparison. Model 5 also wins in further populations (population 25 to 30)

Model 4 did not overweight other models across all the generations; model 3 gained advantage before generation 15 except generation 5. After generation 18 ($\epsilon_{18} = 16.98$), model 5 gained absolute advantage and won in further generations. Also, repeated runs showed similar trends: although uncertainties were present before generation 20, model 5 won with nearly 100% probability in later generations where epsilon values were considerably low, i.e. $\epsilon_t < 16$.

Based on the model comparison results, we could conclude that model 5 was the best model among the tested models. Uncertainties were observed in early generations, where similar patterns were also observed in the model comparison of model 1, 2 and 3. We considered the the violently changing model probability in early generations is related to the inference ability at high epsilon threshold values (epsilon) and possibly the insufficient exploration of the parameter space. However as we only focused on the best model with small final threshold ϵ_T , the easiest way is to extended the number of generation until epsilon converges, or increase the population size for a more statistically confident probability.

For the best model (model 5), the inferred parameter posterior distribution of model 5 is shown in Figure 4.14, the estimated mean values is shown in Table 4.2. The pair-wise joint density distribution is shown in Figure B.4. Most of inferred posteriors show a bell-shaped density distribution, indicating that they are gaussian distributed. A thin and concentrated shape indicate a well inferred parameter, e.g. μ_Φ and $\kappa_{\Phi\beta}$. Some distributions, however, are skewed to zero, e.g. μ_N and $s_{\alpha\Phi}$.

By looking at the joint density plot (Figure B.4), some correlated patterns are observed. For example, an obvious linear relationship on the distributions of μ_β and $s_{\beta\Phi}$. This leads us to consider that reducing the number of parameters in the model is possible. Highly correlated parameters can be reduced, e.g. $s_{\beta\Phi}$ can be expressed as $A\mu_\beta + B$ where A and B can be calculated from the joint distribution by fitting a linear regression.

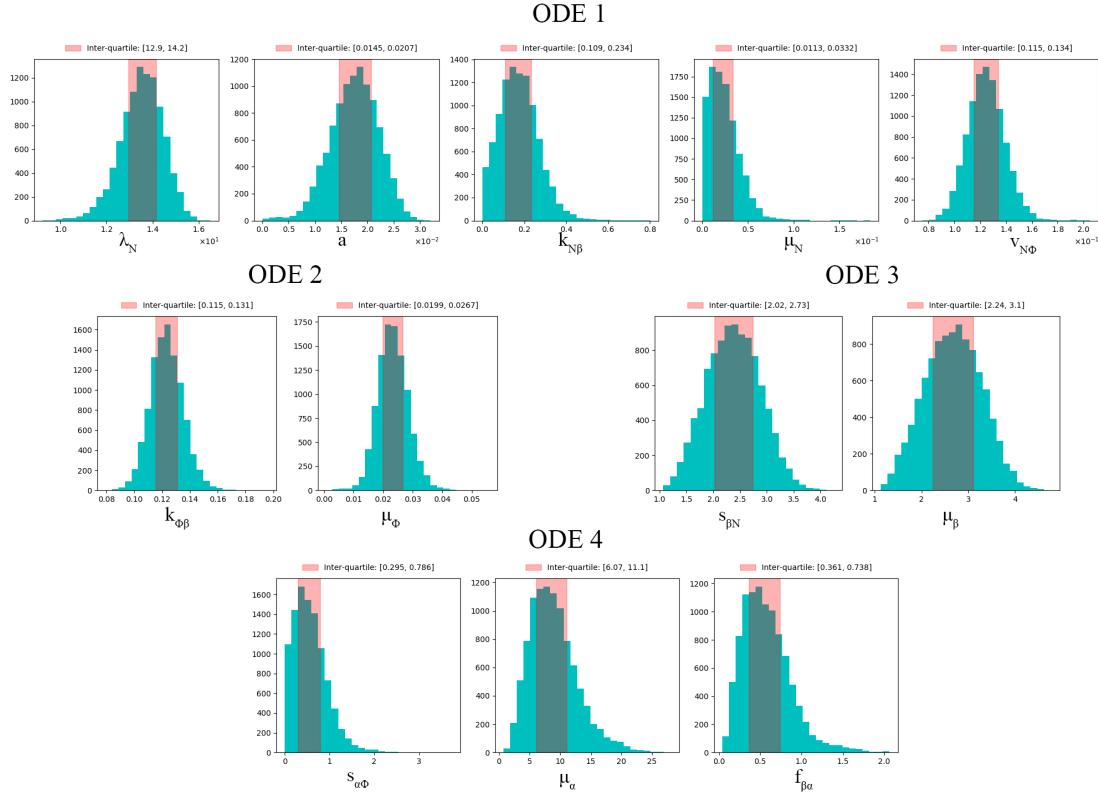


Figure 4.14: Inferred posterior distribution of parameters in model 5. Shaded range indicates 25%–75% quantile of the population

Parameter	Mean value	Inter-quartile interval
λ_N	13.5	[12.9, 14.2]
a	0.0175	[0.0145, 0.0207]
$\kappa_{N\beta}$	0.176	[0.109, 0.234]
μ_N	0.0239	[0.0113, 0.0332]
$\nu_{N\Phi}$	0.125	[0.115, 0.134]
$\kappa_{\Phi\beta}$	0.123	[0.115, 0.131]
μ_Φ	0.0234	[0.0199, 0.0267]
$s_{\beta N}$	2.38	[2.02, 2.73]
μ_β	2.67	[2.24, 3.10]
$s_{\alpha\Phi}$	0.574	[0.295, 0.786]
μ_α	8.91	[6.07, 11.1]
$f_{\beta\alpha}$	0.577	[0.361, 0.738]

Table 4.2: Inferred parameter values of model 5

4.3.3 Sensitivity of parameters

A quick sensitivity analysis of the dynamical systems models can be quantified using principle component analysis (PCA) [8], where the first PCs corresponds to sloppy parameters while the last PCs corresponds to stiff parameters [20]. PCA can only provide rough approximated sensitivity behaviour. By identifying the component loadings, the PCs can be visualised in pie charts and used to indicate the sensitivity.

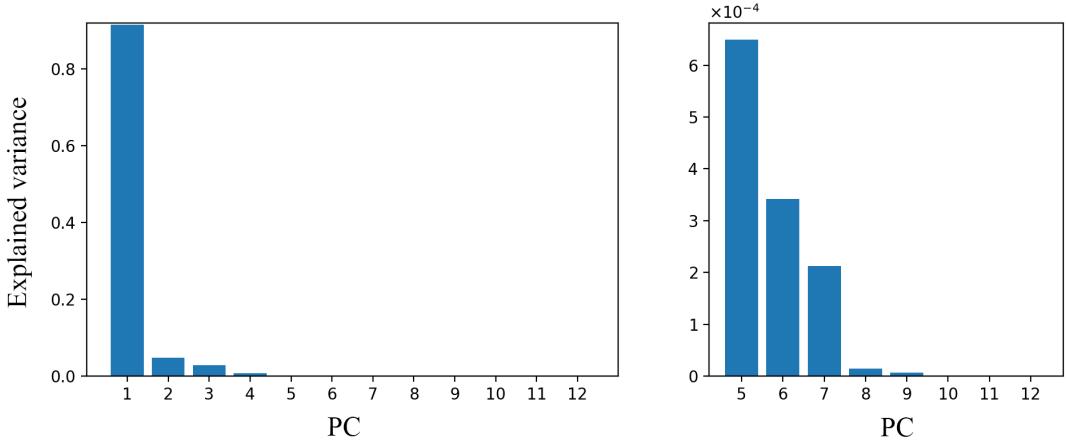


Figure 4.15: Explained total variance in original parameters by each PC. From PC 5 on. little variance is explained by each PC, which is enlarged in Right

As models 5 won in the model selection, a further interest was to explore the parameter-to-model sensitivity of the resultant model. PCA was performed on the last generation of inferred model 5, Figure 4.15 shows the total explained variance by each PC. The first PC explained a major part of the variance (91.5%), and the rest PCs explained less than 5% of the variance each. PC 5 to PC 12 explained much less variance (less than 0.1% of the total variance) than the first four PCs, thus they were considered to extend much narrower regions on the posterior distributions and could be used to identify parameter that the model is sensitive to.

Figure 4.16 shows the fractions of the PCs explained by individual parameters, and a sum of the fractions in the last 8 PCs, i.e. PC 5 to PC 12. The last PC (PC12) is mostly the linear combination of a (from the equation of neutrophil) and μ_Φ (from the equation of macrophage), to which the model is most sensitive. If we conclude the last 8 PCs, it can be found that the biggest portions ($v_{N\Phi}$, a , $\kappa_{N\beta}$, $\kappa_{\Phi\beta}$, μ_N , μ_Φ , $\nu_{N\Phi}$) mostly come from the equation of neutrophil and macrophage i.e. equations of the cells (the un-exploded portion in Figure 4.16 a); if we look at the last 8 PCs individually, the same pattern can be observed. This might suggest that the model is more sensitive to cells' dynamic rather than cytokines' dynamics.

The credible intervals (Figure B.5) of each parameter across generations also agreed with the above conclusion from PCA. The stiff parameters, e.g. a and μ_N , extend

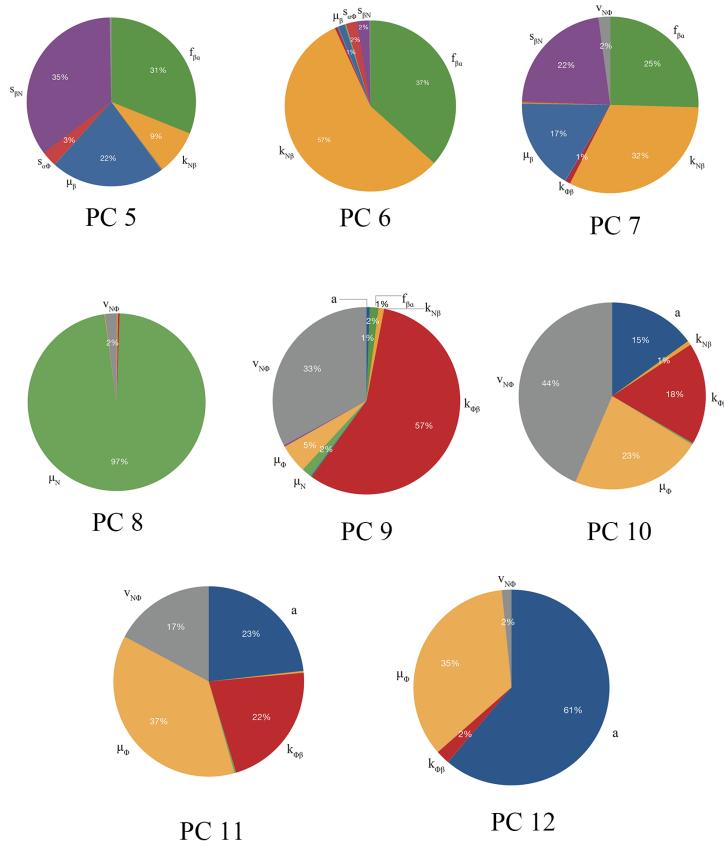
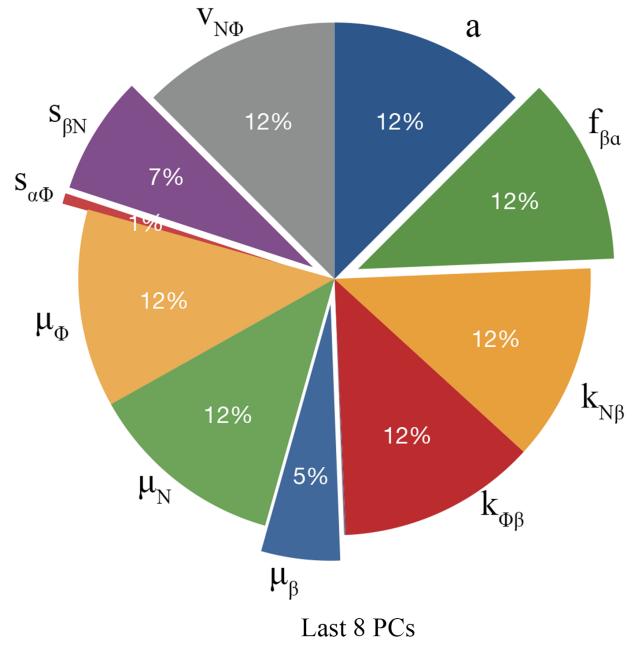


Figure 4.16: Fraction of the length of PCs explained by parameters

narrow credible intervals, while sloppy parameters, e.g. μ_α and $s_{\alpha\Phi}$, have wider credible intervals across the last few generations.

Chapter 5

Performance experiments

The performance experiments were designed to explore the parallel performance of ABC SMC inference framework which were used in the parameter estimation and model selection tasks. Usually ABC SMC is a time-consuming and computation intensive task and ideally executed on large clusters. The scheduling strategy, implementation details and many other factors can affect the parallel efficiency.

5.1 Scaling-up

First experiments are designed to illustrate the scaling-up performance. The program used here is an implementation of ABC SMC on model 5. The details of the ABC SMC settings is listed below

- Prior distribution: default to log-uniform distribution $[1 \times 10^{-6}, 50]$ for all the 12 parameters
- threshold schedule: median epsilon
- No factors, no adaptive distance or adaptive population applied
- Population size is 2000, with 20 generations

For HPC systems like Cirrus, `pyabc` uses multiprocessing for multi-core parallel sampling. By default if the number of cores is not specified, it will automatically read the number of available cores and use them all. Cirrus has a 36-core CPU which support hyperthreading, such that the maximal number of cores available to multiprocessing is 72.

The program is executed on Cirrus, using 8, 16, 24, 36, 54 and 72 cores respectively. Each run is repeated 10 times. The average execution time, required sampling numbers are recorded. Hyperthreading is enabled when using 54 and 72 cores. The access to the node that contains computation cores is exclusive, such that the execution would not be affected by other programs of operations.

The implementation of ABC SMC in `pyabc` enables the parallelisation of sampling, which is the most time-consuming part. The rest part of the program is mostly not parallelised, e.g. database I/O and reductions operations. The sampling process involves sampling, perturbation and test of the acceptance criteria, all of which are computation-intensive.

In practice, using ABC SMC to estimate the parameters of a given model could cost up to several hundred of hours if the computational resources is limited[REF]. The performance experiment result could provide a reference that illustrate that how the efficiency changes when scaling-up or the trade-offs in computational resources' cost and their benefit.

Results

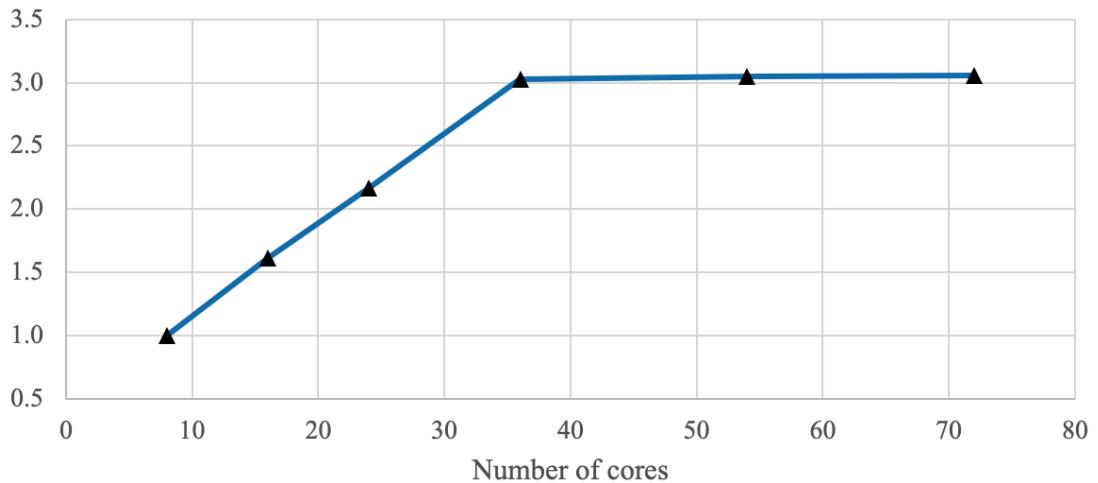


Figure 5.1: Speedup of the program

The recorded execution time using different number of cores were used to calculate the speedup of the program (Figure 5.1). Here 8 cores is regarded as the baseline for speedup and efficiency, as the serial version (using only one core) can take quite a long time to finish 10 repeats. When increasing number of cores, the execution time drops fast at first but the decreasing rate (dropping speed, i.e. the derivative of the execution time against time) is gradually reduced to zero; 36, 54 and 72 cores gives nearly the same average execution time and consequently gains very close speedup values (Table 5.1).

Number of cores	Time (second)	Avg. required samples	Speedup	Efficiency
8	6617	445406.8	1	1
16	3926	590221.0	1.61	0.807
24	2921	573339.8	2.17	0.723
36	2096	446948.7	3.02	0.672
54	2076	640404.9	3.05	0.452
72	2071	768096.0	3.06	0.340

Table 5.1: Performance data

Denote the number of cores used in the ABC SMC runs as variable p . The speedup curve shows a linear trend when $p \leq 36$; when $p \geq 36$ the speedup stays nearly unchanged. $p = 36$ is an inflection point, where 36 is the maximum numbers available physical cores in a compute node of Cirrus. A constant drop of efficiency is observed when increasing p ; smaller p ($p < 36$) gives an efficiency higher than 65% but greater p e.g. $p = 72$ results in a low efficiency (34%).

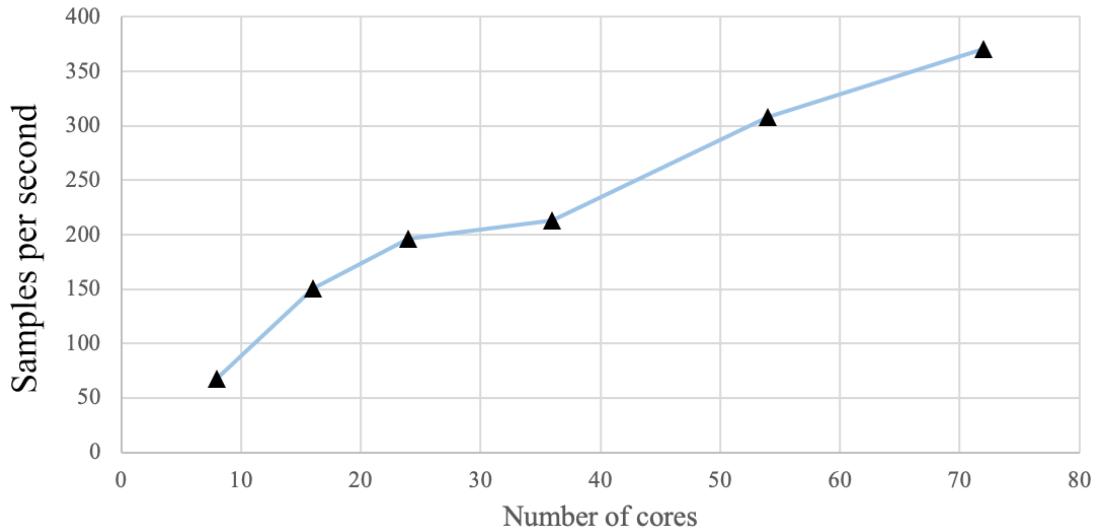


Figure 5.2: Average samples in one second, under different number of cores

5.2 Discussions

A high variance of total required samples was observed in the scaling-up experiments: some single runs required much more samples to finish 20 populations, where they were supposed to have relatively similar average required samples such that we can compare the performance directly via speedup or efficiency.

Due to the variant required samples, our interests switched to the per-second performance, where the average sampling numbers per second under different of number of

cores (p) are plotted (Figure 5.2). The plot shows that even theoretical speedup stops growing when $p > 36$, the average sampling numbers per second performance still get improved when using more than 36 cores. It proves that the implementation used here scales reasonably well without encountering bottleneck when using multi-core parallelisation inside one physical node.

The variant total required samples for different runs under the same setting was most likely the result of different threshold evolve path. Median epsilon schedule and randomness in the sampling process resulted in different sequences of threshold for different runs, which means different repeated runs may have different convergency path [12]. For example, Some runs temporarily converges to a local optimum, while the target epsilon in the next generation cannot possibly be reached if the new particles are still concentrated in the local optimum. In this case, huge amount of sample will be drawn to find enough accepted particles, and new generation will be able to jump out of the local optimum and have different concentrations.

To illustrate this, Two independent run, namely Run A and Run B, were chosen from the experiment of $p = 24$ and visualised in Figure 5.3 and Figure 5.4. These two run used identical options of the algorithm. Figure 5.3 (a) shows the comparison of the required samples, where Run A converged to a local mode, and in population 11 took much more sample tries to move away from the local optimum (Figure 5.3 (b)). Figure 5.4 compares the posterior density distribution before ($t=9$, cyan) and after ($t=13$, red) the jumping away: significant movement of concentrations in posterior density distributions are observed, especially in parameters from ODE 1, 2 and 4. An interesting movement is that there are two concentrations in $\kappa_{\Phi\beta}$ and μ_{Φ} when $t=9$, and the majority of particles are concentrated in the second concentration with high parameter values; when $t=13$, the second concentration distribution disappears and all particles are concentrated near zero.

Moreover, we found that the execution time was not ideally proportional to the sampling tries in that period of time, for both different repeated runs and different generations in a single run. From the same example, Figure 5.3 (b) and (c) show the non-linear relationship between execution time and sample tries in different generations. Apart from 1st and 2nd generation, generation 3 to 20 show a general linear relationship i.e. $T_t \propto n_t$, where t is the population index, T_t is the execution time of generation t and n_t is the required samples in generation t . However, this relationship does not hold for the first two generation. Generation 1 is to draw sample from prior, but it is done with high efficiency as only small amount of time are taken; generation 2 takes unusual large amount of the, but it does not require much more samples. The abnormal execution time of generation 1 and two are observed in all other runs, where for $p = 24$ generation 1 usually takes c. 16s and generation 2 takes c. 1600s. This is believed to be related to the initialisation of the algorithm.

Additionally, by observing CPU frequency and load curve in-time while running the program, we found that the parallelised part is limited the the sampling, calculation and criterion test process; some reduction calculations and preparation work for the next generation (e.g. calculation of the new epsilon values, normalisation of data, fit of

multivariate kernel) are serial. Database I/O is also serial.

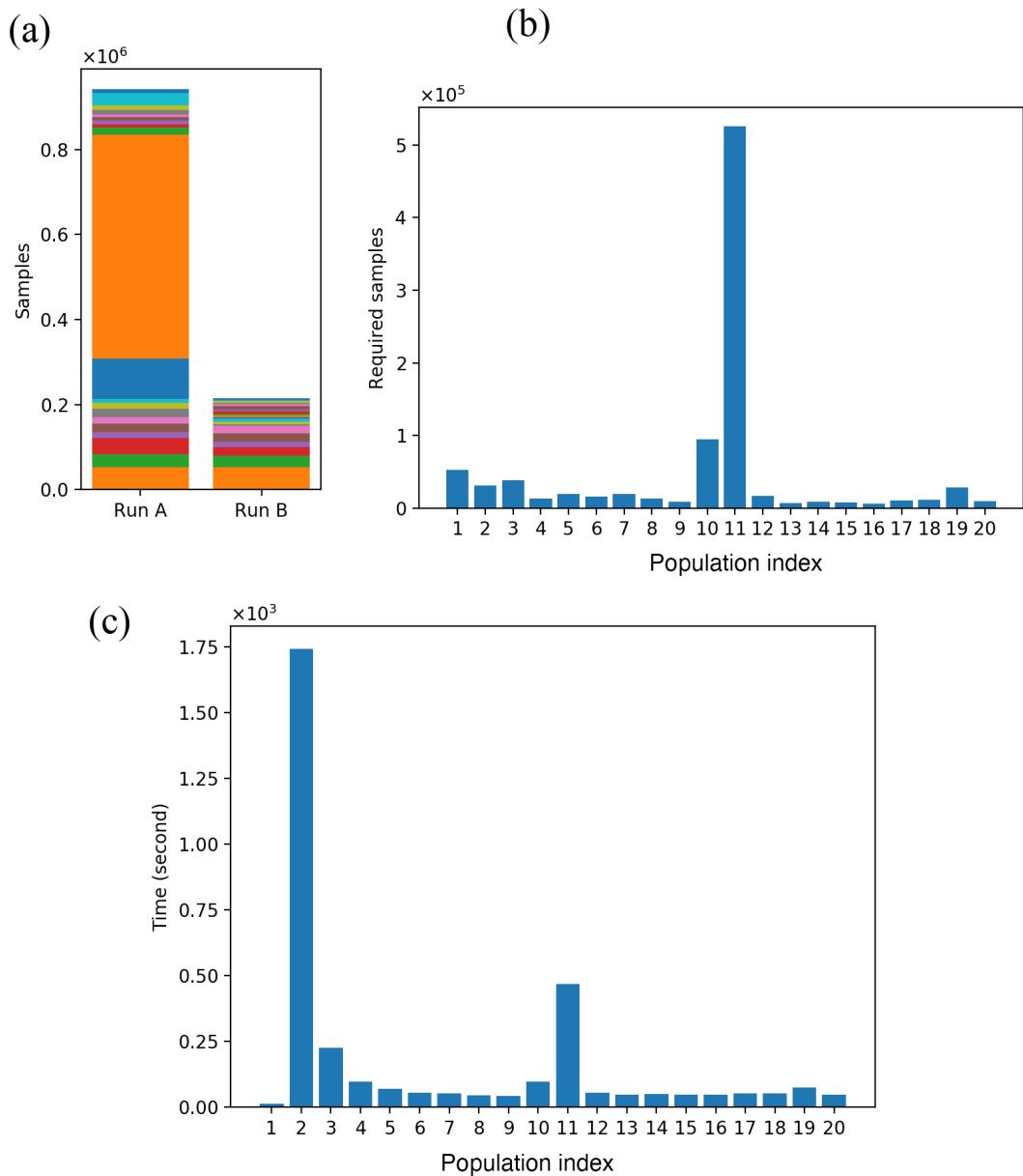


Figure 5.3: (a) Required sampling size for two example runs. Different colour represents different generation (from bottom to top: generation 1 to generation 20). Run A took huge amount of time in population 11. (b) Required sample in each generation of Run A. (c) Time-consumption in each generation of Run A

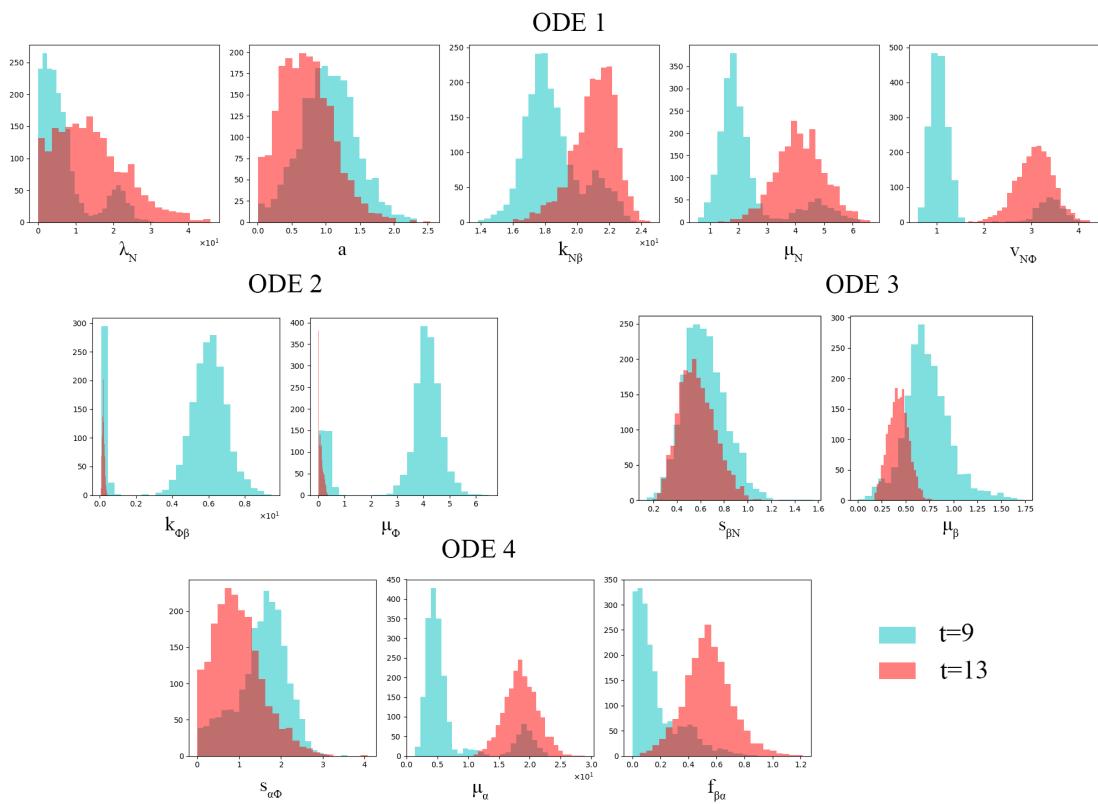


Figure 5.4: Parameter density distribution before (generation 9, cyan) and after (generation 13, red) the concentration movement in generation 11

Chapter 6

Conclusion and future works

To conclude, ABC SMC was successfully applied on the models of zebrafish spinal cord regeneration. These models were all ODE equations that described the interactions and effects between cells and proteins in the lesion site. These equations contains four dependent variable and more than 10 model parameters. Parameter estimation for high-dimensional parameter space is a challenging work for many dynamic system models, as insufficient explore of parameter space can easily lead to local optimal. Using a large population size in ABC SMC in our models could partially relief the concerns in local modes, and give considerable parameter estimates (FIGURE). Also ABC SMC successfully helped in the model comparison and thus can be used for suggestions of possibly best model, although there were uncertainties remained in the process and a large number of population size was need.

ABC SMC implementation is a complex process with respect to both algorithm settings and code development. The ability of parameter inference and robustness of algorithm were firstly studied using synthetic data where the true parameter value was known, and experiments were performed to find proper hyperparameter and implementation settings of ABC SMC. Later inference on the observed data followed the suggestions from the synthetic data experiments and acceptable results were produced.

The inference program was mostly run on multi-core machines, and additional scaling-up performance test illustrated the scaling behaviour. When the number of running Python process exceeds the number of physical cores, hyperthreading would be enabled but gains not significant improvements with respect to execution time, while increase of average sampling size per second was still observed.

From this application, some suggestions on ABC SMC implementation could be drawn. How to set the prior could be the most tricky decision for a high dimensional parameter inference. The distribution and the interval of each parameter can affect the inference results significantly and improper prior cold lead to poor fit. The epsilon trajectory and joint distribution of parameters can be used to identity possible local modes; if stuck, it takes much more samples than usual to jump out out local optimal. To avoid this a large population size and a proper epsilon schedule may help for the uncertainties

observed in model selection and some duplicated runs with high variance, a longer run with more generation could help to identify whether the convergency is consistent such that a reproducible inferred results can be obtained.

In terms of the future works, some more comparison with other algorisms could be of our interest. Least square fitting using MCMC [6] and some other exact inference approaches could be performed on the same model and observed data, as the standard deviation at each measurement point is known. By comparing to these methods, differences in the inference results, algorithm robustness and required computational resources could be revealed and possibly help us in choosing an inference algorism that can balance the trade-off between efficiency and a reasonable good ‘fit’. Also there other implementation options of ABC SMC worth trying, e.g. CUDA acceleration and *Julia* implementations^{1 2}.

In the aspect of models, some further models can be studied. If there were some more experimental evidence on the proposed interaction map, we could proposed more precise models, or calibrate some terms in our existing models. Also, models of other forms, e.g. PDE and stochastic models, could also be studied if a more complex model is needed for the dynamic system. Simplification is also possible of existing models. Some highly-correlated parameters could be reduced, as there exist some linear relationships between parameters (FIGURE).

¹<https://github.com/tanhevg/GpABC.jl>

²<https://github.com/marcjwilliams1/ApproxBayes.jl>

Appendix A

System and software environment

A.1 ABC SMC implementation

A.1.1 Local machine

Local development machine is a Mac laptop, running on macOS 10.15.6. The environment of the development is listed in Table A.1.

Environment	Version
Operating system	macOS Catalina 10.15.6
PyCharm (IDE)	Professional 202.2
Python interpreter	3.7.0
IPython	7.12.0
Clang	6.0 (clang-600.0.57)

Table A.1: Environment on local machine

Version requirements for some critical site packages (Python) used in the code are listed in Table A.2.

Environment	Version
pyABC	$\simeq 0.10.3$
NumPy	$\simeq 1.18.4$
SciPy	$\simeq 1.4.1$
pandas	$\simeq 1.1.0$
matplotlib	$\simeq 3.0.1$
bokeh	$= 1.4.0$

Table A.2: Site packages on local machine

Hardware	Detail
CPU	Quad-Core Intel Core i5 8259U 2.3 GHz
Architecture	64 Bit
Hyper-Threading	Supported
Turbo Boost	Max 3.8 GHz
Memory	16 GB
Graphics card	Iris Plus Graphics 655

Table A.3: Hardwares on local machine

Local development and some runs of small size were done under the hardware listed in Table A.3. When running ABC SMC, the program can make used of 8 python process (Hyper-Threading enabled) and run under 3.8GHz (maximal) for a long time under proper cooling, proved that the program could efficiently use the computation resources for a local personal computer. It is noted that the program took no advantage of graphics card, as multiprocessing along can only exploit the resources inside CPU. The execution time and other performance data presented in Chapter 5 were obtained using remote machines but not local machine.

A.1.2 Remote machine

Cirrus hardware can be found in <https://www.cirrus.ac.uk/about/hardware.html>. The software environment used in the development and experiments are listed in Table A.4. Requirement of site packages is the same as that on local machine (Table A.2)

Environment	Version
Operating system	Red Hat Enterprise Linux 8.1
miniconda environment	4.8.3
Python interpreter	3.7.7
gcc	6.3.0

Table A.4: Environment on local machine

Appendix B

Data and supplementary figures

B.1 Infer-back experiments

B.1.1 Parameter values used to generate synthetic data

Parameter	Value
λ_N	2.20
$\kappa_{N\beta}$	3.96
μ_N	1.72
$\nu_{N\Phi}$	0.219
λ_Φ	1.31
$\kappa_{\Phi\beta}$	0.124
μ_Φ	0.145
$s_{\beta N}$	6.55
$i_{\beta\Phi}$	1.71
μ_β	0.521
$s_{\alpha\Phi}$	10.2
μ_α	19.7

Table B.1: Parameter values used to generate synthetic data for model 1. They were obtained from a preliminarily least-square fitting

B.1.2 Kernel experiment: median epsilon schedule

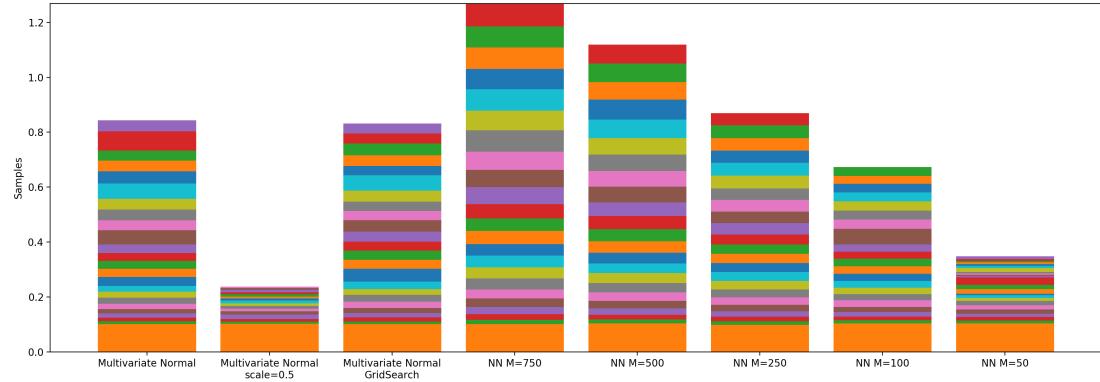


Figure B.1: Total sampling size of different kernels, using median epsilon strategy. Different color represents different generations (bottom to top: population 1 to population 20)

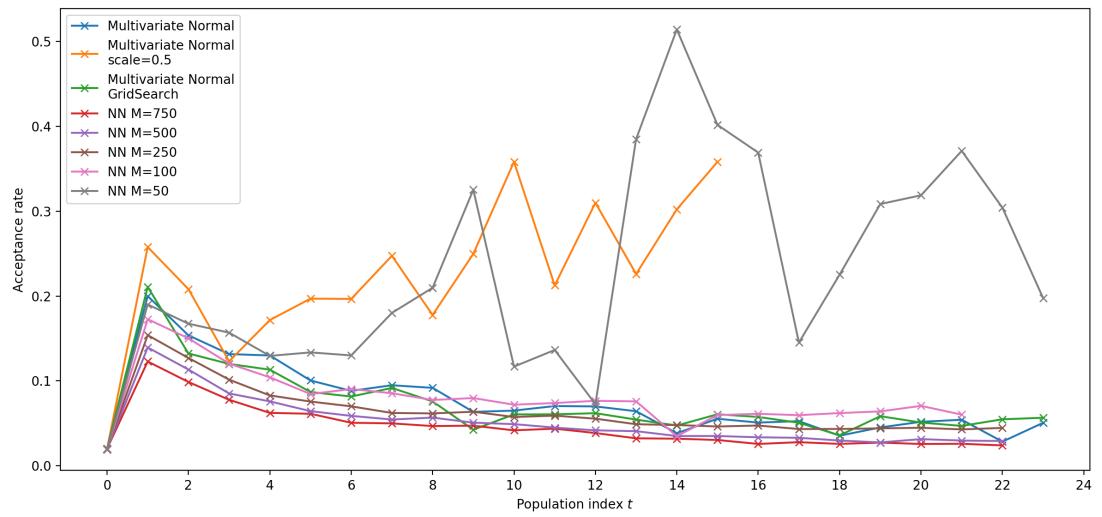


Figure B.2: Acceptance rates of different kernels, using median epsilon strategy. Each population has 2000 particles. The total length of populations is different because some runs reached the final ϵ_T earlier than others

B.2 Parameter inference and model comparison results

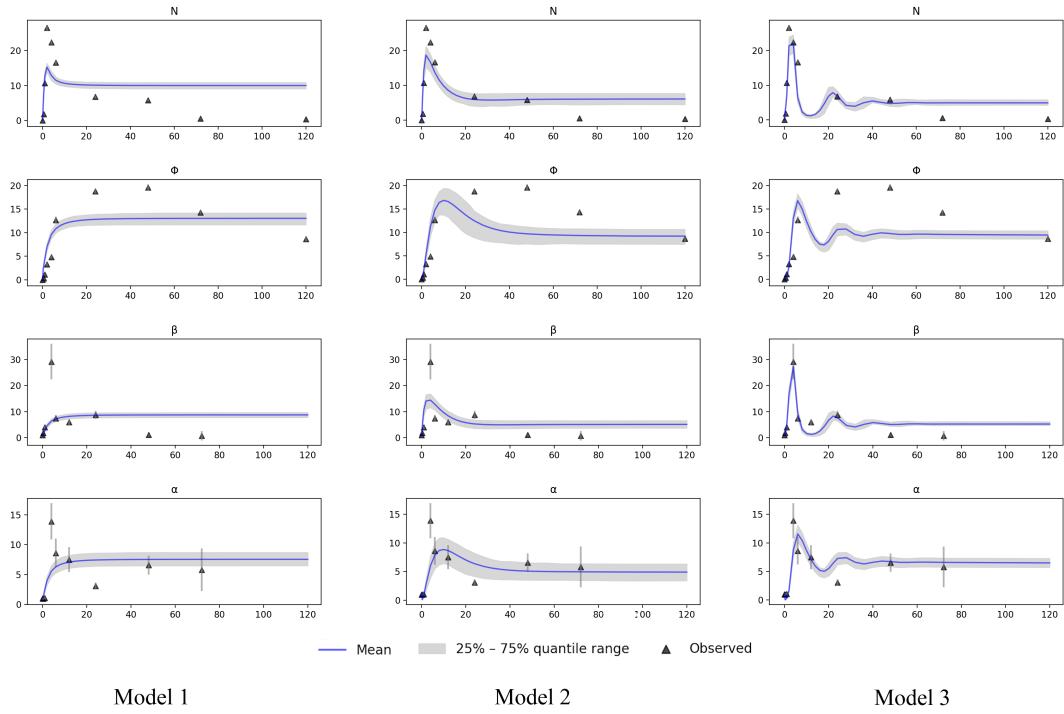


Figure B.3: Simulated data from the last population of model 1, 2 and 3, using uniform prior. Error bar indicates SEM (for N and Φ the SEM is relatively small)

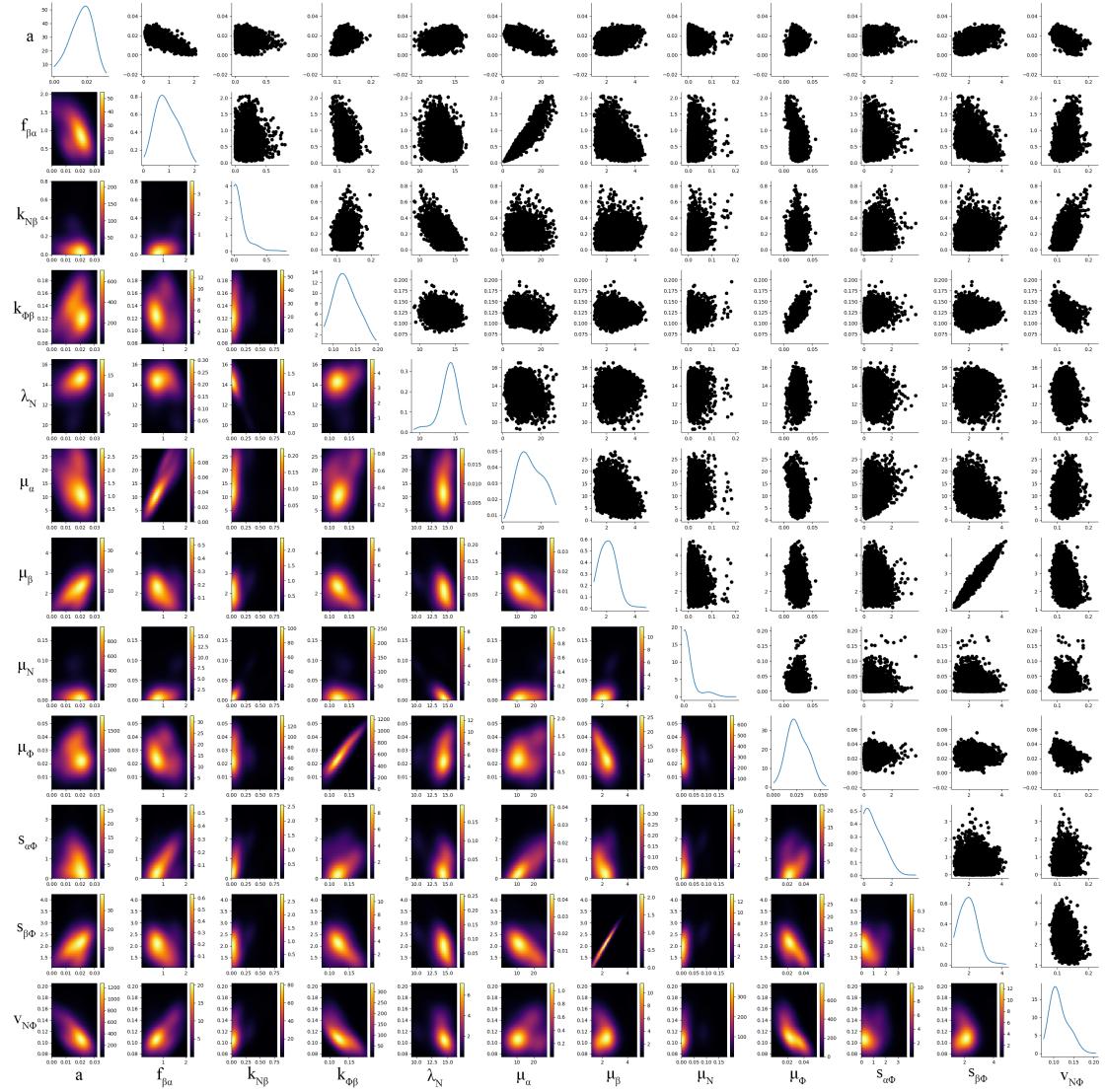


Figure B.4: Joint distribution of each parameter pair. The diagonal is KDE fit of each individual parameter, the upper triangle contains scatter plot of the particles and the lower triangle is the joint density distribution for each parameter pair

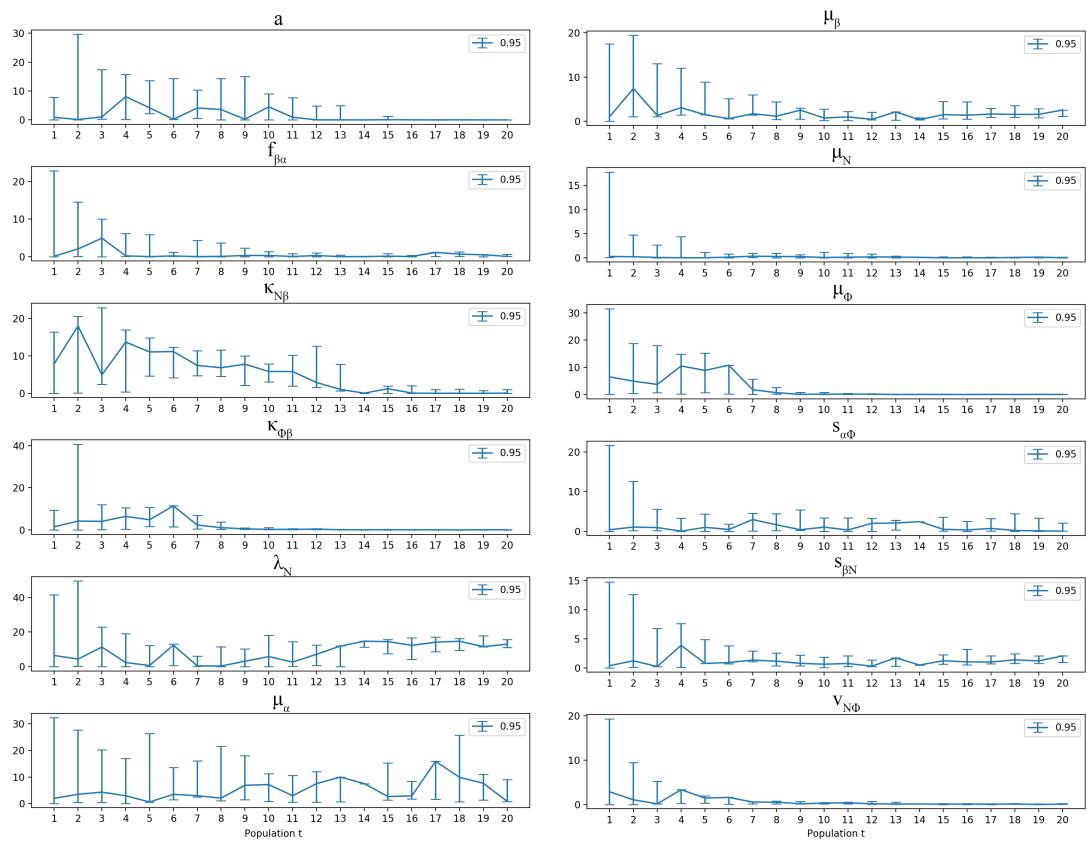


Figure B.5: Credible intervals of parameters in model 5 across all generations

Bibliography

- [1] T. M. Tsarouchas, D. Wehner, L. Cavone, T. Munir, M. Keatinge, M. Lambertus, A. Underhill, T. Barrett, E. Kassapis, N. Ogryzko, *et al.*, “Dynamic control of proinflammatory cytokines $il-1\beta$ and $tnf-\alpha$ by macrophages in zebrafish spinal cord regeneration,” *Nature communications*, vol. 9, no. 1, pp. 1–17, 2018.
- [2] J. Liepe, P. Kirk, S. Filippi, T. Toni, C. P. Barnes, and M. P. Stumpf, “A framework for parameter estimation and model selection from experimental data in systems biology using approximate bayesian computation,” *Nature protocols*, vol. 9, no. 2, pp. 439–456, 2014.
- [3] T. Becker, M. F. Wullimann, C. G. Becker, R. R. Bernhardt, and M. Schachner, “Axonal regrowth after spinal cord transection in adult zebrafish,” *Journal of Comparative Neurology*, vol. 377, no. 4, pp. 577–595, 1997.
- [4] W. D. Anderson, H. K. Makadia, A. D. Greenhalgh, J. S. Schwaber, S. David, and R. Vadigepalli, “Computational modeling of cytokine signaling in microglia,” *Molecular BioSystems*, vol. 11, no. 12, pp. 3332–3346, 2015.
- [5] J. R. Dormand and P. J. Prince, “A family of embedded runge-kutta formulae,” *Journal of computational and applied mathematics*, vol. 6, no. 1, pp. 19–26, 1980.
- [6] W. Gilks, S. Richardson, and D. Spiegelhalter, “Markov chain monte carlo in practice chapman & hall: London,” *Markov Chain Monte Carlo in practice. Chapman and Hall, London.*, 1996.
- [7] S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly, “Inferring coalescence times from dna sequence data,” *Genetics*, vol. 145, no. 2, pp. 505–518, 1997.
- [8] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf, “Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems,” *Journal of the Royal Society Interface*, vol. 6, no. 31, pp. 187–202, 2009.
- [9] P. Joyce and P. Marjoram, “Approximately sufficient statistics and bayesian computation,” *Statistical applications in genetics and molecular biology*, vol. 7, no. 1, 2008.

- [10] M. A. Nunes and D. J. Balding, “On optimal selection of summary statistics for approximate bayesian computation,” *Statistical applications in genetics and molecular biology*, vol. 9, no. 1, 2010.
- [11] A. Minter and R. Retkute, “Approximate bayesian computation for infectious disease modelling,” *Epidemics*, vol. 29, p. 100368, 2019.
- [12] D. Silk, S. Filippi, and M. P. Stumpf, “Optimizing threshold-schedules for approximate bayesian computation sequential monte carlo samplers: applications to molecular systems,” *arXiv preprint arXiv:1210.3296*, 2012.
- [13] E. Klinger and J. Hasenauer, “A scheme for adaptive selection of population sizes in approximate bayesian computation-sequential monte carlo,” in *International Conference on Computational Methods in Systems Biology*, pp. 128–144, Springer, 2017.
- [14] D. Prangle *et al.*, “Adapting the abc distance function,” *Bayesian Analysis*, vol. 12, no. 1, pp. 289–309, 2017.
- [15] S. Filippi, C. P. Barnes, J. Cornebise, and M. P. Stumpf, “On optimality of kernels for approximate bayesian computation using sequential monte carlo,” *Statistical applications in genetics and molecular biology*, vol. 12, no. 1, pp. 87–107, 2013.
- [16] T. Toni and M. P. Stumpf, “Simulation-based model selection for dynamical systems in systems and population biology,” *Bioinformatics*, vol. 26, no. 1, pp. 104–110, 2010.
- [17] A. C. Daly, J. Cooper, D. J. Gavaghan, and C. Holmes, “Comparing two sequential monte carlo samplers for exact and approximate bayesian inference on biological models,” *Journal of The Royal Society Interface*, vol. 14, no. 134, p. 20170340, 2017.
- [18] A. Weyant, C. Schafer, and W. M. Wood-Vasey, “Likelihood-free cosmological inference with type ia supernovae: approximate bayesian computation for a complete treatment of uncertainty,” *The Astrophysical Journal*, vol. 764, no. 2, p. 116, 2013.
- [19] E. Klinger, D. Rickert, and J. Hasenauer, “pyabc: distributed, likelihood-free inference,” *Bioinformatics*, vol. 34, no. 20, pp. 3591–3593, 2018.
- [20] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, “Universally sloppy parameter sensitivities in systems biology models,” *PLoS Comput Biol*, vol. 3, no. 10, p. e189, 2007.