|epcc|

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

Chaolin Han

July 12, 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

List of Tables

List of Figures

Acknowledgements

This template is a slightly modified version of the one developed by Prof. Charles Duncan for MSc students in the Dept. of Meteorology. His acknowledgement follows:

This template has been produced with help from many former students who have shown different ways of doing things. Please make suggestions for further improvements.

Chapter 1

Introduction

This should contain a description of your project and the problem you are trying to solve. Where appropriate you should also include references to work which has already been done on your topic and anything else which lets you set your work in context.

One of the things you will need to do is to ensure that you have a suitable list of references. To do this you should see [1] or some other suitable reference. Note the format of the citation used here is the style favoured in this department. Here is another reference [2] for good measure.

You will also want to make sure you have no spelling or grammatical mistakes. To help idwentify spelling mistukes you caan use the commands *ispell* or *spell*. See the appropriate manual pages. Remember that spelling mistakes are not the only errors which can occur. Spelling checkers will not find errors which are, in fact, valid words such as *there* for *their*, nor will they find repeated words which sometimes occur if your concentration is broken when typing. **There is no substitute for thorough proof reading!**

Chapter 2

Background

2.1 Zebrafish spinal cord regeneration

2.2 Mathematical modelling

2.2.1 Hard bits

You might want to include an equation here:

$$\delta N_\nu = (\delta N_\nu)_{ex} + (\delta N_\nu)_{au} \tag{2.1}$$

2.2.2 Even harder bits

This might be one of the places where you might want to refer to equation 2.1. You will usually need to use the Latex command twice to make cross-references like this work properly. The cross-reference information is stored in the *.aux* file so don't delete it.

Numbering

You can keep subdividing but eventually you get to a level where numbering stops. This text is in a subsubsection which is not numbered by default.

More on numbering: This text is in a paragraph which is also not numbered by default and the "title" of the paragraph is not on a separate line. If you want to increase the depth to which sections are numbered you should see the section on setting the secnumdepth counter in the manual.

2.3 Bayesian inference

2.4 Software tools

Chapter 3

Mathematical modelling

Mathematical models can describe different kinds of dynamic system, and can be used as a guide to prediction and analysis. An ideal model in this case, can represent reasonable interactions/effects between cells and cytokines and recover the observed trend against time.

Models for the regeneration process is in the form of ordinary differential equations, specifically the time differential form. Terms in the ODE are mostly explainable and corresponds interaction paths.

3.1 Observed data

Our models is built regarding the existing experiments data form Tsarouchas et al.[3]. The measurement includes the number of three kinds of cells (neutrophil N, macrophage Φ and microglia) and the relative concentration of four cytokines (il-1β, tnf-α, tgf-β1a and tgf-β3). As proposed in [3], neutrophil and macrophage play important roles in the promotion of spinal cord regeneration with il-1β and tnf-α being the mediation. According to this, our current models focus on the changes of four variables N, Φ, β (il-1β) and α (tnf-α). N and Φ is of the unit 'number of cells', β and α is of the unit 'relative concentration'.

It is noted that the variance of the measured data is relatively high. The summary statistic used for the parameter estimations is mean of measurement, assuming that measured data is Gaussian-like distributed. The distribution of the measured data points is plotted and examined to see if the measurement mean could represent the distribution. The result is that at most time points the measurement values are Gaussian-distributed, some distributions are skewed. One abnormal distribution is observed at time point 120 h post-lesion (hpl) for macrophage where there are two concentrations. Mean value can summarise most data measures and thus is still used as the target summarised observed data. A plot of the mean of the four variables is shown in Figure 3.1

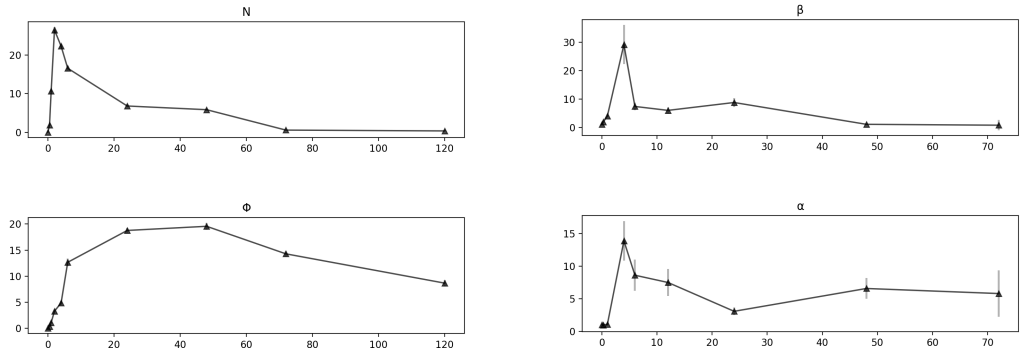

Figure 3.1: Mean of the observed data for neutrophil N, macrophage Φ, il-1β and tnf-α, from experiment results of [3]. Error bars indicate standard error of mean

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 be more close to the true process, recover more features of the observed data or test our hypothesis.

All these models assumes the interactions is within two kinds of cells (neutrophil and macrophage) and two kinds of cytokines (il-1β and tnf-α) and use the data presented in Figure 3.1 for parameter estimation. Interactions or effects from other cells or cytokines is not considered as there might not be corresponding data.

3.2.1 Basic model

A preliminary model is proposed according to [3] and used to build and test the code. A interaction map illustrate the model is shown in Figure 3.2. This model is a simplification a the process described in [3]. To describe the parameters' units, we denote the unit of N and Φ i.e. number of cells as 'cell', and denote the unit of β and $\aleph$ i.e. relative mRNA expression as 'unit' for simplicity.

It is assumed that there are negative feedbacks for all the variables. MORE DESCRIBE.

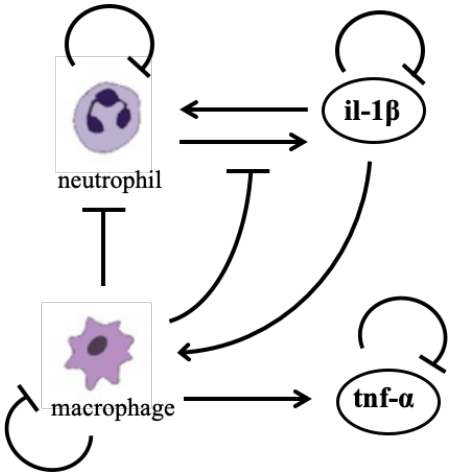

Figure 3.2: Interactions modelled in the basic model (model 1) based on Tsarouchas et al.[3]. Lines ended with arrow represent promoting effect, lines ended with T-connectors represent inhibition

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

Parameter	Definition	Units
λ_N	Self-increase rate of neutrophil	$cell/h$
$\kappa_{N\beta}$	Promoting effect coefficient by il-1β	$cell/(unit\cdot h)$
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-1β	$cell/(unit\cdot h)$
μ_Φ	Coefficient of negative feedback of Φ	h^{-1}
$s_{\beta N}$	Production rate from N	$unit/(cell\cdot h)$
$i_{\beta\Phi}$	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	$unit/(cell\cdot h)$
μ_α	Coefficient of negative feedback of α	h^{-1}

Table 3.1: Parameters introduced in the basic model (model 1)

3.2.2 Alternative models

Model 2 and model 3 As the observed data indicates, this dynamic system has a steady state where the inflammation is resolved and immune cells should not be present at the injury site. Regarding this, the self-increase term λ cannot be constant, thus a exponentially decay λ term is introduced and model 2 is proposed as Eqn. 3.2. The inhibition of il-1β being produced by neutrophil, i.e. $i_{\beta\Phi}$ is considered to be ignored, in which case the relative expression of il-1β 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 has one extra parameter a which is a coefficient in the exponentially decay determining the decay speed, with the unit h^{-1}.

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

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

Model 3 can be regarded as a simplification of model 2, as it can be regarded as model 2 with parameter $i_{\beta\Phi} = 0$. For these three models, model 1 is a naive one that is proposed at very first time and used as a 'template' to build and test parameter inference framework. As the implementation is successful, model 2 and 3 is proposed. 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 are close to steady states for neutrophil and macrophage. model 3 is a small simplification of model 2 and theoretically less general than model 2.

Model 4 and model 5 After the first model selection experiment, it is found that some significant features presented in the observed data is not recovered by any of the model.

According to that, attempts are tried to introduce more interactions within the dynamic system considering the both biological and mathematical context. Extra promoting effect to the expression of tnf-α is considered, by either add a phenomenological term (which means the same effect as directly promoting but the underlying mechanism is unclean) or add a term that represents a promoting effect to the production process of tnf-α, namely model 4 (Eqn. 3.4) and model 5 (Eqn. 3.5).

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

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

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 Model evaluation and comparison

Chapter 4

Implementations and Experiments

4.1 Parameter estimation

Here are some results.

4.2 Model comparison

Here are some results.

4.3 Performance experiments

The performance experiments are designed to explore the parallel performance of ABC SMC implementations. Usually ABC SMC is a time-consuming and computation intensive task and usually is executed on large clusters. The scheduling strategy, implementation details and many other factors can affect the parallel efficiency.

4.3.1 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 [1×10^{-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

HOW PYABC parallelise the sampling

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, 12, 16, 24, 36, 54, 72 cores respectively. Each run is repeated 5 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 must 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.

4.3.2 Profiling

The performance could also be analysed given a profiling report. The second experiment profiles the program to reveal the detailed time consumption for each operation and the possible bottleneck, according to which we could find the hot-spot of program and given possible suggestions on improving the performance.

In this case, profile tools `cProfile` and `yappi` is used in PyCharm IDE.

Chapter 5

Results Analysis

5.1 ABC SMC results

5.2 Performance experiments

5.3 Discussions

This is the place to put your conclusions about your work. You can split it into different sections if appropriate. You may want to include a section of future work which could be carried out to continue your research.

Chapter 6

Future works

Chapter 7

Conclusions

Appendix A

Stuff which is too detailed

Appendices should contain all the material which is considered too detailed to be included in the main bod but which is, nevertheless, important enough to be included in the thesis.

Appendix B

Stuff which no-one will read

Some people include in their thesis a lot of detail, particularly computer code, which no-one will ever read. You should be careful that anything like this you include should contain some element of uniqueness which justifies its inclusion.

Bibliography

[1] L.Lamport. *1986 Latex User's Guide and Reference Manual.* Addison Wesley. pp242.

[2] F.Bloggs. *1993 Latex Users do it in Environments* Int. Journal of Silly Findings. pp 23-29.

[3] Tsarouchas, T.M., Wehner, D., Cavone, L., Munir, T., Keatinge, M., Lambertus, M., Underhill, A., Barrett, T., Kassapis, E., Ogryzko, N. and Feng, Y., 2018. Dynamic control of proinflam- matory cytokines Il-1β and Tnf-α by macrophages in zebrafish spinal cord regeneration. Nature communications, 9(1), pp.1-17.