

UNIVERSITY OF SOUTHAMPTON
FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

**Fast approximate Bayesian computation for inference in non-linear
differential equations**

by

Sanmitra Ghosh

Thesis for the degree of Doctor of Philosophy

September 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Doctor of Philosophy

FAST APPROXIMATE BAYESIAN COMPUTATION FOR INFERENCE IN
NON-LINEAR DIFFERENTIAL EQUATIONS

by **Sanmitra Ghosh**

Complex biological systems are often modelled using non-linear differential equations which provide a rich framework for describing the dynamic behaviour of many interacting physical variables representing quantities of biological importance. Approximate Bayesian computation (ABC) using a sequential Monte Carlo (SMC) algorithm is a Bayesian inference methodology that provides a comprehensive platform for parameter estimation, model selection and sensitivity analysis in such non-linear differential equations. However, this method incurs a significant computational cost as it requires explicit numerical integration of differential equations to carry out inference. In this thesis we propose a novel method for circumventing the requirement of explicit integration, within the ABC-SMC algorithm, by using derivatives of Gaussian processes to smooth the observations from which parameters are estimated. We evaluate our methods using synthetic data generated from model biological systems described by ordinary and delay differential equations. Upon comparing the performance of our method to existing ABC techniques, we demonstrate that it produces comparably reliable parameter estimates at a significantly reduced execution time. To put emphasis on the practical applicability of our fast ABC-SMC algorithm we have used it extensively in the task of inverse modelling of a phenomenon pertaining to plant electrophysiology. Particularly we model the electrical responses in higher plants subjected to periods of ozone exposure. We investigate the generation of calcium responses at local sites following a stimulation and model electrical signals as a plant-wide manifestation of such responses. We propose a novel mathematical model that describes the experimentally observed responses to ozone. Furthermore, we pose the modelling task as an inverse problem where much of our insight is gained from the data itself. We highlight throughout the inverse modelling process the usefulness of the proposed fast ABC-SMC method in fitting, discriminating and analysing models described as non-linear ordinary differential equations. We carry out all these tasks using noisy experimental datasets, that provide limited information, to derive novel insights about the underlying biological processes.

Contents

Nomenclature	xvii
Declaration of Authorship	xix
Acknowledgements	xxi
1 Introduction	1
1.1 Motivation	1
1.2 Recent developments in Bayesian inference	2
1.3 Electrophysiological responses in plants	3
1.3.1 Types and natures of electrical signals in plants	4
1.4 Thesis contributions: Increasing the computational efficiency of ABC-SMC	7
1.4.1 Mathematical modelling of electrical responses	7
1.5 Thesis structure	8
1.6 Publication	9
2 ABC methods for learning in differential equation models	11
2.1 Introduction	11
2.2 Bayesian parameter estimation and model selection	11
2.3 The basic ABC framework	13
2.3.1 ABC-SMC for parameter estimation in ODE	14
2.4 Algorithmic settings	15
2.4.1 Tolerance schedule	16
2.4.2 Perturbation kernel	16
2.4.2.1 Multivariate perturbation kernel	17
2.4.2.2 Fisher information to construct kernel	18
2.4.3 Particle population size	19
2.5 Model selection using ABC	20
2.5.1 Model selection approaches in ABC-SMC	21
2.6 Limitations of ABC-SMC	21
2.6.1 Advantages	23
2.7 Conclusion and thesis goals	24
3 Gaussian processes for speeding up the ABC-SMC algorithm	25
3.1 Introduction	25
3.2 Gradient based parameter estimation in differential equations	25
3.3 Gaussian processes	27
3.3.1 Noisy Observations	29

3.3.2	Maximum Likelihood estimation	30
3.4	Gaussian process regression	31
3.5	Derivative of a Gaussian process	31
3.6	ABC-SMC with Derivative GP	32
3.6.1	Illustrations through an example model	33
3.6.2	The GP-ABC-SMC algorithm	34
3.6.3	The GP-ABC-SMC algorithm for DDEs	36
3.6.4	Working with hidden variables	37
3.6.5	Comparison with other GP based algorithms	37
3.7	Conclusion	39
4	Experimental evaluation of the GP-ABC-SMC algorithm	41
4.1	Introduction	41
4.1.1	ODE: The predator prey model	42
4.1.2	DDE: The Hes1 model	45
4.1.3	ABC variability	49
4.1.3.1	ABC variability: Effect of population size	51
4.1.4	Signal transduction cascade	52
4.2	Model selection with GP-ABC-SMC	61
4.3	Chaotic attractor: The Mackey-Glass equation	63
4.4	Computational limitations of GP-ABC-SMC	67
4.5	Conclusion	67
5	Mathematical modelling of plantwide electrical responses in higher plants exposed to ozone	69
5.1	Introduction	69
5.2	Ozone as stimulus: Experiment design and analysis	71
5.2.1	Experimental setup	72
5.2.2	Electrical responses to ozone as stimulus	74
5.3	Phenomenological model of calcium current	81
5.3.1	Cellular current model	81
5.3.2	Macroscopic current model	83
5.3.3	Ligand gated channel	86
5.4	Model fitting to experimental data using GP-ABC-SMC	86
5.4.1	Preprocessing	86
5.4.2	Prior distributions	88
5.4.3	Results	91
5.5	Choosing the best model: ABC model selection	100
5.6	Posterior sensitivity analysis	101
5.7	Population approach	105
5.8	Conclusion	106
6	Summary and main contributions	109
6.1	Summary	109
6.1.1	Chapter 1	109
6.1.2	Chapter 2	110
6.1.3	Chapter 3	110

6.1.4	Chapter 4	110
6.1.5	Chapter 5	111
6.2	Future work	112
6.2.1	Feedback between GP and ODE parameters	112
6.2.2	Geometric tolerance schedule	113
6.2.3	Partially observed systems	115
6.2.4	Further work regarding the modelling of plant electrical responses	116
A	Multivariate Gaussian Identities	119
A.1	Schur complement	119
A.2	Multivariate Gaussian	120
B	ABC-SMC	123
C	Fisher information matrix for ODEs	127
C.1	Evaluating the Fisher information matrix for ODEs	127
C.2	Approximation of the Hessian matrix for ODEs	128
	References	131

List of Figures

1.1	Typical shape of the action potential.	5
1.2	VP in sunflower plant Stankovic et al. (1997) triggered by flaming is measured by three electrodes placed 5, 15 and 25 cm from the stimulated site on the stem.	6
1.3	Defence responses induced by electrical signalling (Christmann and Grill, 2013). Transmission of electrical signals to local and distant sites after herbivore attack trigger defence response as jasmonate hormone increases. These electrical signals are generated by the glutamate-receptor-like (GLR) ion channels.	6
2.1	In the ABC-SMC algorithm a sample from the prior distribution is passed through filtering steps where the probability of the particles to represent the data best is updated during the algorithm via intermediate distributions. The final distribution is an approximation of the true posterior distribution. The SMC algorithm follows a particle-filtering approach. . .	16
3.1	Draws from a GP with a squared exponential kernel with hyperparameters σ_{ker}^2 and l^2 set to 1	28
3.2	Noisy state trajectory of the Mackey-Glass model.	34
3.3	The estimated smoothed state trajectory of the Mackey-Glass model. . . .	34
3.4	The estimated velocity field of the Mackey-Glass model.	35
4.1	Histograms of the final particle populations generated by the GP-ABC-OLCM and the ABC-SMC-OLCM algorithms, approximating the marginal posterior distribution for each of the parameters of Lotka Volterra model. The black line marks the ground truth in each plot.	43
4.2	Histograms of the final particle populations generated by the GP-ABC-OLCM and the ABC-SMC-OLCM algorithms, approximating the marginal true posterior distribution for each of the parameters of Hes 1 respectively. Inference based on Dataset 1 is shown here. The black line marks the ground truth in each plot.	48
4.3	The boxplots represent the distributions of the mean and variances (across 50 runs) of the final population representing the marginal approximate posterior parameter distributions learnt by the GP-ABC-SMC and the GP-ABC-OLCM from the three artificial datasets of Lotka Volterra model.	50
4.4	Distributions of the mean and variances learnt by the GP-ABC-SMC and the GP-ABC-OLCM from the three artificial datasets of Hes1 model. . .	51

4.5	Reconstructed and true state trajectories of the Lotka Volterra model. The results corresponding to the three datasets (1, 2, 3) are shown using blue, red and magenta colours respectively. Reconstructed trajectories are represented as curves and observations as stars. The ground truth is the black (dashed and circled) curve.	52
4.6	Reconstructed and true state trajectories of the Hes1 model.	53
4.7	Posterior marginal densities for the Lotka Volterra model parameters for different population sizes used in GP-ABC-SMC. We have used kernel density estimation for plotting these densities. The black vertical lines denote the ground truth and the black thick curve in both plots shows the respective posterior densities for $N = 100$	54
4.8	Results of GP-ABC-OLCM for the signal transduction cascade ($[S]$ – $[R_{pp}]$) in plots (a)–(e) respectively). In all the plots observations are the black stars, the true state trajectory is the red (dashed) curve and the blue curve shows the reconstructed trajectory. We have also plotted the GP mean function as the magenta curve and the 95% confidence region is shown as the shaded area. The reconstructed trajectory is generated by numerically integrating (equation 4.3) with the parameters set to the mean of the posterior distribution estimated by the GP-ABC-OLCM algorithm.	56
4.9	The ESS values indicating number of healthy or alive particles at each of the SMC steps. Except the first SMC step the number of alive particles fall well below 50 particles indicating particle degeneracies when using narrow priors (blue dashed curve) centred on the true values of the parameters. While using a wider prior we see a marked improvement in the ESS above 90% (brown dashed curve).	58
4.10	Posterior distributions shown as histogram plots. The standard deviations <i>std</i> for each parameters' final particle populations are shown on top of each of the plot. Clearly the posterior variances are bigger than those using MCMC in Table 4.	58
4.11	Results of GP-ABC-OLCM for the signal transduction cascade using a wide support for the priors. In all the plots the true state trajectory is the black (dashed) curve and the magenta curve shows the reconstructed trajectory generated by numerically integrating (equation 4.3) with the parameters set to the mean of the posterior distribution estimated by the GP-ABC-OLCM algorithm. The yellow thin curves are reconstructions based on each of the $N = 100$ posterior parameter sample from the last step of GP-ABC-OLCM.	60
4.12	Common-cold data from <i>Tristan da Cunha</i> collected in October 1967 (Toni et al., 2009).	62
4.13	Histograms show populations 1 – 6 of the model particles m for the third run of GP-ABC-SMC. Population 6 represents the final posterior marginal estimate of m . The third model get eliminated from population 3 onwards.	63
4.14	Histograms show the posterior marginal densities of the learned parameters. Posteriors in the top & bottom rows correspond to Model 2 and Model 1 parameters respectively.	64

4.15	Comparison of the actual common-cold data with the reconstructed time courses from Model 1 & 2. The reconstruction is done using mean of the final particle populations generated in the third run of GP-ABC-SMC. In terms of model selection none of these models stand out from the other. Plot (a) shows the time courses of $I(t)$ and (b) shows $R(t)$	65
4.16	Posterior distributions of the parameters of the Mackey-Glass model as histogram of final populations (for each parameter respectively) of GP-ABC-OLCM.	66
4.17	Reconstructed time course of the variable x along with the ground truth trajectory and its noisy version.	66
4.18	Comparison between the original attractor (left) and its reconstruction (right) based on the outputs of GP-ABC-OLCM.	67
5.1	Refractory period of the rapid transient decrease in stomatal conductance as found in experiments reported in Vahisalu et al. (2010). In the first experiment (a) successive ozone expositions have no further effects. However, when successive ozone expositions are separated in time for ~ 90 minutes or more (b) then a second episode of conductance decrease can be observed.	72
5.2	Experimental setup showing a tomato plant inside a plastic transparent box, kept inside a Faraday cage. The placement of the electrodes on the stem is also shown.	74
5.3	Circuit connection diagram for a dual channel instrumentation amplifier-data acquisition system	75
5.4	Typical electrical responses from the tomato plant. The blue and red curves indicate the amplitudes recorded with the 1st (bottom) and 2nd (top) electrodes on the stem. The black dashed curves (vertical) indicates the time location of the application of ozone stimuli.	76
5.5	Electrical responses pertaining to the (estimated) excitation component $\hat{V}_d(t)$ obtained by smoothing is shown as the green curve for the same batch of experiment as shown previously. The blue curve indicates the net response $V(t)$ and the red curve is the smoothed response representing moving average $V_{b_{MA}}(t)$ of the base amplitude $V_b(t)$. All these amplitudes are based on the recording of the 1st (bottom) electrode.	78
5.6	Differential response time course for a batch ozone exposures on a cucumber plant. The black curve shows differential excitation component while the blue curves show the corresponding amplitudes recorded at the two electrodes on the stem. The blues curves here show the net amplitude $\hat{V}_{d_1}(t)$ and $\hat{V}_{d_2}(t)$. The blown up sections show the differential amplitudes after the second and third ozone exposures within that batch.	79
5.7	Typical electrical responses from the tomato plant using a shorter inter exposition time. The blue curve shows the differential amplitude obtained after smoothing. Here the stimulus is repeated after every 15 minutes.	80
5.8	Interactions of ionic currents within a Guard cell after ozone exposition. A sample time course of the cellular calcium current is also shown as a magenta curve.	82
5.9	Hypothetical mechanism behind the net effect of cellular microscopic currents generated at two different locations (where the top two electrodes are inserted) resulting in a macroscopic calcium current along the phloem.	84

5.10	The time series on the left portrays one episode represented using the original sampling rate (10 samples / second) and the plot on the right is its downsampled version.	87
5.11	Histogram plots for each of the model parameters. Histograms are constructed from the collection of point estimates found after minimizing the r.h.s of equation 5.11. In total 54 episodes from different batches of experiments are used.	89
5.12	Reconstructed trajectories of the dynamical variables in equation 5.6. . .	90
5.13	Comparison of smoothed time course $\widehat{\delta \mathbf{I}_{ca}}(t)$ (black curve) achieved using two different GP covariance kernels, the squared exponential kernel in (a) and the Matern kernel in (b) on a representative episode (red curve). . . .	91
5.14	The posterior densities of each of the model parameters as obtained after applying the GP-ABC-SMC on episodes 4 (see Figure 5.16(d)) and 5 (see Figure 5.16(e)) from batch 2 corresponding to cucumber plant. We have used kernel density estimates rather than histogram to ensure that the spread of the support of these densities are clearly visible.	93
5.15	Model fitting results after applying the GP-ABC-SMC algorithm on episodes from batch 1, using the voltage gated model. Experimental data $\delta \hat{V}_d(t)$ is shown as blue curves. The reconstructed time courses of $\delta \mathbf{I}_{ca}(\phi, t)$ are plotted as yellow curves. The smoothed time courses of $\widehat{\delta \mathbf{I}_{ca}}(t)$ as obtained through the GP regression are plotted as the red curves.	95
5.16	Model fitting results after running the GP-ABC-SMC algorithm based on experiments on the cucumber plant, that is batch 2. Notice how the reconstructed curves are more spread around the data in the plots (e-h) indicating a regime change or shrinkage of the electrical responses.	96
5.17	Model fitting results for the ligand gated model based on experiments on the tomato plant, that is batch 1.	98
5.18	Model fitting results for the ligand gated model based on experiments on the cucumber plant, that is batch 2.	99
5.19	The number of particles assigned to each model, the voltage and the ligand gated variant, are plotted as bar graphs. We have plotted these graphs for all the SMC time steps for episode 1 from batch 1 (tomato plant, see Figure 5.15(a)).	101
5.20	Results of the sensitivity analysis are plotted through the principal components and their associated explained variances. The loadings, as the linear parameter combinations, of the 1st (comp 1) and the last (comp 2) principal components are plotted as bar charts in (a) and (b). The height in both direction of the bars indicate the contributions of each of the parameters for both (1st and the last) the components. We have also plotted the (%) of variance explained by all the components in decreasing order.	103
5.21	Results of the sensitivity analysis on episodes from batch 2.	104
5.22	The black thick curve representing the average of all the episodes and most of the coloured thick curves representing the smoothed (through GP regression) time courses of $\widehat{\delta \mathbf{I}_{ca}}(t)$ falls within the ranges of the population of simulated calcium currents plotted with thin magenta curves.	106

6.1	The block diagram on top represents the single-output GP modelling framework, wherein separate GP priors are placed on each state of a coupled ODE. In the MTGP framework, a single prior is placed on all the states through the Kronecker product covariance.	116
6.2	Time course of electrical potential obtained after adding 5ml of 0.05M H_2SO_4 to the soil.	117

List of Tables

4.1	Estimated parameters of the Lotka Volterra predator-prey model denoted by the mean and standard deviation of the particles of the final population for datasets 1-3 respectively in each row.	44
4.2	Run-time and the ratio of total number of particles accepted (θ^{**} in Algorithm 1) to that of generated (θ^*) for the four ABC-SMC algorithms when applied to the three artificial datasets pertaining to the Lotka Volterra predator-prey model (left) and Hes1 model (right). The values for run-time are rounded to nearest integers.	44
4.3	Estimated parameters of the Hes1 loop model.	47
4.4	Estimated parameters of the signal transduction cascade by all the GP based approaches including the GP-ABC-OLCM. The estimates for GP-ODE , AGM and GM are taken from Wang and Barber (2014).	55
4.5	Total number of times m indicates the corresponding models in the final population generated in each run of GP-ABC-SMC. Model 1, 2, and 3 correspond to the ODEs in equation 4.8, 4.9 and 4.10.	62
4.6	Estimated parameters of the Mackey-Glass equations.	64
5.1	Total number of particles assigned to each model as obtained after running the GP-ABC-SMC for model selection.	100

Nomenclature

$\mathcal{M}_\theta$	A generative model
Y^s	Simulated data generated from a model
Y^d	Observed data
ϵ	A tolerance value
Δ	A distance function between the simulated and observed data
$\mathbf{X}(t)$	Trajectory of a differential equation
$\mathbf{V}^d(t)$	Empirical velocity field obtained through smoothing
τ	Algorithmic iteration
$\boldsymbol{\theta}^*$	Importance sampled particle
$\boldsymbol{\theta}^{**}$	Perturbed particle
w	Importance weight
K_τ	A transition kernel that performs a random walk
$\#m$	Frequency of a model $\{\mathcal{M}_\theta^m\}$ for being able to meet the tolerance
$K(x, x'; \phi)$	A covariance function on inputs x and x' , parametrized by ϕ
$\mathbf{K}_{\mathbf{x}\mathbf{x}^*}$	Shorthand notation to denote covariance between input vectors $\mathbf{x}$ and $\mathbf{x}^*$
$\mathcal{GP}$	A Gaussian process
Σ_τ	Covariance of a multivariate perturbation kernel
ESS	Effective sample size
$V_d(t)$	Excitation component of electrical response
$V_b(t)$	Base component of electrical response
$\hat{V}_d(t)$	smoothed excitation component of electrical response
I_{oz}	Ozone current affecting a single cell
I_{ion}	Cellular ionic current
I_{ca}	Cellular calcium current
V_m	Membrane potential
$\mathbf{I}_{ca}$	Average calcium current for a collection of cells
$\delta \mathbf{I}_{ca}$	Difference of calcium currents at two extracellular locations
$I(\boldsymbol{\theta})$	Fisher information matrix
H	Hessian matrix
$\mathbf{I}_{ion}$	Average ionic current for a collection of cells
$\mathbf{I}_{oz}$	Average ozone current for a collection of cells
$\mathbf{V}_m$	Average membrane potential for a collection of cells

V_o	Resting membrane potential
$\ \cdot\ ^2$	Euclidean norm
$O(t)$	Probability of an ion channel being in open state
$C(t)$	Probability of an ion channel being in closed state
$U(t)$	Measured ozone concentration

Declaration of Authorship

I, **Sanmitra Ghosh** , declare that the thesis entitled *Fast approximate Bayesian computation for inference in non-linear differential equations* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: as listed in section [1.6](#)

Signed:.....

Date:.....

Acknowledgements

First and foremost I would like to thank my supervisors Dr. Srinandan Dasmahapatra and Prof. Koushik Maharatna for their guidance and support and for the time they have dedicated towards supervising me over the course of my doctoral studies. In particular I want to thank Dr. Dasmahapatra for all his help, patience, and unwavering optimism. Thanks to Prof. Maharatna for his enthusiasm, encouragement and insights throughout. I want to express my deepest gratitude to my supervisors for introducing me to the quest for scientific truth that transcends the boundaries of disciplines.

I would also like to thank Shre Chatterjee, Dr. Saptarshi Das and Dr. Dwaipayan Biswas for many helpful and thought provoking discussions on plant electrophysiology and the role of machine learning and signal processing in this field.

I want to thank the entire PLEASED team Dr. Andrea Vitaletti, Dr. Elisa Masi, Dr. Luisa Santopolo, Dr. Ilaria Colzi and Prof. Stefano Mancuso for sharing the experimental data that is used in this thesis.

I thank my parents, my uncles and my sister for their love and support and endless encouragement!

Finally, thank to my wife, Shivangi, for her unconditional love and support! I thank her for being optimistic at the time of crisis and for gifting me many joyous moments during the course of my doctoral studies.

To Thami...

Chapter 1

Introduction

1.1 Motivation

The time evolution of the variables modelled in a variety of science and engineering branches are often described by ordinary differential equations that are characterised by model structure – the functions of the dynamical variables – and model parameters. In many practical problems it is necessary to infer both the model parameters and an appropriate model structure from the experimental observations. We refer to these problems as inverse problems. The task of estimating model parameters and choosing the most appropriate among competing models from experimental observations, often corrupted with noise, is essentially a statistical inference problem of paramount importance. It should be noted however that there exist vast differences between the principles governing many biophysical phenomena that we wish to understand and predict, and our description of the same in terms of mathematical models. A philosophical perspective adopted by many scientists to overcome the aforementioned differences (or lack of precise knowledge) is the association of uncertainty. Uncertainty thus can be associated with a model structure, the evidence (in form of experimental data) as well as future predictions. However, in practice systematic characterisation of model uncertainty has been given much less consideration with the majority of effort being focused on parameter estimation from an optimisation perspective. The Bayesian methodology ([Gelman et al., 2003](#); [Bernardo and Smith, 2001](#)) of inference is well suited for the quantification of uncertainty in the estimates of parameters as well as uncertainty over a set of candidate models. To apply Bayesian techniques we need to integrate marginal likelihoods, which can be computationally intractable in non-linear differential equation models. For this reason some form of approximation such as Monte Carlo integration is generally preferred for parameter inference. Markov chain Monte Carlo algorithms are used most often to carry out approximate Bayesian inference in non-linear differential equation models ([Girolami, 2008](#)). This thesis presents the process of developing a novel approximate Bayesian inference algorithm that is built upon many recent developments

in Bayesian inference and other allied fields such as machine learning. This novel algorithm, we argue through rigorous experiments, combines the benefits provided by the state-of-the-art Bayesian parameter and model selection methods in terms of quality of estimates, capability of handling uncertainty, and computational efficiency within a single framework. Furthermore, a significant part of this thesis presents an application scenario where we have applied this algorithm for a novel inverse problem pertaining to plant electrophysiology. We have proposed electrophysiological models to explain novel experimental results. To the best of our knowledge this is the first instance of Bayesian inverse modelling in plant electrophysiology and through this problem we have highlighted many useful features of this algorithm in shedding light on a complicated physiological process. In the next section we will provide details on some of the more recent developments in Bayesian inference in the context of differential equations on which our proposed algorithm is built. Then in later sections we will introduce, very briefly, aspects of plant electrophysiology leading to our chosen inverse problem.

1.2 Recent developments in Bayesian inference

A major practical hindrance to the widespread use of Bayesian methods in this context is rooted in the computational burden of MCMC algorithms which require repeated numerical integration of differential equations. For complex models such computational burden appears prohibitive. Thus a number of methods have been proposed that alleviate this computational bottleneck by circumventing the need of numerical integration. These methods are collectively known as gradient matching methods. Any such gradient based inference algorithm operates by smoothing the gradients or the velocity field of the governing models without performing integration. The actual process of inference is however based on optimisation. [Macdonald et al. \(2016\)](#) provides a comprehensive review and historical perspective of many such methods. The idea of gradient based smoothing has been incorporated in recent methods that rely on supervised learning techniques built upon a Bayesian framework. Such methods and algorithms have been shown to work extremely well for complex non-linear ordinary differential equations in [Calderhead et al. \(2008\)](#), [Dondelinger et al. \(2013\)](#), [Wang and Barber \(2014\)](#). These methods employ Gaussian processes to perform supervised learning (also known as regression) achieving significant improvements in smoothing. Moreover, since Gaussian processes are probabilistic models they are easier to implement within the sampling schemes of conventional MCMC algorithms. With the use of smoothing such methods avoid numerical integration required within MCMC sampling schemes providing a computationally faster approach towards Bayesian inference of parameters of differential equation.

Bayesian inference algorithms also suffer from the fact that none of these algorithms provide a holistic platform for uncertainty quantification. As for example many variants

of MCMC algorithms such as the Metropolis-Hastings require modifications and some amount of engineering in order to be applied for the task of parameter estimation as well as model selection. Moreover, such MCMC algorithms do not address the issue of sensitivity analysis. Thus designing a single inference platform within the Bayesian framework that can quantify uncertainty associated with model structure, parameters and sensitivity is highly desirable.

Approximate Bayesian computation (ABC) based on sequential Monte Carlo (SMC) is a recently developed approximate inference technique that provides a consistent platform for uncertainty quantification. It has been applied to the task of both parameter estimation and model selection in [Toni et al. \(2009\)](#). The ABC-SMC algorithm has been shown to work well for the examples considered in [Toni et al. \(2009\)](#). ABC-SMC produces reliable estimates of parameters and has been used to discriminate between a set of candidate models using Bayesian model selection criteria. Moreover, ABC-SMC enables the calculation of parameter sensitivities ([Toni et al., 2009](#)). ABC methods prove to be most useful for large and intractable models with complex likelihood functions that are difficult to evaluate. ABC methods are also known as likelihood-free inference as in ABC evaluation of likelihood is replaced by simulation from a generative model. Note that inference using ABC does not have the theoretical guarantee of recovering the true posterior estimates. Thus, in the context of ABC we will refer MCMC algorithms as exact Monte Carlo algorithms. Also, this method, like other Monte Carlo methods, incurs a significant computational cost as it requires explicit numerical integration of differential equations to carry out inference.

We have noticed that most parameter estimation algorithms proposed in recent literature are tested only on benchmark model systems. This is because such models present formidable known challenges that any inference algorithm must be able to conquer at the least. However, we believe the applicability and usefulness of any inference algorithm can be fully understood and appreciated by solving previously unsolved or novel inverse problems. Having applied the modified ABC-SMC algorithm to solve an electrophysiological inverse problem (mentioned previously) we realized its true usefulness (statistical as well as computational). We will turn to this problem next with a brief introduction to the various electrical signals found in plants.

1.3 Electrophysiological responses in plants

Plants have the ability to sense the surrounding environment. Plants show fundamental physiological responses to many environmental perturbations such as change in incident light, temperature, moisture, pollutants, pesticides etc. These physiological responses are elongation growth, respiration, moisture absorption, photosynthesis and transpiration, pollination, fertilization etc. Investigation of these responses, led by plant scientists,

revealed that most of these physiological processes are coordinated by electrical signals found within plants. Analogies between plant and animal electrical signals have been drawn by many plant scientists. However, fundamental knowledge about these electrical activities within the plants is yet to be gained.

1.3.1 Types and natures of electrical signals in plants

The electrical signals in plants are found to be the potential difference between inside and outside of the plasma membrane. The difference in electrical potential between interior and exterior of a biological cell membrane has been a focal point among biophysicists for over half a decade. This is the fundamental electrical signal found in plants and it is formally defined as the membrane potential. The communication of changes in this potential difference across the plant anatomy is manifested in different forms on the basis of plant physiology and is categorized into two categories; the action potential (AP) and the variation potential (VP). These signals have unique morphological identifiers as well as specific stimulus-response characteristics.

Action potential

Action potential is induced by non-damaging stimuli (cold, mechanical, and electrical stimuli). Action potential can move across the entire plant body transmitting information across and for this reason it is considered the most widespread signalling phenomenon in plants (Pickard, 1973). AP is generally described by its different phases shown in Figure 1.1. The first phase is the firing phase or the depolarizing phase in which the membrane voltage rises above the resting potential and this phenomenon is called the depolarization of the membrane potential. This depolarization voltage is generally found to be approximately between 50mV to 75mV in higher plants (Pickard, 1973). After this rise, the membrane voltage returns back towards the resting potential gradually and this is known as the repolarization and also as the falling phase. After the falling phase the AP can be divided into an absolute refractory period and a relative refractory period. In the absolutely refractory period it is impossible to evoke another AP giving the same stimulus to the same site but during the relative refractory period strong stimulus can evoke further APs from the same site. Sometimes during the repolarizing phase the membrane potential shoots below the resting potential for some duration of time and then again stabilizes at the resting potential.

Variation potential: The slow wave

Variation or slow wave potentials are propagating electrical signals like the APs which also consist of a transient change in membrane potential (the depolarization and the repolarization). However there exist few fundamental differences between the AP and the VP. VPs have delayed repolarization and a large range of variation. Also, these signals vary with the intensity of the stimulus. VPs are generated in response to damaging

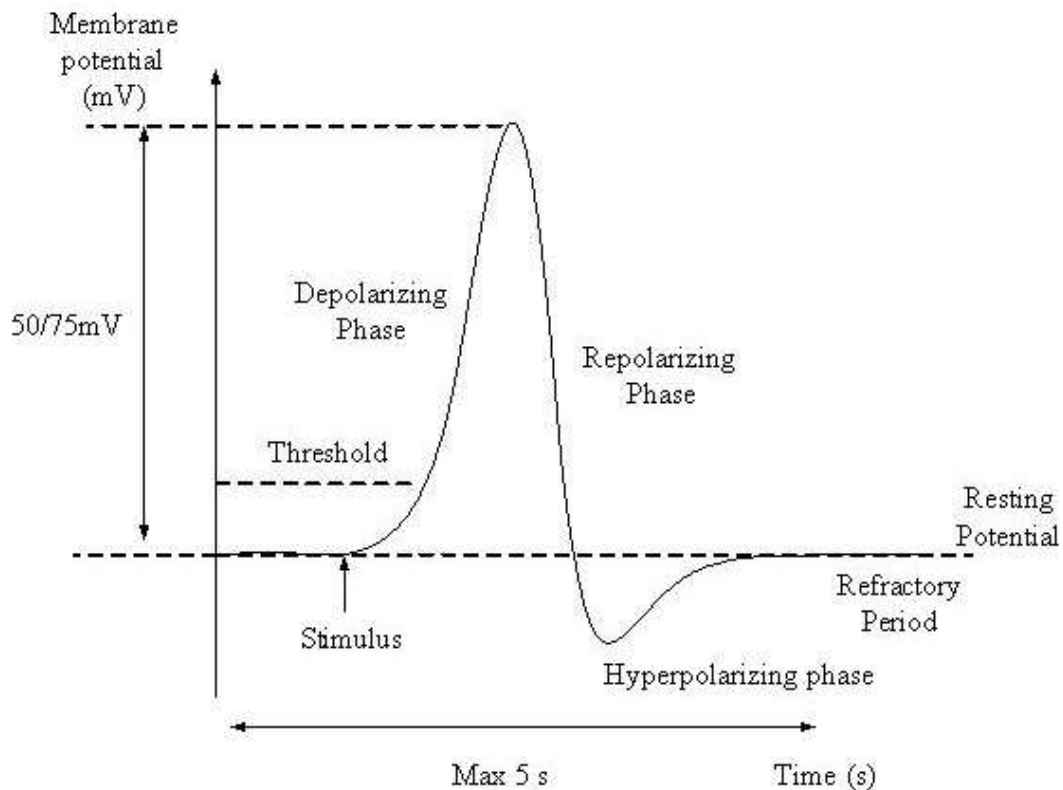


Figure 1.1: Typical shape of the action potential.

stimuli like mechanical wounding, flaming or organ excision by either a local change in hydraulic pressure or by the transmission of chemicals through the dead xylem (Fromm and Lautner, 2007). VPs are studied in a wide variety of plants such as cucumber and peas seedlings (Stahlberg and Cosgrove, 1992) and also in woody plants such as *Vitis Vinifera* (Mancuso, 1999). The VP is characterized by three aspects: 1) Decrease in the amplitude and speed as a function of the distance from the site of injury; 2) Ability to pass dead regions of tissue; 3) Dependence on xylem tension (Fromm and Lautner, 2007). Stankovic et al. (1997) suggested from their experiments on a sunflower plant, that a locally applied wounding of the sunflower plant triggers a surge in pressure in the xylem which activates some mechanosensitive/stretch-responsive ion channels in adjacent living cells. Activation of these ion channels creates ion fluxes evoking a depolarization of the local membrane voltage, which is observed extracellularly as a VP. Figure 1.2 shows the VPs generated in the sunflower plants (Stankovic et al., 1997).

Jasmonate inducing electrical signals

Apart from the aforementioned signals, long distance electrical signals as part of a defence mechanism against wounding has been discovered in (Mousavi et al., 2013; Zimmermann et al., 2009; Felle and Zimmermann, 2007). In Mousavi et al. (2013) transmission of such long distance electrical signal generated by the activity of glutamate-receptor-like (GLR) ion channels in response to herbivore attacks has been discovered in *Arabidopsis*

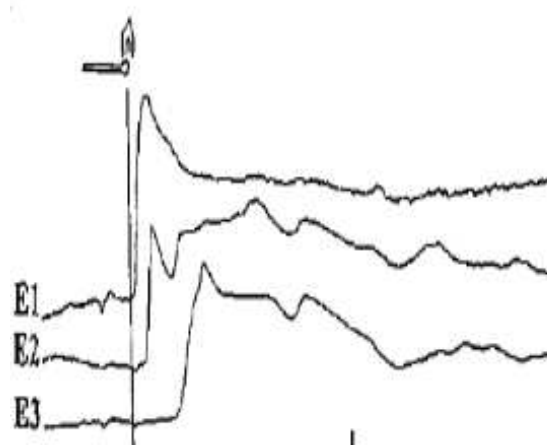


Figure 1.2: VP in sunflower plant [Stankovic et al. \(1997\)](#) triggered by flaming is measured by three electrodes placed 5, 15 and 25 cm from the stimulated site on the stem.

thaliana. These signals induce the increase in hormone jasmonate which triggers defence responses at local and distant sites. Figure 1.3 shows a cartoon illustrating this phenomenon termed as the “Electrical Defence” in [Christmann and Grill \(2013\)](#). The

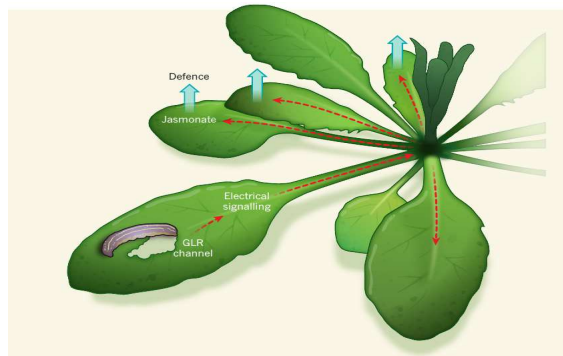


Figure 1.3: Defence responses induced by electrical signalling ([Christmann and Grill, 2013](#)). Transmission of electrical signals to local and distant sites after herbivore attack trigger defence response as jasmonate hormone increases. These electrical signals are generated by the glutamate-receptor-like (GLR) ion channels.

work reported in [Mousavi et al. \(2013\)](#) has raised many questions on the generation and maintenance of these electrical signals. The role of calcium ions in these processes is also being probed.

1.4 Thesis contributions: Increasing the computational efficiency of ABC-SMC

We propose a novel algorithm based on ABC-SMC using gradient based smoothing as an alternative to numerical integration for simulation from the generative model. Specifically, we apply Gaussian process regression for smoothing as in [Calderhead et al. \(2008\)](#) while preserving the inferential advantages of ABC-SMC over conventional MCMC. We bring together these two contrasting approaches to parameter estimation in order to build a holistic platform for handling uncertainty without any computational bottleneck. We will show through experiments on many model systems that the computational efficiency gained in our approach outperforms not only the original ABC-SMC method but also other MCMC methods that employ Gaussian processes. All our model systems are previously used for benchmarking various inference algorithms and are also relevant to many developments particularly in the field of computational biology.

1.4.1 Mathematical modelling of electrical responses

Our work is concerned with the study of the electrophysiological responses in plants due to harmful environmental pollution. In this thesis we attempt to provide a phenomenological characterisation of electrical responses in plants subjected to a particular form of environmental pollution: exposure to ambient ozone. We model the time courses of electrical responses obtained experimentally to arrive at the aforementioned phenomenological characterisation. We have used ozone as our chosen stimulus because it is a global air pollutant that harms plant productivity, and has significant negative impact on agriculture. It enters the plant through stomatal opening present in the leaves and also through pores all over the plant's stem, generating other reactive oxygen species causing oxidative stress, which in turn decreases photosynthesis, plant growth, and biomass accumulation ([Ainsworth et al., 2012](#)) and reduces plant productivity. Due to the harmful (wounding) nature of ozone as a stimulus, we believe the resulting electrical responses shall have traits of the jasmonate inducing signals. Through our proposed mathematical models we attempt to investigate the role of calcium ions behind the generation of these electrical responses.

In this process we will apply the modified fast ABC-SMC algorithm to validate and choose a suitable model in light of the experimental observations. Furthermore, we will use ABC-SMC to quantify the uncertainties associated with many aspects of our proposed model in hope of gaining a thorough understanding of the mechanism behind the stimulus-response phenomenon.

1.5 Thesis structure

The thesis is organised in six chapters (including the present one). In the following we provide a brief outline of the rest of the chapters (2-6):

- Chapter 2, **ABC methods for learning in differential equation models**—In this chapter we present a literature review of the ABC-SMC methods for parameter estimation and model selection in non-linear differential equations. ABC itself is in a burgeoning phase with many developments published in the past couple of years. In this chapter we primarily focus on ABC methods in the context of inference in differential equations. However, we have indicated other developments in the broader field of ABC that we believe is important in establishing our chosen subject.
- Chapter 3, **Gaussian processes for speeding up the ABC-SMC algorithm**—We introduce the theory of Gaussian process regression and apply the same for smoothing the trajectories of an ordinary differential equation to obtain the empirical velocity fields. We then modify the ABC-SMC algorithm by incorporating Gaussian process smoothing to replace numerical integration required for repeated simulations. Furthermore, we introduce briefly other Gaussian process based MCMC algorithms published in recent literature and compare these with the modified ABC-SMC algorithm. Thus in this chapter we introduce our major contribution, a fast ABC-SMC algorithm which we call the GP-ABC-SMC.
- Chapter 4, **Experimental evaluation of the GP-ABC-SMC algorithm**—To show the computation benefits of the GP-ABC-SMC we compare the same with the ABC-SMC algorithm that uses numerical integration. Furthermore, we also compare the GP-ABC-SMC with other algorithms that use Gaussian processes. The experiments were performed using artificial data generated from benchmark model systems. We have taken extreme care while generating such datasets to avoid any systematic bias towards a specific algorithm under comparison.
- Chapter 5, **Mathematical modelling of plant-wide electrical responses in higher plants exposed to ozone**—In this chapter we develop a novel mathematical model that explains the generation of electrical responses in higher plants subjected to ozone exposure. The first part of this chapter introduces experiments that involve the aforementioned ozone treatments. We hypothesise that the electrical signals are manifestation of a plant-wide calcium wave. We then provide a mathematical model that quantifies the dynamics of this calcium wave. In the second part of this chapter we apply the GP-ABC-SMC algorithm to fit this calcium wave model to the experimentally observed electrical signals. Furthermore, we use the output of the GP-ABC-SMC algorithm to further probe and predict the behaviour of our proposed model.

- Chapter 6, **Summary and main contributions**—In this chapter we summarize our two major achievements: i) developing a fast Bayesian inference algorithm for non-linear differential equations, and ii) solving a novel inverse problem in plant electrophysiology. We finish this chapter and the thesis with some indication of future work.

1.6 Publication

The primary contribution of this thesis, the GP-ABC-SMC algorithm, has been published in the following article:

S. Ghosh, S. Dasmahapatra, and K. Maharatna. Fast approximate bayesian computation for estimating parameters in differential equations. *Statistics and Computing* [Online]. Available :<http://dx.doi.org/10.1007/s11222-016-9643-4>.

Chapter 2

ABC methods for learning in differential equation models

2.1 Introduction

The two major tasks related to the inverse modelling, where we strive to explain experimentally measured time courses of physical variables using a dynamical system, are estimating the parameters that control the model behaviour and selecting the best explanation or hypothesis of the experimental observations through selecting a suitable model among many candidates. Both these tasks when posed as statistical questions could be arguably – ‘best’ answered in the framework of Bayesian statistics. In this chapter we will review the fundamentals of ABC methods especially based on sequential Monte Carlo (SMC) for this purpose. Since ABC is a new field which is evolving at a rapid pace we will thus mostly focus on the developments of ABC based methods in the context of inference in deterministic dynamical systems.

2.2 Bayesian parameter estimation and model selection

In the Bayesian framework parameter estimation is carried out by inferring the posterior probability distribution $p(\theta|Y^d)$ of the parameters θ given the data Y^d . Following the Bayes formula the posterior is given by

$$p(\theta|Y^d) = \frac{p(Y^d|\theta)\pi(\theta)}{\int p(Y^d|\theta)\pi(\theta)d\theta}, \quad (2.1)$$

where $p(Y^d|\theta)$ is the likelihood function and $\pi(\theta)$ is the prior distribution of the parameters θ .

Model selection within the Bayesian framework is carried out by providing evidence in support of a model. Given data Y^d and two competing hypotheses H_1 and H_2 , we can calculate the probability of each model hypothesis given the data. The posterior odds (ratio) of these probabilities are given by

$$\underbrace{\frac{p(H_1|Y^d)}{p(H_2|Y^d)}}_{\text{Posterior odds}} = \underbrace{\frac{p(Y^d|H_1)}{p(Y^d|H_2)}}_{\text{Bayes factor}} \underbrace{\frac{p(H_1)}{p(H_2)}}_{\text{Prior odds}}. \quad (2.2)$$

If there is no preference a priori for a particular model, the prior probabilities of the model may be set to equal. Hence for $p(H_1) = p(H_2)$, the Bayes factor given by

$$B_{1,2} = \frac{p(Y^d|H_1)}{p(Y^d|H_2)} \quad (2.3)$$

can be used to choose among the competing hypotheses. The likelihood of the data given a model also known as the integrated or marginal likelihood, obtained by integrating over the parameter space:

$$p(Y^d|H_j) = \int p(Y^d|\theta_j, H_j)\pi(\theta_j|H_j)d(\theta_j) \quad (2.4)$$

where a specific model hypothesis H_j is parametrized by θ_j and $\pi(\theta_j|H_j)$ is the prior density of the parameter θ_j under the same model hypothesis. It is thus easy to understand, intuitively, that the Bayes Factor simply weighs the evidence provided by the data in support of each model through the integrated or marginal likelihood.

In Kass and Raftery (1995) the following qualitative interpretation of Bayes Factor was given: *1 to 3 is barely worth a mention, 3 to 10 is substantial, 10 to 30 is strong, 30 to 100 is very strong and over a 100 is decisive evidence in favour of model H_1 . Values below 1 take the inverted interpretation in favour of model H_2 .*

Although the Bayesian framework provides a cohesive and probabilistic platform for parameter estimation as well as model selection, in the case of a complex model such as a non-linear ODEs the marginal likelihood becomes analytically intractable and thus we have to resort to approximation schemes for inferring the posterior. Bayesian numerical methods based on Markov chain Monte Carlo (MCMC) sampling is generally used for approximate inference in this context (Vyshemirsky and Girolami (2008), Gelman et al. (2003), Calderhead and Girolami (2009)). In this thesis we will be using a somewhat different approach of approximate inference known as the Approximate Bayesian Computation (ABC). ABC methods use a simulation-based procedure to eliminate the computation of the likelihood function in cases where it is intractable. ABC based on sequential Monte Carlo (SMC) is one such approximate inference technique that has been applied to different classes of dynamical systems described by deterministic or stochastic

differential equations for both parameter estimation and model selection in [Toni et al. \(2009\)](#). In the following section we would introduce the reader to a brief review of ABC methods in practice.

2.3 The basic ABC framework

ABC methods are widely used in population genetics ([Pritchard et al. \(1999\)](#), [Beaumont et al. \(2002\)](#)) since their introduction by [Pritchard et al. \(1999\)](#) in the form of the ABC rejection sampler. ABC methods have been conceived with the aim of inferring posterior distributions without evaluating the likelihood function. ABC methods exploit the numerical efficiency of simulation techniques by replacing the calculation of likelihood with a comparison between the observed and simulated data and are proven to be most useful for large models with complex likelihood surfaces that are difficult to evaluate.

The operating principle of ABC methods lies in replacing the evaluation of likelihoods with a simulation based procedure for inference, by using a generative model $\mathcal{M}_\theta$ with parameters θ drawn from a prior distribution $\pi(\theta)$ to simulate observations $Y^s \sim \mathcal{M}_\theta$ that are compared with the observed data Y^d . If the likelihood $p(Y^d|\mathcal{M}_\theta)$ of observed data Y^d is intractable or infeasible to compute, then we can use the ABC algorithm to obtain samples from the following modified posterior density

$$p_\epsilon(\theta, Y^s|Y^d) = \frac{\mathbb{1}(\Delta(Y^d, Y^s) \leq \epsilon)(Y^s \sim \mathcal{M}_\theta)\pi(\theta)}{\int_\theta \int_{Y^s} \mathbb{1}(\Delta(Y^d, Y^s) \leq \epsilon)(Y^s \sim \mathcal{M}_\theta)\pi(\theta)d\theta dY^s} \quad (2.5)$$

where $\epsilon > 0$ is a tolerance level, Δ is a distance function, $\mathbb{1}$ is the indicator function and $p_\epsilon(\theta, Y^s|Y^d) = p(\theta, Y^s|\Delta(Y^d, Y^s) \leq \epsilon)$. A good (enough) approximation of the true marginal posterior distribution is obtained when the distance $\Delta(Y^d, Y^s)$ is within a predetermined small tolerance ϵ , *i.e.*,

$$p_\epsilon(\theta|Y^d) = \int_{Y^s} p_\epsilon(\theta, Y^s|Y^d)dY^s \approx p(\theta|Y^d) \quad (2.6)$$

Since ABC (including ABC-SMC) requires the generation of a number of simulated observations $Y^s \sim \mathcal{M}_\theta$, the generation of observations could be a computationally expensive process. Thus although ABC-SMC mitigates the intractability of evaluating the likelihood function through simulation, repeated simulation from complex models for inference can itself be burdensome. For the case of dynamical systems such simulations require explicit numerical solutions of non-linear differential equations.

ABC methods generally have the following algorithmic form:

1. Sample a candidate parameter vector θ from a prior distribution $\pi(\theta)$ and for each $\theta \sim \pi(\theta)$, simulate a dataset $Y^s \sim \mathcal{M}_\theta$ from a generative model $\mathcal{M}_\theta$.

2. Compute a distance $\Delta(Y^s, Y^d)$ between the simulated dataset, Y^s and the experimental data Y^d . If $\Delta(Y^d, Y^s) \leq \epsilon$, where $\epsilon \geq 0$ is the error tolerance of accepted solutions, then accept θ and reject otherwise.

This scheme is repeated until N parameter values are accepted, which represent a sample from the approximate posterior distribution $p_\epsilon(\theta|Y^d)$. Exact posterior can be obtained from this scheme when $\epsilon = 0$.

2.3.1 ABC-SMC for parameter estimation in ODE

If the prior distribution is very different from the posterior distribution, the basic ABC framework is very inefficient as it spends a considerable amount of time sampling from areas of low likelihood in parameter space, which makes the acceptance rate extremely low. In order to improve upon poor acceptance rates and facilitate exploration of the parameter space, ABC algorithms based on the SMC sampling method were proposed in [Sisson et al. \(2007\)](#), [Del Moral et al. \(2006\)](#) and [Sisson et al. \(2009\)](#) and sequential importance sampling (SIS) in [Toni et al. \(2009\)](#), [Beaumont et al. \(2009\)](#). [Toni et al. \(2009\)](#) applied the ABC algorithm based on SIS for parameter estimation and model selection for a variety of dynamical systems including non-linear ODEs and DDEs, which will also be the focus in this paper. Although all the variants of ABC algorithms that come under the SMC category can potentially be used for inference in dynamical systems, we will specifically focus our attention on the ABC approach as adopted in [Toni et al. \(2009\)](#).

We shall apply ABC-SMC to models of the evolution of state $\mathbf{X}(t) = (X_1(t), \dots, X_K(t))$ that are governed by ODEs or DDEs $\frac{d\mathbf{X}(t)}{dt} = f(\mathbf{X}(t - t_d), \boldsymbol{\theta})$, where t_d stands for the time delay in DDEs, with $t_d = 0$ for ODEs, and $\boldsymbol{\theta}$ is a vector of parameter values. We express the integrated solution of the differential equations $\mathbf{X}(t, \mathbf{X}_{\text{in}}; \boldsymbol{\theta})$ as a map $\psi_t(\mathbf{X}_{\text{in}}; \boldsymbol{\theta})$ that generates state trajectories $\mathbf{X}(t)$ given a set of parameters $\boldsymbol{\theta}$ and initial conditions $\mathbf{X}_{\text{in}} \triangleq \mathbf{X}(t_d \leq t \leq 0)$. To generate the samples Y^s , we obtain the solutions $\mathbf{X}(t, \mathbf{X}_{\text{in}}; \boldsymbol{\theta})$ considering the differential equation as the generative model $\mathcal{M}_\theta$:

$$(Y^s \sim \mathcal{M}_\theta) \Leftrightarrow Y^s = \mathbf{X}(t, \mathbf{X}_{\text{in}}; \boldsymbol{\theta}). \quad (2.7)$$

to be used in the ABC framework.

A collection of parameter values, called particles $\boldsymbol{\theta}$ are sampled from the prior $\pi(\boldsymbol{\theta})$ to instantiate the generative model $\mathcal{M}_\theta$. To decide whether a particular choice of parameters $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$ is accepted, we need to compare if the simulated trajectory $Y^s \sim \mathcal{M}_\theta$ is within a tolerance level ϵ of the observed trajectory Y^d , for which we introduce the distance function

$$\Delta(Y^d, Y^s) = \sum_{i=1}^L \sum_{k=1}^K (Y_k^d(t_i) - X_k(t_i))^2, \quad (2.8)$$

where we assumed that the data were collected, and the state evaluated, at discrete time points $t^L \triangleq \{t_i\}_{i=1,\dots,L}$. Note that for dynamical systems such distance functions are generally built by considering the entire time-series data instead of some sufficient statistics.

The sequential stage of this algorithm involves replacing a single tolerance value ϵ by a sequence of tolerance values ϵ_τ , where $\tau = 0, \dots, S_{MC}$ denotes the sequential steps, and $\epsilon_\tau > \epsilon_{\tau+1}$. The particles θ_τ are indexed by τ labelling the tolerance level, and are sampled from the intermediate distribution $p(\theta_{\tau-1} | \Delta(Y^d, Y^s) \leq \epsilon_{\tau-1})$ obtained from the previous sequential step, thus introducing a step-wise procedure for generating parameters from a sequence of increasingly informative distributions, starting at $\tau = 0$ with the prior distribution $\pi(\theta)$. To accept or reject the sampled particles for sequence index τ , the generated trajectories from the model with parameters θ_τ must be closer to the observed data Y^d than those generated from the model with parameters $\theta_{\tau-1}$ in step $\tau - 1$. The generative mechanism for the particles θ_τ differs from that of $\theta_{\tau-1}$ in that they are sampled from the N particles $\{\theta_{\tau-1}^{(i)}\}_{i=1,\dots,N}$ with importance weights (Toni et al., 2009) $w_{\tau-1}$ and each $\theta^* \sim \{\theta_{\tau-1}^{(i)}\}_{i=1,\dots,N}$ is moved in the parameter space by using some transition kernel $K_\tau(\theta | \theta^*)$ (Toni et al., 2009). By movement we mean a random walk. This movement in the parameter space is usually called perturbation and therefore the transition kernel is called a perturbation kernel.

For each $\tau = 0, \dots, S_{MC}$, the N particles meeting the acceptance criterion $\Delta(Y^s, Y^d) \leq \epsilon_\tau$ represent a point-wise approximation for the posterior distribution over the parameter values:

$$p_{\epsilon_\tau}(\theta | Y^d) \approx \frac{1}{N} \sum_{i=1}^N w_\tau^{(i)} \delta(\theta - \theta_\tau^{(i)}), \quad (2.9)$$

Essentially this algorithm works like a particle filter (see figure 2.1) wherein the particles sampled from the prior are filtered and passed through a series of intermediate distributions until they represent an approximation of the target i.e the true posterior distribution. ABC-SMC is fundamentally based on SIS and thus it draws upon the notion of importance sampling by assigning weights to the particles in a population and then sampling from the weighted distribution. Weighting the particles is done to address the discrepancy between the true and approximate intermediate distributions. For a sufficiently large number of particles, the population approach can avoid the problem of getting stuck in areas of low probability unlike the basic ABC algorithm mentioned previously. The pseudo-code for the ABC-SMC algorithm is listed in Algorithm 1.

2.4 Algorithmic settings

The success of ABC-SMC algorithm both in terms of computational complexity and quality of the solution depends on the choice of the ϵ_τ schedule (the term “schedule”

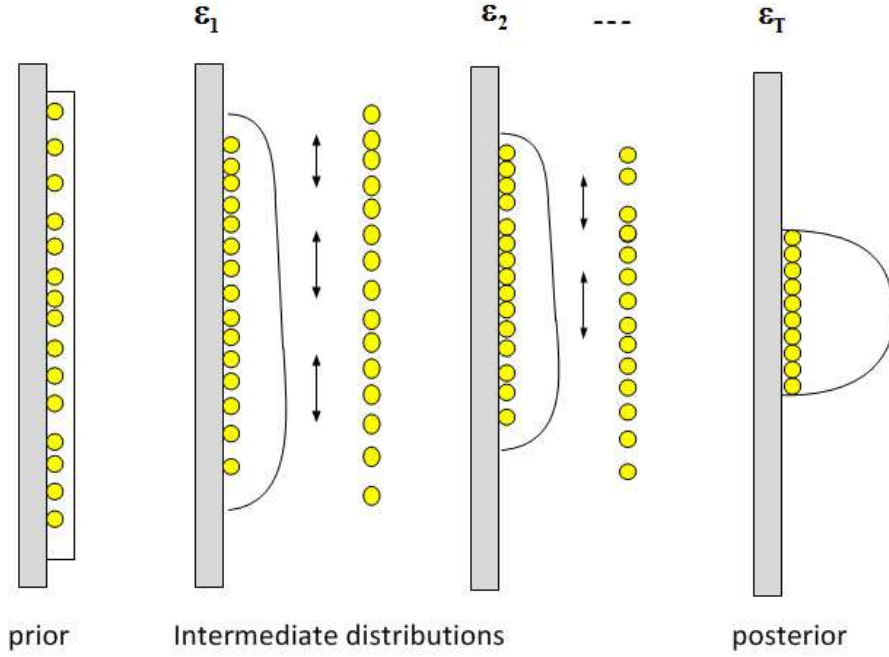


Figure 2.1: In the ABC-SMC algorithm a sample from the prior distribution is passed through filtering steps where the probability of the particles to represent the data best is updated during the algorithm via intermediate distributions. The final distribution is an approximation of the true posterior distribution. The SMC algorithm follows a particle-filtering approach.

denotes a decreasing sequence of tolerances) and the perturbation kernel K_τ . In this section we will briefly describe these two algorithmic settings. A detailed discussion concerning the effects of these settings can be found in [Filippi et al. \(2013\)](#).

2.4.1 Tolerance schedule

Until recently, tolerance values were manually tuned in practice based on prior empirical knowledge about the model. An adaptive choice of the tolerance values has been proposed in [Del Moral et al. \(2012\)](#) and [Drovandi and Pettitt \(2011\)](#). In an adaptive schedule the value of the tolerance ϵ_τ is chosen as the α -th quantile, where $0 \leq \alpha \leq 1$ of the distances between the observed data Y^d and simulated data $Y_{\tau-1}^s$ generated at the previous algorithmic time.

2.4.2 Perturbation kernel

Perturbation kernels hold the key to the acceptance rates in ABC-SMC and the speed of the algorithm as exploited in [Filippi et al. \(2013\)](#). Perturbation kernels can be broadly divided into two categories: a component-wise perturbation kernel and a multivariate

Algorithm 1 ABC-SMC as proposed in [Toni et al. \(2009\)](#)

-
1. Given Y^d , $\pi(\theta)$, $\mathcal{M}_\theta$.
 2. Initialise $\epsilon_\tau > 0$, $\tau = 0, \dots, S_{MC}$, $\epsilon_\tau > \epsilon_{\tau+1}$. Set $\tau = 0$.
 3. Set $i = 1$.
 4. **if** $\tau = 0$ **then**
 5. sample θ^{**} independently from $\pi(\theta)$:
 $\theta^{**} \sim \pi(\theta)$
 6. **else**
 7. from the previous population $\{\theta_{\tau-1}^{(i)}\}_{i=1, \dots, N}$
 sample $\theta^* \sim \{\theta_{\tau-1}^{(i)}\}_{i=1, \dots, N}$ with associated weights $w_{\tau-1}^*$ and use the perturbation kernel $K_\tau(\theta|\theta^*)$ to produce $\theta^{**} \sim K_\tau(\theta|\theta^*)$.
 8. **end if**
 9. **if** $\pi(\theta^{**}) = 0$ **then**
 10. go to 4.
 11. **else**
 12. Simulate a candidate dataset Y^s from the model $\mathcal{M}_\theta$ with parameter θ^{**} : $Y^s \sim \mathcal{M}_\theta|_{\theta \leftarrow \theta^{**}}$.
 13. **end if**
 14. **if** $\Delta(Y^d, Y^s) \geq \epsilon_\tau$ **then**
 15. go to 4.
 16. **else**
 17. Set $\theta_\tau^{(i)} \leftarrow \theta^{**}$ and calculate the weight for particle $\theta_\tau^{(i)}$,

$$w_\tau^{(i)} = \begin{cases} 1, & \text{if } \tau = 0 \\ \frac{\pi(\theta_\tau^{(i)})}{\sum_{j=1}^N w_{\tau-1}^{(j)} K_\tau(\theta_\tau^{(i)}|\theta_{\tau-1}^{(j)})}, & \text{if } \tau > 0 \end{cases}$$
 18. **end if**
 19. **if** $i < N$ **then**
 20. Set $i \leftarrow i + 1$ and go to 4.
 21. **else**
 22. Normalise the weights.
 23. **end if**
 24. **if** $\tau < S_{MC}$ **then**
 25. Set $\tau \leftarrow \tau + 1$ and go to 3.
 26. **else**
 27. **return** particles $\theta_{S_{MC}}^{(i)}$ at $\tau = S_{MC}$.
 28. **end if**
-

perturbation kernel. In a component-wise perturbation kernel $\theta \sim \mathcal{N}(\theta, \Sigma_\tau)$ where Σ_τ is a diagonal covariance matrix whose diagonal entries $\sigma_{\tau,j}^2$, $j = 1, \dots, d$ are chosen adaptively according to the previous population labelled by $\tau - 1$ ([Beaumont et al., 2009](#); [Didot et al., 2011](#); [Filippi et al., 2013](#)).

2.4.2.1 Multivariate perturbation kernel

A component-wise perturbation kernel is, by construction, unable to generate particles with correlated components; therefore, for models with strongly correlated parameters

the ABC-SMC sample generator will not be able to reflect the structure of the posterior and the acceptance rate will be low. Thus, in order to capture such correlations the particles can be perturbed according to a multivariate normal distribution with a non-diagonal covariance matrix Σ_τ that depends on the covariance of the particles as reflected in the population in the previous sequential step ($\tau - 1$) (Filippi et al., 2013). Furthermore, a multivariate perturbation kernel operating on a subset of size N' of the N particles (a local kernel) was also shown (Filippi et al., 2013) to produce a noticeable improvement in the acceptance rate. In order to define this kernel we will introduce some notation. Let $Y_\tau^{s(i)}$ denote the simulated data generated from $\mathcal{M}_\theta$ with particle $\theta \leftarrow \theta_\tau^{(i)}$, $i = 1, \dots, N$ from a population of size N at algorithmic time τ . The corresponding importance weights are denoted as $w_\tau^{(i)}$. We collect all such particles (along with the weights) from algorithmic time $\tau - 1$ for which $Y_\tau^{s(i)}$ is not only within distance $\epsilon_{\tau-1}$ of the observed data Y^d but also within distance ϵ_τ of it. We denote such particles as $\tilde{\theta}_{\tau-1}^{(j)}$:

$$\left\{ \tilde{\theta}_{\tau-1}^{(j)} \right\}_{1 \leq j \leq N'} = \left\{ \theta_{\tau-1}^{(i)} \mid \Delta(Y^d, Y_{\tau-1}^{s(i)}) \leq \epsilon_\tau, 1 \leq i \leq N \right\}, \quad (2.10)$$

with associated normalised weights $\tilde{w}_{\tau-1}^{(j)} \triangleq (w_{\tau-1}^{(j)} / \bar{w})$, with $\bar{w} \triangleq \sum_j w_{\tau-1}^{(j)}$.

Having defined the pairs $(\tilde{\theta}_{\tau-1}^{(j)}, \tilde{w}_{\tau-1}^{(j)})$ we can now use a multivariate normal distribution $\mathcal{N}(\theta_{\tau-1}^{(i)}, \Sigma_\tau^i)$, with a local covariance Σ_τ^i (termed the optimal local covariance in Filippi et al., 2013), to perturb a particle $\theta_{\tau-1}^{(i)}$, where local refers to particle i . This covariance is given by

$$\Sigma_\tau^i = \sum_{j=1}^{N'} \tilde{w}_{\tau-1}^{(j)} \left(\tilde{\theta}_{\tau-1}^{(j)} - \theta_{\tau-1}^{(i)} \right) \left(\tilde{\theta}_{\tau-1}^{(j)} - \theta_{\tau-1}^{(i)} \right)^T. \quad (2.11)$$

2.4.2.2 Fisher information to construct kernel

Another construction of a local covariance matrix is possible that utilizes the Fisher information matrix (FIM) (R. Rao, 1945; MacKay, 2003). The FIM $\mathbf{I}(\theta)$ defined as

$$\mathbf{I}(\theta) = -E_x \left[\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) \right] \quad (2.12)$$

measures the amount of information that the observable random variable $\mathbf{x}$ carries about the parameter $\theta \in \mathbb{R}^D$ where $f(\mathbf{x}|\theta)$ is the likelihood function. This specific form of the information matrix is known as the expected information and just the Hessian of the likelihood term, without the expectation operator, is known as the observed information. If x has a multivariate Gaussian distribution $\mathcal{N}(\mu(\theta), \Sigma)$, where $\mu(\theta)$ is the mean vector and Σ is the covariance matrix (not a function of θ), then the information matrix $\mathbf{I}(\theta)$ is given by the following identity (see Porat and Friedlander (1986) for the derivation):

$$\mathbf{I}(\theta) = \frac{\partial \mu^T}{\partial \theta_j} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_l}, \quad (2.13)$$

where $j, l = 1, \dots, D$ indexes a D -dimensional parameter vector $\boldsymbol{\theta}$. An important fact about the information matrix, relevant to covariance construction, is that the variance of the unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ (considering a point estimate) is lower bounded by the reciprocal of the Fisher information:

$$\text{var}(\hat{\boldsymbol{\theta}}) \geq \frac{1}{\mathbf{I}(\boldsymbol{\theta})} \quad (2.14)$$

The inequality given in equation 2.14 is known as the Cramer-Rao bound (Cramér, 1999; R. Rao, 1945). This inequality straightforwardly suggests the use of the inverse of FIM $\mathbf{I}^{-1}(\boldsymbol{\theta})$ as a covariance for a multivariate normal perturbation kernel of the kind discussed in the previous sections. One advantage here is that such a covariance, when constructed using the parameters obtained by importance sampling, will always capture the maximum information about the model behaviour. This will lead to the generation of a variety of simulated data Y^s that might lie close to the observed data Y^d which in turn will increase the acceptance rate of SMC. A perturbation kernel based on the FIM was first proposed in Filippi et al. (2013). For an ODE system the FIM can be evaluated (approximately) using the ODE sensitivity equations as in Filippi et al. (2013) and Girolami and Calderhead (2011). We have provided a brief description of such construction of the FIM in C.1.

2.4.3 Particle population size

The ABC-SMC algorithm as described in Algorithm 1 is based upon the gradual adaptation of the intermediate distributions $p_{\epsilon_\tau}(\boldsymbol{\theta}_\tau) = p(\boldsymbol{\theta}, Y^s | \Delta(Y^d, Y^s) \leq \epsilon)$ using the tolerances ϵ_τ . To obtain samples from these intermediate distributions a corresponding series of importance distributions $q_\tau(\boldsymbol{\theta}_\tau)$ is used. Now success of any importance sampling mechanism (including ABC-SMC) is based on the crucial choice of a suitable importance distribution (see Del Moral et al. (2006)). ABC-SMC builds upon the SIS framework by sequentially adapting $q_\tau(\boldsymbol{\theta}_\tau)$ starting from the prior density $q_0(\boldsymbol{\theta}_0) = \pi(\boldsymbol{\theta})$. In a standard SIS setting (see for example Del Moral et al. (2006)) the importance density is given by

$$q_\tau(\boldsymbol{\theta}_\tau) = \int q_{\tau-1}(\boldsymbol{\theta}_{\tau-1}) k_{\tau-1}(\boldsymbol{\theta}_{\tau-1}, \boldsymbol{\theta}_\tau) d\boldsymbol{\theta}_{\tau-1}, \quad (2.15)$$

where $k_{\tau-1}(\boldsymbol{\theta}_{\tau-1}, \boldsymbol{\theta}_\tau)$ is a transition kernel for algorithmic step τ . Within ABC-SMC the adaptation of the importance distribution is given by

$$\begin{aligned} q_\tau(\boldsymbol{\theta}_\tau) &= \mathbb{1}(\pi(\boldsymbol{\theta}_\tau) \geq 0) \times \int p_{\epsilon_{\tau-1}}(\boldsymbol{\theta}_{\tau-1}) K_{\tau-1}(\boldsymbol{\theta}_\tau | \boldsymbol{\theta}_{\tau-1}) d\boldsymbol{\theta}_{\tau-1} \\ &= \mathbb{1}(\pi(\boldsymbol{\theta}_\tau) \geq 0) \times \int w_{\tau-1}(\boldsymbol{\theta}_{\tau-1}) q_{\tau-1}(\boldsymbol{\theta}_{\tau-1}) K_{\tau-1}(\boldsymbol{\theta}_\tau | \boldsymbol{\theta}_{\tau-1}) d\boldsymbol{\theta}_{\tau-1} \\ &\approx \mathbb{1}(\pi(\boldsymbol{\theta}_\tau) \geq 0) \frac{1}{N} \sum_{\boldsymbol{\theta}_{\tau-1}^{(i)} \sim p_{\epsilon_{\tau-1}}} w_{\tau-1}(\boldsymbol{\theta}_{\tau-1}^{(i)}) K_{\tau-1}(\boldsymbol{\theta}_\tau | \boldsymbol{\theta}_{\tau-1}^{(i)}), \end{aligned} \quad (2.16)$$

where N is the number of samples used to approximate the integral in the above equation, $w_{\tau-1}(\boldsymbol{\theta}_{\tau-1}) = p_{\epsilon_{\tau-1}}(\boldsymbol{\theta}_{\tau-1})/q_{\tau-1}(\boldsymbol{\theta}_{\tau-1})$ and $K_{\tau-1}$ are the weights and perturbation kernel respectively for algorithmic step τ and $\pi(\boldsymbol{\theta})$ is the prior distribution. Essentially equation 2.16 describes $q_{\tau}(\boldsymbol{\theta}_{\tau})$ as a perturbed version of the intermediate distribution $p_{\epsilon_{\tau-1}}(\boldsymbol{\theta}_{\tau-1})$ found in the previous algorithmic step $\tau - 1$. This is a reasonable choice leading to samples of $p_{\epsilon_{\tau}}(\boldsymbol{\theta}_{\tau})$ obtained from $p_{\epsilon_{\tau-1}}(\boldsymbol{\theta}_{\tau-1})$. However, having the perfectly engineered tolerance schedule and perturbation kernel as described in the previous sections does not guarantee the best importance distribution as inaccuracy can be introduced through the Monte Carlo approximation of the integral in equation 2.16 when N is set very low. Although a relatively higher value of $N \geq 1000$ is desirable to mitigate any approximation error, higher values of N increases the computational burden in case of complex models for which we want to use ABC-SMC in the first place. In practice we have found a relatively low value of the sample size $N = 300$ used in Toni et al. (2009).

2.5 Model selection using ABC

Model selection using ABC methods is largely an open problem. Among various model selection approaches a consensus has not yet been reached about the most suitable method that can be applied to a large number of problems. However, there are two basic approaches to model selection in ABC, both requiring the estimation of a Bayes factor, that have been used most widely. The first approach is based on a rejection sampling idea (Grelaud et al., 2009) that uses the relative frequency (the ratio) of satisfying a tolerance by each model as a proxy for the Bayes factor. Also, in this approach a nested model (embedding all other models) is used and thus only a single run of an ABC (with some modifications to accommodate the nested model) algorithm is required for estimating the Bayes factor. In Toni et al. (2009) this idea is adopted within ABC-SMC and further developed and used for estimating the Bayes factor. The second approach is premised upon approximating the marginal likelihood

$$p(Y^d|\mathcal{M}_{\theta}) = \int p(Y^d|\boldsymbol{\theta}, \mathcal{M}_{\theta})\pi(\boldsymbol{\theta}, \mathcal{M}_{\theta})d\boldsymbol{\theta} \quad (2.17)$$

of the data Y^d given a model $\mathcal{M}_{\theta}$ parametrized by $\boldsymbol{\theta}$ and then using this for estimating the Bayes factor. This approach is used in Murakami (2014) and had been adopted in ABC-SMC algorithmic framework in Toni et al. (2009). The first approach based on estimating the Bayes factor directly has been shown to work well for dynamical systems problem in Toni et al. (2009). Thus we have chosen to use this model selection technique for our purposes and will introduce the methodology in the following section.

2.5.1 Model selection approaches in ABC-SMC

ABC-SMC can be used for model selection, using the rejection sampling idea, by including a discrete parameter $m = 1, \dots, M$ representing the models $\{\mathcal{M}_\theta^m\}$, where M is the number of models. We denote the model specific parameters as $\theta(m) \in \mathbb{R}^{D_m}$, where D_m is the number of parameters for the m -th model. In each population we start by sampling the model indicator m^* from a prior distribution $\pi(m)$ defined on the indicator. For a specific model $\mathcal{M}_\theta^{m^*}$ we propose new particles θ^{**} by perturbing the particle θ^* sampled from the previous population $\{\theta^{(i)}(m^*)_{\tau-1}\}_{i=1,\dots,N}$ specific to m^* in a similar way to the parameter estimation algorithm (Algorithm 1). The weights for particles $\theta(m)$ are also calculated in a similar way. From the collection of model indicators at each SMC step we can calculate the corresponding frequency $\#m$ of a particular model $\{\mathcal{M}_\theta^m\}$ for being able to meet the tolerance at that step of SMC. The ratio of these frequencies in the final step represent the Bayes factor between two competing models. Thus for models $\mathcal{M}_\theta^m$ and $\mathcal{M}_\theta^p$, where $m, p = 1, \dots, M$, the Bayes factor is given by

$$B_{m,p} = \frac{\#m}{\#p}. \quad (2.18)$$

This method of model selection can be carried out within the framework of Algorithm 1 with the aforementioned amendments in the sampling steps. Thus we present the pseudo-code for the model selection using ABC-SMC in Algorithm 2.

Model selection using ABC has its limitations. When in ABC distance functions are based on sufficient statistics that is in place of the data Y^d, Y^s some summary statistics $\mathcal{S}(Y^d, Y^s)$ is used to build a distance function $\Delta(\mathcal{S}(Y^s), \mathcal{S}(Y^d)) \leq \epsilon_\tau$ a correct approximation to the true Bayes factor cannot be obtained. This was pointed out in Robert et al. (2011) where a proof of this fact is shown. For dynamical systems problem, especially those involving ODEs, since no summary of the data is used this argument stands irrelevant. However, for model selection in dynamical systems problem Bayes factor is used (Toni et al., 2009; Murakami, 2014) and the accuracy of this quantity in comparison with the true Bayes factor as well as those estimated using MCMC (Vysheirsky and Girolami, 2008) has not been explored sufficiently.

2.6 Limitations of ABC-SMC

The development of ABC is in a nascent stage and thus although a lot of work has been done to solidify the theoretical foundations and advancing the methodology to tackle bigger challenges, it still has several theoretical as well as practical limitations. A comprehensive review of ABC methods is carried out in Marin et al. (2012) which elucidates some of its major limitations. As some of these limitations are application

Algorithm 2 ABC-SMC for model selection as proposed in [Toni et al. \(2009\)](#)

-
1. Given Y^d , $\pi(\theta)$, $\{\mathcal{M}_\theta^m\}, \pi(m)$.
 2. Initialise $\epsilon_\tau > 0$, $\tau = 0, \dots, S_{MC}$, $\epsilon_\tau > \epsilon_{\tau+1}$. Set $\tau = 0$.
 3. Set $i = 1$.
 4. Sample m^* from $\pi(m)$.
 5. **if** $\tau = 0$ **then**
 6. sample θ^{**} independently from $\pi(\theta(m^*))$:
 $\theta^{**} \sim \pi(\theta(m^*))$
 7. **else**
 8. from the previous population $\{\theta^{(i)}(m^*)_{\tau-1}\}_{i=1, \dots, N}$
sample $\theta^* \sim \{\theta^{(i)}(m^*)_{\tau-1}\}_{i=1, \dots, N}$ with associated weights $w^*(m^*)_{\tau-1}$ and use the perturbation kernel $K_\tau(\theta|\theta^*)$ to produce $\theta^{**} \sim K_\tau(\theta|\theta^*)$.
 9. **end if**
 10. **if** $\pi(\theta^{**}) = 0$ **then**
 11. go to 4.
 12. **else**
 13. Simulate a candidate dataset Y^s from the model $\mathcal{M}_\theta^{m^*}$ with parameter θ^{**} : $Y^s \sim \mathcal{M}_\theta^{m^*} |_{\theta \leftarrow \theta^{**}}$.
 14. **end if**
 15. **if** $\Delta(Y^d, Y^s) \geq \epsilon_\tau$ **then**
 16. go to 4.
 17. **else**
 18. Set $\theta_\tau^{(i)} \leftarrow \theta^{**}$ and $m_\tau^{(i)} \leftarrow m^*$. Calculate the weight for particle $\theta^{(i)}(m^{(i)})_\tau$,
$$w^{(i)}(m^{(i)})_\tau = \begin{cases} 1, & \text{if } \tau = 0 \\ \frac{\pi(\theta_\tau^{(i)})}{\sum_{j=1}^N w_{\tau-1}^{(j)} K_\tau(\theta_\tau^{(i)} | \theta_{\tau-1}^{(j)})}, & \text{if } \tau > 0 \end{cases}$$
 19. **end if**
 20. **if** $i < N$ **then**
 21. Set $i \leftarrow i + 1$ and go to 4.
 22. **else**
 23. Normalise the weights for each m .
 24. **end if**
 25. **if** $\tau < S_{MC}$ **then**
 26. Set $\tau \leftarrow \tau + 1$ and go to 3.
 27. **else**
 28. **return** model indicators $m_{S_{MC}}^{(i)}$ at $\tau = S_{MC}$.
 29. **end if**
-

specific thus we would instead discuss here the case of ABC-SMC for learning non-linear differential equations, which is the central theme of this thesis.

The first and most general problem of ABC-SMC is that its approximation of the true posterior distribution is accurate when the final tolerance $\epsilon_{S_{MC}}$ is zero. However, in practice this is never set to zero in differential equation problem. It then becomes necessary to quantify the discrepancy between the estimated and the true posterior as well as finding the guarantee, if it exists, that the estimation will converge to the true solution, the actual posterior distribution. There had been some investigation of the convergence properties of ABC, in a general setting, in recent literature ([Barber](#)

et al., 2015). However, convergence properties of ABC-SMC in particular (largely) and especially in the context of inverse problem involving non-linear differential equations has remained unsolved. There are other challenges when applying ABC-SMC for learning in differential equations that are specific to the SIS implementation of the ABC-SMC algorithm. These are as follows:

1. **Computational bottleneck:** For dynamical systems, each simulated observation Y^s generated by the ABC-SMC algorithm is obtained through explicit numerical solution of non-linear differential equations. Thus, this process of generating simulated observations is computationally expensive. In particular, the acceptance criterion $\mathbb{1}(\Delta(Y^d, Y^s) \leq \epsilon)$ can lead to the generation of many unused simulations $Y^s \sim \mathcal{M}_\theta$, and various methods have been proposed in Filippi et al. (2013) to improve the acceptance rate and reduce the run-time of the algorithm. These methods, through clever choices of perturbation kernels, reduce the number of simulations in the SMC steps, but for complex models each simulation is still associated with expensive computations.
2. **Choice of tolerances:** The usage of ABC-SMC algorithm as described in Algorithm 1 requires a crucial choice of the tolerance schedule. The tolerances can be chosen by hand through multiple trials or adaptively as mentioned previously. Both these choices do not incorporate any knowledge about the error surface on which the distance function $\Delta(Y^s, Y^d) \leq \epsilon_\tau$ is constructed. Incorporating this knowledge while choosing tolerances remains challenging and has not been done yet. However, without understanding geometric properties of the error surface it remains difficult to choose the tolerances in order to save computational time wasted in simulations while producing reliable estimates of the posterior.

2.6.1 Advantages

The limitations highlighted in the previous section pose a formidable challenge to practitioners and thus alleviating such hindrances becomes necessary. Regardless of this fact ABC-SMC, as reviewed in this chapter, does present itself as a handy tool for carrying out Bayesian inference in differential equations. It provides a unified platform to carry out both parameter estimation and model selection. Furthermore, the samples generated in the SMC stages can be used as diagnostics of the model itself and the final parameter population can be used to learn about parameter sensitivities (Secrier et al., 2009). However, the most salient feature of this algorithm is that it can handle complex non-linear models without introducing changes to the algorithm, except for the simulation mechanism. Thus, if it is possible to simulate a dynamical systems model then its parameters can be learnt using this algorithm.

2.7 Conclusion and thesis goals

In this chapter we have carried out a detailed review of the ABC-SMC algorithm for inference in dynamical systems described as ODEs and DDEs. Out of its two major limitations almost no work had been done on alleviating its major drawback – the huge computational burden resulting from massive number of simulations requiring numerical integration. One of the major goals of this thesis will be solving this issue using Gaussian processes. In the next chapter we will introduce the Gaussian process and show how it can be applied in the context of saving time incurred in simulation within ABC-SMC. Although we have shed light on the second limitation – construction of optimal tolerances, in this thesis and subsequent chapters we will be entirely focusing on the first issue as we believe speeding up ABC-SMC could potentially lead us towards solving allied challenges.

Chapter 3

Gaussian processes for speeding up the ABC-SMC algorithm

3.1 Introduction

In this chapter we propose a method of speeding up the ABC-SMC algorithm for parameter estimation and model selection in deterministic models described by ordinary differential equations (ODE) or delay differential equations (DDE) by reducing the time incurred in simulation. We achieve this speedup by: (i) completely circumventing the process of integrating the differential equation by operating on the derivative space and (ii) by smoothing the derivatives using Gaussian processes (GP). It should be noted that using Gaussian processes as functional emulators in the derivative space, as a concept, has been proposed in [Calderhead et al. \(2008\)](#), [Dondelinger et al. \(2013\)](#) for speeding up parameter estimation in deterministic differential equations. For parameter estimation, GP-based gradient matching has been used for ODEs and DDEs using a population Monte Carlo sampling ([Calderhead et al., 2008](#)); an adaptive variant of this approach is proposed for ODEs ([Dondelinger et al., 2013](#)). See [Wang and Barber \(2014\)](#) for a review and comparison between these approaches. The novelty of our proposed method is the fusion of GP regression with ABC-SMC. Our algorithm for fast parameter estimation can be easily incorporated into methods for model selection and recovering parameter sensitivities for deterministic differential equations.

3.2 Gradient based parameter estimation in differential equations

We have mentioned previously that the computational bottleneck stems from the explicit integration carried out in each simulation step. In order to avoid the integration

one could essentially use a gradient based estimation. If the temporal variations in observations $Y^d(t)$ are believed to be less smooth than the underlying state evolution that is modelled by differential equations, we shall introduce the target state variable $\widehat{\mathbf{X}}(t)$ to be the smoothed version $\widehat{\mathbf{X}}(t) \triangleq \mathcal{S}(Y^d)$ of the observations. Here, $\mathcal{S}$ represents any smoothing procedure, and we will use Gaussian Process (GP) regression to perform the smoothing below. In the ABC framework, we shall accept the trajectories $\mathbf{X}(t)$ from the model $\mathcal{M}_\theta$ (see equation 2.7) if they are close to $\widehat{\mathbf{X}}(t)$. Once we have $\widehat{\mathbf{X}}(t)$ we can compute its numerical derivative to obtain the empirical velocity field $\mathbf{V}^d(t)$ of the dynamical system $\mathcal{M}_\theta$:

$$\text{for } \widehat{\mathbf{X}}(t) \triangleq \mathcal{S}(Y^d), \mathbf{V}^d(t) \triangleq \frac{d}{dt} \widehat{\mathbf{X}}(t). \quad (3.1)$$

In addition, while the left hand side of the ODE $\frac{d}{dt} \mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta})$ is estimated by the empirical derivative $\mathbf{V}^d(t)$, it should be matched by the model vector field $\mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta})$ on the right hand side, when evaluated on the smoothed state data $\mathbf{f}(\mathbf{X}(t) = \widehat{\mathbf{X}}(t), \boldsymbol{\theta})$.

Upon introducing a new distance measure between $\mathbf{V}^d(t)$ obtained from the smoothing and $\mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta})$ obtained from the vector field we can eliminate the original distance metric for ABC-SMC between observed, Y^d , and the simulated, $Y^s \sim \mathcal{M}_\theta$, trajectories, thus unburdening ABC-SMC of ODE integration at each simulation step. The gradient based method was first suggested in Varah (1982) where a spline-based smoothing was used to denoise the observed data. In this method a cost function was built using the distance metric in derivative space and optimisation was used to minimise this cost function in order to obtain point estimates. Recent developments of this methods are described in Ramsay et al. (2007). All these approaches suffer from similar problems of using additional regularisation parameters for smoothing and often the estimates are sub-optimal point estimates. Although porting the derivative based distance within an ABC scheme alleviates the computational bottleneck, this approach suffers from an inherent shortcoming that is rooted in obtaining a numerical derivative as this might lead to information loss.

In our approach, we replace the numerical differentiation with a zero mean Gaussian Process (GP) prior on the state $\mathbf{X}(t)$ given by

$$p(\mathbf{X}(t)|\boldsymbol{\phi}) \sim \mathcal{GP}(0, K(t, t'; \boldsymbol{\phi})), \quad (3.2)$$

where $K(t, t'; \boldsymbol{\phi})$ denotes a covariance function with hyperparameters $\boldsymbol{\phi}$. Once such a prior is established then Gaussian Process regression techniques can be applied to estimate both the state vector $\widehat{\mathbf{X}}(t)$ and also the derivative process $\mathbf{V}^d(t)$. Using GP regression the derivative process can be inferred within a probabilistic framework. Hence, we propose to use a distance function in the derivative space where the state $\widehat{\mathbf{X}}(t)$ and derivative $\mathbf{V}^d(t)$ is modelled using GP regression, within the ABC-SMC algorithm. In this way our proposed method is based on the GP construction in the derivative space as

in [Calderhead et al. \(2008\)](#) and [Wang and Barber \(2014\)](#), combined with the ABC-SMC algorithm as proposed in [Toni et al. \(2009\)](#).

We like to point out the fact that there is a distinction between the approach of [Calderhead et al. \(2008\)](#) to that of [Dondelinger et al. \(2013\)](#) and [Wang and Barber \(2014\)](#). The former approach is a two step method where in a GP is fitted to the data and then the smoothed state and derivative information are used to sample the ODE parameters from a distribution $p(\boldsymbol{\theta}|\widehat{\mathbf{X}}(t), \mathbf{V}^d(t))$. Thus the GP inference process has no information about the ODE dynamics. The latter however induces a coupling between the GP parameters and the ODE parameters by sampling the state from a distribution $p(\widehat{\mathbf{X}}(t)|\boldsymbol{\theta}, \boldsymbol{\phi})$ conditioned on both the GP and ODE hyperparameters. Such a conditioning forces the ODE system dynamics to influence the GP inference. Thus for a very noisy dataset the GP can adapt itself using information from the ODE to better estimate the state trajectories, which in turn improves the ODE parameter inference. Our proposed method is built on the former two step approach as in [Calderhead et al. \(2008\)](#). Next, we will introduce the theory of Gaussian processes for regression and will apply this to the ABC-SMC algorithm.

3.3 Gaussian processes

Gaussian process (GP) is a Bayesian non-parametric model that specifies a distribution over functions, $p(f)$, where f is a function mapping some input space $\mathcal{X}$ to $\mathbb{R}$ ($f : \mathcal{X} \rightarrow \mathbb{R}$) ([O'Hagan and Kingman \(1978\)](#), [MacKay \(1998\)](#), [Neal \(1998\)](#), [Rasmussen and Williams \(2006\)](#)). Let $\mathbf{f} = (f(x_1), \dots, f(x_n))$ be an n -dimensional vector of function values evaluated at n points $x_i \in \mathcal{X}$. Now we can formally define the Gaussian process as follows:

Definition 1. $p(\mathbf{f})$ is a Gaussian process if for any finite subset $\{x_1, \dots, x_n\} \subset \mathcal{X}$, the marginal distribution over the finite subset $p(\mathbf{f})$ has a multivariate Gaussian distribution.

Gaussian processes are characterised by a mean function $\mu(x)$ and a covariance function or kernel, $K(x, x')$, of a real process $f(x)$ as

$$\begin{aligned}\mu(x) &= \mathbb{E}[f(x)], \\ K(x, x') &= \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x')))],\end{aligned}\tag{3.3}$$

and the Gaussian process can be expressed as

$$f(x) \sim \mathcal{GP}(\mu(x), K(x, x')).\tag{3.4}$$

For a finite number of inputs $\mathbf{x} = (x_1, \dots, x_n)$ we can generate a sample function $\mathbf{f}$ ¹ from a GP prior with the covariance matrix $K(\mathbf{x}, \mathbf{x})$. This prior given by

$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}(x), K(\mathbf{x}, \mathbf{x})) \quad (3.5)$$

is simply a multivariate normal distribution following the definition of GP. For example we can sample from a GP with the squared exponential function given by

$$K_{SE}(x, x') = \sigma_{\text{kernel}}^2 \exp\left(\frac{1}{2} \frac{(x - x')^2}{l^2}\right), \quad (3.6)$$

with hyperparameters σ_{kernel}^2 and l^2 (variance and characteristic lengthscale). The length-scale l determines the length of monotonicity in a function. In general it is not possible to extrapolate more than l units. Figure 3.1 shows sample functions drawn from a GP.

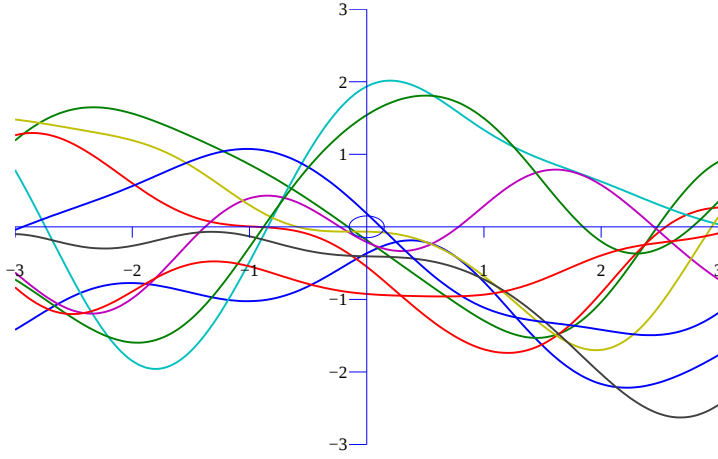


Figure 3.1: Draws from a GP with a squared exponential kernel with hyperparameters σ_{kernel}^2 and l^2 set to 1

Now we can, instead of sampling from the prior distribution, gather information from an observed function $\mathbf{f} = (f(x_1), \dots, f(x_n))$ for a set of n known or training inputs $\mathbf{x} = (x_1, \dots, x_n)$ to predict an unknown function $\mathbf{f}^*$ evaluated at n^* number of test inputs given as $\mathbf{x}^*$. The joint distribution of the training and the test output (functions) is given by:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(\mathbf{x}) \\ \boldsymbol{\mu}(\mathbf{x}^*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{xx}} & \mathbf{K}_{\mathbf{xx}^*} \\ \mathbf{K}_{\mathbf{x}^*\mathbf{x}} & \mathbf{K}_{\mathbf{x}^*\mathbf{x}^*} \end{bmatrix}\right), \quad (3.7)$$

where the shorthand notations² $\mathbf{K}_{\mathbf{xx}}$ and $\mathbf{K}_{\mathbf{x}^*\mathbf{x}^*}$ denote the $n \times n$ and $n^* \times n^*$ covariance matrices associated with the n training and n^* test inputs respectively. $\mathbf{K}_{\mathbf{xx}^*}$ and $\mathbf{K}_{\mathbf{x}^*\mathbf{x}}$

¹Here we term the vector of function values $\mathbf{f} = (f(x_1), \dots, f(x_n))$, evaluated at n points $x_i \in \mathcal{X}$, as a sample function indicating a draw from the Gaussian process.

²These shorthand notations have been introduced to present the GP equations compactly

are the corresponding $n \times n^*$ and $n^* \times n$ cross covariances. Now the important conditional distribution is given as

$$p(\mathbf{f}^*|\mathbf{f}) \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}), \quad (3.8)$$

where the mean function $\mathbf{m}$ and the covariance function $\mathbf{\Sigma}$ are (detailed derivation of these forms have been shown in the [A.1](#) and [A.2](#))

$$\begin{aligned} \mathbf{m} &= \boldsymbol{\mu}(\mathbf{x}^*) + \mathbf{K}_{\mathbf{x}^*\mathbf{x}} \mathbf{K}_{\mathbf{xx}}^{-1} (\mathbf{f} - \boldsymbol{\mu}(\mathbf{x})), \\ \mathbf{\Sigma} &= \mathbf{K}_{\mathbf{x}^*\mathbf{x}^*} - \mathbf{K}_{\mathbf{x}^*\mathbf{x}} \mathbf{K}_{\mathbf{xx}}^{-1} \mathbf{K}_{\mathbf{xx}^*}. \end{aligned} \quad (3.9)$$

3.3.1 Noisy Observations

In almost all practical cases the realistic observations are corrupted by noise. Let us first consider such a case where the observations $\mathbf{y}$ are generated from a vector of latent function values $\mathbf{f}$ evaluated at the training input points $\mathbf{x}$ (mentioned previously), corrupted by additive i.i.d Gaussian noise $\epsilon = \mathcal{N}(0, \sigma_n^2)$. We have

$$\mathbf{y} = \mathbf{f} + \epsilon, \quad (3.10)$$

where,

$$\text{cov}(\mathbf{y}_{\mathbf{p}}, \mathbf{y}_{\mathbf{q}}) = \mathbf{K}_{\mathbf{x}_{\mathbf{p}}\mathbf{x}_{\mathbf{q}}} + \sigma_n^2 \delta_{\mathbf{pq}} \quad (3.11)$$

or

$$\text{cov}(\mathbf{y}) = \mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I}, \quad (3.12)$$

where $\delta_{\mathbf{pq}}$ is Kronecker delta product which is one iff $\mathbf{p} = \mathbf{q}$ and zero otherwise. Now the joint distribution of the noisy observations and a test function $\mathbf{f}^*$ is given as:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}(\mathbf{x}) \\ \boldsymbol{\mu}(\mathbf{x}^*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I} & \mathbf{K}_{\mathbf{xx}^*} \\ \mathbf{K}_{\mathbf{x}^*\mathbf{x}} & \mathbf{K}_{\mathbf{x}^*\mathbf{x}^*} \end{bmatrix} \right). \quad (3.13)$$

The posterior distribution of the test function conditioned on the observations is:

$$p(\mathbf{f}^*|\mathbf{y}) \sim \mathcal{N}(\mathbf{m}', \mathbf{\Sigma}') \quad (3.14)$$

, where

$$\begin{aligned} \mathbf{m}' &= \boldsymbol{\mu}(\mathbf{x}^*) + \mathbf{K}_{\mathbf{x}^*\mathbf{x}} (\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu}(\mathbf{x})), \\ \mathbf{\Sigma}' &= \mathbf{K}_{\mathbf{x}^*\mathbf{x}^*} - \mathbf{K}_{\mathbf{x}^*\mathbf{x}} (\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{K}_{\mathbf{xx}^*}. \end{aligned} \quad (3.15)$$

3.3.2 Maximum Likelihood estimation

The marginal likelihood of the data $p(\mathbf{y}|\Phi)$ is given by:

$$p(\mathbf{y}|\Phi) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\Phi)d\mathbf{f}, \quad (3.16)$$

where Φ is the set of parameters (this includes the hyperparameters of the covariance kernel and the observation noise) and $p(\mathbf{f}|\Phi) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{f}), \mathbf{K}_{\mathbf{xx}})$ is a Gaussian prior distribution. From equations 3.10 and 3.12, the likelihood as the probability density function of the observations conditioned on the latent function is is another normal distribution $p(\mathbf{y}|\mathbf{f}) = \mathcal{N}(\mathbf{f}, \sigma_n^2 \mathbf{I})$. In most practical cases we assume the prior mean $\boldsymbol{\mu}(\mathbf{f}) = 0$. Thus we have

$$\begin{aligned} p(\mathbf{y}|\Phi) &= \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\Phi)d\mathbf{f} \\ &= \int \mathcal{N}(\mathbf{f}, \sigma_n^2 \mathbf{I})\mathcal{N}(0, \mathbf{K}_{\mathbf{xx}})d\mathbf{f}. \end{aligned} \quad (3.17)$$

The above integral can be evaluated analytically and thus we have the marginal likelihood given by

$$p(\mathbf{y}|\Phi) = \mathcal{N}(0, \mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I}), \quad (3.18)$$

and the logarithm of the likelihood is given as

$$\log p(\mathbf{y}|\Phi) = -\frac{1}{2}\mathbf{y}^T(\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log |\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I}| - \frac{n}{2}\log(\pi). \quad (3.19)$$

The gradient of the log likelihood with respect to Φ can be obtained by

$$\begin{aligned} \frac{\partial \log p(\mathbf{y}|\Phi)}{\partial \Phi} &= -\frac{1}{2}\mathbf{y}^T(\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})^{-1}\frac{\partial(\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})}{\partial \Phi}(\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})^{-1}\mathbf{y} \\ &\quad - \frac{1}{2}\text{tr} \left((\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})^{-1}\frac{\partial(\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I})}{\partial \Phi} \right). \end{aligned} \quad (3.20)$$

The maximum likelihood estimate (MLE) can be obtained by solving

$$\frac{\partial \log p(\mathbf{y}|\Phi)}{\partial \Phi} = 0. \quad (3.21)$$

Closed form solution of such equations are intractable and that is why we have to resort to numerical methods. Conjugate gradient is one such algorithm that is frequently used (Shewchuk, 1994) in this context. In our work we have used *minimise*, a MATLAB routine developed in Rasmussen and Nickisch (2010), which is a stable and widely used implementation of the conjugate gradient algorithm for MLE in Gaussian processes.

3.4 Gaussian process regression

The predictive mean function $\mathbf{f}^*$ for noisy observations, assuming the prior mean $\mu(\mathbf{f}^*)$ to be zero, is given by $\mathbf{m} = \mathbb{E}[\mathbf{f}^*] = \mathbf{K}_{\mathbf{x}^*\mathbf{x}}(\mathbf{K}_{\mathbf{xx}} + \sigma_n^2\mathbf{I})^{-1}\mathbf{y}$. To avoid numerical instability direct inversion of the matrix $\mathbf{K}_{\mathbf{y}} = (\mathbf{K}_{\mathbf{xx}} + \sigma_n^2\mathbf{I})$ should be avoided. A more robust alternative is to compute the Cholesky decomposition given by

$$\mathbf{K}_{\mathbf{y}} = \mathbf{L}\mathbf{L}^T. \quad (3.22)$$

We can compute the predictive mean and covariance function as shown in Algorithm 3 given in [Rasmussen and Williams \(2006\)](#). This algorithm which computes the predictive mean and covariance functions, given the training and test data and the hyperparameters to build a covariance function, is known as the Gaussian process regression. It should be noted that the inference of the hyperparameters using the MLE as shown in the previous section is not a part of this algorithm. It takes $O(N^3)$ time to compute the Cholesky decomposition and then $O(N^2)$ time to solve for $\alpha = \mathbf{K}_{\mathbf{y}}^{-1} = \mathbf{L}^{-T}\mathbf{L}^{-1}\mathbf{y}$. The predictive mean and covariance functions can then be computed in $O(N^2)$ time.

Algorithm 3 Gaussian process regression

1. $\mathbf{L} = \text{cholesky}(\mathbf{K}_{\mathbf{xx}} + \sigma_n^2\mathbf{I})$.
 2. $\alpha = \mathbf{L}^T \setminus (\mathbf{L} \setminus \mathbf{y})$.
 3. $\mathbf{m} = \mathbf{K}_{\mathbf{x}^*\mathbf{x}}\alpha$.
 4. $\mathbf{v} = \mathbf{L} \setminus \mathbf{K}_{\mathbf{x}^*\mathbf{x}}$.
 5. $\Sigma = \mathbf{K}_{\mathbf{x}^*\mathbf{x}^*} - \mathbf{v}^T\mathbf{v}$.
-

3.5 Derivative of a Gaussian process

Differentiation being a linear operator, the derivative of a Gaussian process is another Gaussian process ([Solak et al., 2002](#)). This makes it possible to include derivative observation in the GP model, or to compute prediction about derivatives. We have

$$\mathbb{E} \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right] = \frac{\partial \mathbb{E}[\mathbf{f}]}{\partial \mathbf{x}}. \quad (3.23)$$

And likewise the covariance between partial derivative and a function value can be written as

$$K \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}, \mathbf{f}^* \right) = \frac{\partial}{\partial x} K(\mathbf{x}, \mathbf{x}^*), \quad (3.24)$$

and the covariance between partial derivatives follows

$$K \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}, \frac{\partial \mathbf{f}^*}{\partial \mathbf{x}^*} \right) = \frac{\partial^2}{\partial x \partial x^*} K(\mathbf{x}, \mathbf{x}^*). \quad (3.25)$$

For example considering the squared exponential covariance function given in equation 3.6, we can write equation 3.24 as

$$K\left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}, \mathbf{f}^*\right) = -\frac{(x - x^*)}{l^2} \exp\left(\frac{1}{2} \frac{(x - x^*)^2}{l^2}\right) \quad (3.26)$$

The conditional distribution of the derivative of a test function $\frac{\partial \mathbf{f}^*}{\partial \mathbf{x}^*}$ having observed a function $\mathbf{f}$ is given as

$$p\left(\frac{\partial \mathbf{f}^*}{\partial \mathbf{x}^*} | \mathbf{f}\right) = \mathcal{N}(\mathbf{m}, \Sigma), \quad (3.27)$$

where we have the mean function $\mathbf{m}$ and covariance function Σ (obtained by using equation 3.23 and 3.24 in conjunction with equation 3.9 and considering the prior mean of the test and training function to be zero) given as

$$\begin{aligned} \mathbf{m} &= \frac{\partial \mathbf{K}_{\mathbf{xx}^*}}{\partial \mathbf{x}^*} \mathbf{K}_{\mathbf{xx}}^{-1} \mathbf{f} \\ \Sigma &= \frac{\partial^2 \mathbf{K}_{\mathbf{x}^* \mathbf{x}^*}}{\partial \mathbf{x}^* \partial \mathbf{x}^*} - \frac{\partial \mathbf{K}_{\mathbf{x}^* \mathbf{x}}}{\partial \mathbf{x}^*} \mathbf{K}_{\mathbf{xx}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{xx}^*}}{\partial \mathbf{x}^*}. \end{aligned} \quad (3.28)$$

Likewise for noisy observations we have

$$\begin{aligned} \mathbf{m} &= \frac{\partial \mathbf{K}_{\mathbf{x}^* \mathbf{x}}}{\partial \mathbf{x}^*} [\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{f} \\ \Sigma &= \frac{\partial^2 \mathbf{K}_{\mathbf{x}^* \mathbf{x}^*}}{\partial \mathbf{x}^* \partial \mathbf{x}^*} - \frac{\partial \mathbf{K}_{\mathbf{x}^* \mathbf{x}}}{\partial \mathbf{x}^*} [\mathbf{K}_{\mathbf{xx}} + \sigma_n^2 \mathbf{I}]^{-1} \frac{\partial \mathbf{K}_{\mathbf{xx}^*}}{\partial \mathbf{x}^*}. \end{aligned} \quad (3.29)$$

3.6 ABC-SMC with Derivative GP

In this section we apply the machinery presented in the previous sections to the task of inferring parameters in differential equation models whose solution is the state trajectory $\mathbf{X}(t)$. If we assign a GP prior to the state evaluated at time points $t^L \triangleq \{t_i\}_{i=1, \dots, L}$, then the set of values of the state $\mathbf{X}(t^L)$ takes on a Gaussian prior distribution:

$$p(\mathbf{X}(t^L) | t^L) = \mathcal{N}(\mathbf{X}(t^L) | 0, K(t^L, t^L)). \quad (3.30)$$

The modelling task is to represent the experimental data as $Y^d = \{\mathbf{X}(t^L) + \boldsymbol{\eta}^L\}$ where $\boldsymbol{\eta}^L$ refers to L i.i.d. samples from $\mathcal{N}(0, \sigma^2 \mathbb{I}_L)$, here $\mathbb{I}_L$ is a $L \times L$ identity matrix, and $\mathbf{X}(t)$ satisfies a differential equation. We can use GP regression to obtain the expectation and variance of the posterior (given training data Y^d at t^L) state $\mathbf{X}(t^*)$ for some test input time point t^* as in section 3.4 (Rasmussen and Williams, 2006):

$$\begin{aligned} E[\mathbf{X}(t^*) | Y^d] &= K(t^*, t^L) (K(t^L, t^L) + \sigma^2 \mathbb{I})^{-1} Y^d, \\ Var[\mathbf{X}^*] &= K(t^*, t^*) - K(t^*, t^L) (K(t^L, t^L) + \sigma^2 \mathbb{I})^{-1} K(t^L, t^*). \end{aligned} \quad (3.31)$$

This expected posterior state variable $\mathbf{X}(t^*)$ for arbitrary choice of t^* models the smoothed evolution of the state $\widehat{\mathbf{X}}(t)$ introduced above, and where it is assumed that observational

noise accounts for deviations from the smoothed time course. The smoothed state estimation enables us to compute the velocity field using the derivative GP (as in section 3.5):

$$E\left[\frac{d}{dt}\mathbf{X}\right] = \frac{\partial K(t, t)}{\partial t} (K(t, t) + \sigma^2 \mathbb{I})^{-1} E[\mathbf{X}]. \quad (3.32)$$

This completes the procedure for deriving the empirical velocity field (equation 3.1) $\mathbf{V}^d(t) = E\left[\frac{d}{dt}\mathbf{X}\right]$.

3.6.1 Illustrations through an example model

Having observed a noisy experimental data for which we have an explanation in the form of an ODE or DDE our first task in hand is to calculate the velocity field $\mathbf{V}^d(t)$ using equation 3.31 and 3.32. To illustrate the intermediate stages we chose a model system: The Mackey-Glass delay differential equation given by

$$\dot{x} = \beta \frac{x_{t_d}}{1 + x_{t_d}^n} - \gamma x, \quad \gamma, \beta, n > 0 \quad (3.33)$$

where β , γ , t_d and n are parameters, and x_{t_d} represents the value of the variable x at time $(t - t_d)$. Glass and Mackey (1979) proposed this delay differential equation to model non-linear feedback control in physiology. The above equation displays several interesting dynamic behaviours such as limit cycle oscillations and aperiodic solutions. Furthermore for certain combination of the parameters, $\gamma = 1$, $\beta = 2$, $t_d = 2$ and $n = 9.65$, this system exhibits deterministic chaos. The complex oscillatory pattern exhibited by the chaotic state trajectory x poses a crucial challenge to both the state and velocity estimation processes in presence of observational noise. Thus we consider a noise corrupted version of x as an ideal test example to elucidate the velocity estimation process. We generated 100 samples between the interval $[0 : 0.1 : 10]$, from the chaotic regime of the system with mentioned parameter values that can generate chaos. We further added random noise with standard deviation $\sigma_x = 0.1$ to create an artificial experimental data. Figure 3.2 shows the noisy data.

Because of the presence of multiple time-scales of oscillations and chaos we have chosen a GP prior with the following covariance kernel called the Matern kernel:

$$K_{Matern}(t, t') = \left(1 + \frac{\sqrt{3}(t - t')}{l}\right) \exp\left(-\frac{\sqrt{3}(t - t')}{l}\right), \quad (3.34)$$

where the hyperparameter l denotes lengthscale. This is a stationary covariance with lesser smoothness assumptions on the function to be estimated. Using GP regression as in Algorithm 3 along with equation 3.31 we estimate the smoothed state trajectory. This is plotted in figure 3.3 along with the estimated confidence region. Finally using equation 3.32 we estimate the velocity field or the derivative process. The estimation

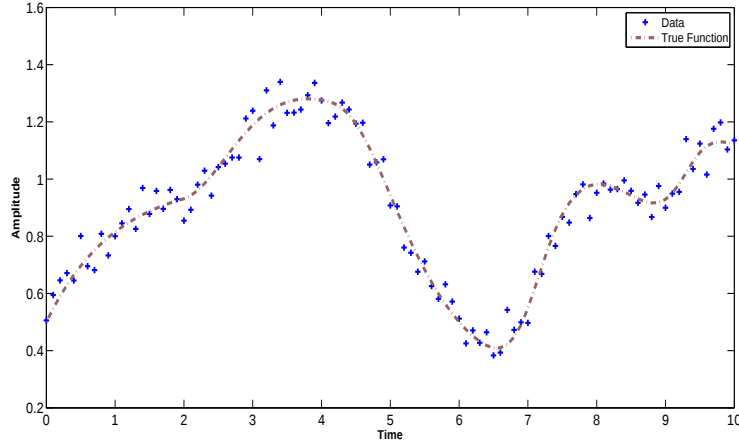


Figure 3.2: Noisy state trajectory of the Mackey-Glass model.

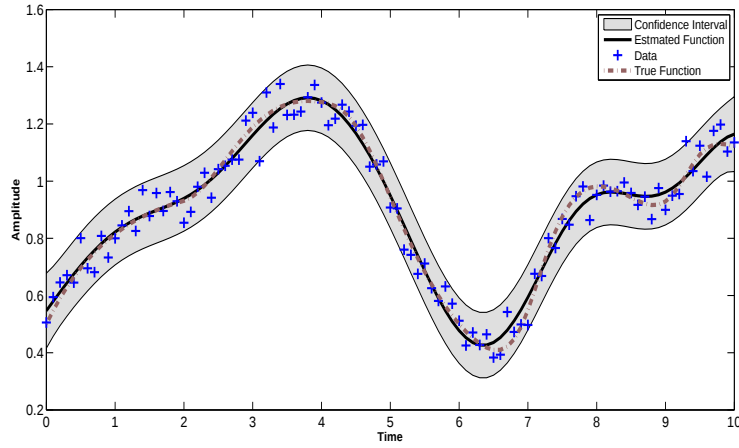


Figure 3.3: The estimated smoothed state trajectory of the Mackey-Glass model.

of the state and its derivative is compared with the ground truth in both figure 3.3 and figure 3.4 respectively.

3.6.2 The GP-ABC-SMC algorithm

To apply this derivative process within the ABC framework we need to define a distance metric $\Delta(\mathbf{V}^d(t), \mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta}))$ between the smoothed velocity field derived from the observed data, and the velocity field postulated in a differential equation model, where the expected state estimation $\widehat{\mathbf{X}}(t)$ has been substituted for the state variable. Hence our proposed fast alternative ABC-SMC based on GP gradient distance (GP-ABC-SMC) works as follows:

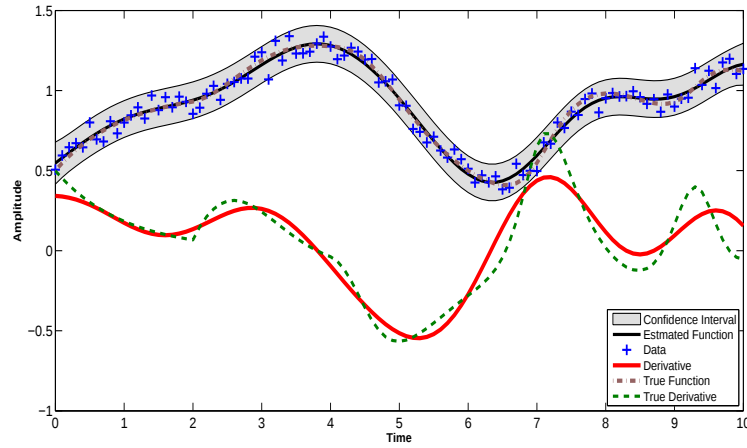


Figure 3.4: The estimated velocity field of the Mackey-Glass model.

1. Having given data Y^d as a noisy observation of the true state variable $\mathbf{X}(t)$, assign a GP prior on $\mathbf{X}(t)$ using (equation 3.2) and choose a covariance function, with some unknown hyperparameters, needed to define the GP prior.
2. Learn the hyperparameters of the covariance function from the original noisy experimental data Y^d using maximum likelihood estimation and then run GP regression to obtain an estimation of the smoothed state evolution $\widehat{\mathbf{X}}(t) = \mathbb{E}[\mathbf{X}]$ using (equation 3.31) and the experimental time points t as both the training and test input points.
3. Construct the first derivative of the covariance matrix and estimate the derivative process $\mathbf{V}^d(t) = \mathbb{E}[\frac{d}{dt}\mathbf{X} |_{\mathbf{X}=\widehat{\mathbf{X}}}]$ using (equation 3.32).
4. Run the ABC-SMC algorithm with a modified distance metric $\Delta(\mathbf{V}^d(t), \mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta})) \leq \epsilon_\tau$ for tolerance schedule $\{\epsilon_\tau\}$, where at each simulation step the simulated data $Y^s = \mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta})$ is generated by evaluating the velocity field on the right hand side of the differential equation. This yields the posterior distribution of the parameters $p(\boldsymbol{\theta} | Y^d)$ (equation 2.9).

Note that we cannot interchange the order of expectation with the non-linear function $\mathbf{f}$ appearing on the right hand side of any non-linear ODE:

$$\mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta}) = \mathbf{f}(\mathbb{E}[\mathbf{X}(t)], \boldsymbol{\theta}) \neq \mathbb{E}[\mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta})], \quad (3.35)$$

as we did while deriving the empirical velocity field $\mathbf{V}^d(t)$. For this reason the GP-ABC-SMC as described above does not propagate the uncertainty associated with the state $\mathbf{X}(t)$ and employs the GP as an interpolant. However, we can modify the GP-ABC-SMC distance function to include $\mathbb{E}[\mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta})] = \frac{1}{K} \sum_{k=1}^K \mathbf{f}(\mathbf{X}^k(t), \boldsymbol{\theta})$ where $\mathbf{X}^k$ is sampled from the posterior of the GP on the state vector with mean and variance given in equation 3.31.

Since we only need to invert the covariance matrix once for sampling the posterior states thus the computational complexity for taking the expectation of the right hand side of an ODE is $\mathcal{O}(KL)$, where K is the number of state samples and L is the number of time points. We have used the interpolating version of the GP-ABC-SMC in our experiments and avoided the full uncertainty propagation to achieve maximum speedup.

Also note that no explicit solution of differential equation is required to generate the simulated data within the iterations of the GP-ABC-SMC algorithm. Also note the fact that, in order to run the GP-ABC-SMC algorithm, no knowledge of the initial condition is required. For the sake of conformity with the ABC terminologies, we will persist in using the terms Y^d for observed data and Y^s for simulated data, as before. However, within the context of GP-ABC-SMC algorithm, observed and simulated data refer to $\mathbf{V}^d(t)$ and $\mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta})$ respectively. That is what Y^d and Y^s will refer to for the rest of the paper.

For the specific case of gradient matching applications, such as the proposed GP-ABC-SMC, it may so happen that the experimental time points are irregularly spaced. Then the covariance pertaining to the training points can become ill-conditioned. In such cases we can choose the training and test time points $t_{equidist}, t_{equidist}^*$ in (equation 3.31) and (equation 3.32) to be equidistant to perform the GP regression, although the experimental time points t^L are irregularly spaced. The resultant state $\mathbf{X}(t_{equidist})$ and derivative $\mathbf{V}^d(t_{equidist})$ trajectories can be used along with the right hand side $\mathbf{f}(\widehat{\mathbf{X}}(t_{equidist}), \boldsymbol{\theta})$ to create the distance metric $\Delta(\mathbf{V}^d(t_{equidist}), \mathbf{f}(\widehat{\mathbf{X}}(t_{equidist}), \boldsymbol{\theta})) \leq \epsilon_\tau$ without using an ill-conditioned covariance matrix altogether.

3.6.3 The GP-ABC-SMC algorithm for DDEs

In the case of a DDE model given by

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{f}(\mathbf{X}(t), \mathbf{X}_{t_d}; \boldsymbol{\theta}), \quad (3.36)$$

the delayed state $\mathbf{X}_{t_d} = \mathbf{X}(t - t_d)$ consists of two parts: $\mathbf{X}(t \geq 0)$ and $\mathbf{X}(t < 0)$. Applying GP-ABC-SMC in this case requires the smoothed estimate of this delayed state vector, which we denote here as $\widehat{\mathbf{X}}_{t_d}$, including both its parts to compute the model right hand side $\mathbf{f}(\widehat{\mathbf{X}}(t), \widehat{\mathbf{X}}_{t_d}; \boldsymbol{\theta})$ in order to build the ABC distance function. For applying gradient matching algorithms, such as ours, this delayed state vector is obtained from the smoothed estimate $\widehat{\mathbf{X}}(t)$. However, unlike an ODE system we cannot avoid the estimation of the initial values of the delayed state while applying the GP-ABC-SMC. Specifically, we need to estimate the initial history function $\widehat{\mathbf{X}}_h = \widehat{\mathbf{X}}(t < 0)$, which is not available from the smoothed estimate of the state $\widehat{\mathbf{X}}(t \geq 0)$. In most practical cases the initial history function is taken as a constant function. For such cases we can obtain the estimated initial history function as a vector with constant elements obtained by copying

the first element of the smoothed state estimate as obtained by the GP regression:

$$\{\widehat{\mathbf{X}}_h(t_i)\}_{i=1,\dots,D} = \widehat{\mathbf{X}}(t_0), \quad (3.37)$$

where $D = \frac{t_d}{t_i - t_{i-1}}, i = 1, \dots, L$ is delay in terms of discrete samples. Using this as the initial history we can then use the first $L - D$ elements of the smoothed state $\widehat{\mathbf{X}}_{lagged} = \{\widehat{\mathbf{X}}(t_i)\}_{i=1,\dots,L-D}$ to create the delayed state vector as:

$$\widehat{\mathbf{X}}_{t_d} = \left(\widehat{\mathbf{X}}_h, \widehat{\mathbf{X}}_{lagged} \right)^T. \quad (3.38)$$

Having obtained the delayed state vector the rest of the steps of GP-ABC-SMC remain the same as in the case of ODEs. This method of estimating the delayed state throws away all information pertaining to $\mathbf{X}(t < 0)$ for models where the initial history function is anything other than a constant. In such cases GP regression can be used to extrapolate the history vector using covariance kernels specifically designed for extrapolation proposed in [Wilson and Adams \(2013\)](#).

3.6.4 Working with hidden variables

The GP-ABC-SMC algorithm can be extended to the case of partially observed system where some of the state variables cannot be measured. Let us consider the state vector $\mathbf{X} = (X_o, X_h)$, consisting of an observed $X_o(t)$ and a hidden $X_h(t)$ state variable. We can expand the parameter vector by including an additional parameter for the initial value of the hidden state $X_h(t = 0)$. We would sample this initial value along with other parameters and then evolve the hidden state by using the GP mean function estimate of the visible state $X_o(t)$ using Euler or Runge-Kutta steps. In this way we can construct the entire smoothed state vector $\widehat{\mathbf{X}} = (\widehat{X}_o, \widehat{X}_h)$ and subsequently the right hand side of the model $Y^s = \mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta})$. Finally we create the ABC distance function between the empirical velocity field $\mathbf{V}^d(t) = E[\frac{d}{dt}X_o(t)]$ obtained from the observed state and right hand side of the model $Y^s = \mathbf{f}((\widehat{X}_o(t), \widehat{X}_h(t)), \boldsymbol{\theta})$ using the smoothed estimate of $\widehat{X}_o(t)$ and the Euler construction of $X_h(t)$.

3.6.5 Comparison with other GP based algorithms

It was mentioned previously (see section 3.2) that the proposed GP-ABC-SMC algorithm is similar to [Calderhead et al. \(2008\)](#) in terms of a two step approach. First an interpolant is fit to data and then the interpolant's gradient is matched to the model function. However, there exist fundamental methodological distinctions between how GP-ABC-SMC matches the gradient to that of [Calderhead et al. \(2008\)](#) and [Dondelinger et al. \(2013\)](#). In [Calderhead et al. \(2008\)](#) the gradient matching is done by combining

the conditional distribution of the derivative $p(\dot{\widehat{\mathbf{X}}}|\widehat{\mathbf{X}}, \phi)$ given the GP and the model function $p(\dot{\widehat{\mathbf{X}}}|\mathbf{f}(\widehat{\mathbf{X}}, \boldsymbol{\theta}), \gamma)$ as:

$$p(\dot{\widehat{\mathbf{X}}}|\widehat{\mathbf{X}}, \boldsymbol{\theta}, \phi) \propto p(\dot{\widehat{\mathbf{X}}}|\widehat{\mathbf{X}}, \phi)p(\dot{\widehat{\mathbf{X}}}|\mathbf{f}(\widehat{\mathbf{X}}, \boldsymbol{\theta}), \gamma). \quad (3.39)$$

The above construction, known as product of experts ([Mayraz and Hinton, 2002](#)), introduces an unnatural noise term γ to make up for the discrepancy between the model function evaluated on the true state $\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$ and the smoothed state $\mathbf{f}(\widehat{\mathbf{X}}, \boldsymbol{\theta})$. Marginalizing over $\dot{\widehat{\mathbf{X}}}$, the joint distribution of the observations, the smoothed state and the parameters is then given by

$$p(Y^d, \widehat{\mathbf{X}}, \boldsymbol{\theta}, \phi, \gamma) = p(\boldsymbol{\theta}, \gamma|\widehat{\mathbf{X}}, \phi)p(\widehat{\mathbf{X}}|\phi, Y^d)p(\phi|Y^d)p(Y^d). \quad (3.40)$$

The term $p(\boldsymbol{\theta}, \gamma|\widehat{\mathbf{X}}, \phi)$ is problematic because of the conditioning of $\boldsymbol{\theta}$ on the smoothed state $\widehat{\mathbf{X}}$ obtained by the GP regression. In [Dondelinger et al. \(2013\)](#) the joint distribution $p(Y^d, \widehat{\mathbf{X}}, \boldsymbol{\theta}, \phi, \gamma)$ is factored as ³:

$$p(Y^d, \widehat{\mathbf{X}}, \boldsymbol{\theta}, \phi, \gamma) = p(Y^d|\widehat{\mathbf{X}})p(\widehat{\mathbf{X}}|\boldsymbol{\theta}, \phi, \gamma)p(\boldsymbol{\theta})p(\phi)p(\gamma). \quad (3.41)$$

The term $p(\widehat{\mathbf{X}}|\boldsymbol{\theta}, \phi, \gamma)$ introduces a coupling between the parameters of the GP and the ODE, thus no longer being a two step approach. However, both these approaches are essentially built upon the product of experts assumption that introduces a fictitious noise term γ . Our ABC approximation on the other hand is bereft of any such approximate noise models and probabilistic relationship between $\dot{\widehat{\mathbf{X}}}$ and $\mathbf{f}(\widehat{\mathbf{X}}, \boldsymbol{\theta})$. We only assume that in the absence of noise both these quantities should be same. In the presence of noise we rely upon the ABC approximation that a good match between the two gives us the posterior approximation that is close to the true posterior distribution.

The methodology found in both [Calderhead et al. \(2008\)](#) and [Dondelinger et al. \(2013\)](#) involves sampling of the GP posterior states $\widehat{\mathbf{X}}$ at each step of the MCMC involving the inversion of a $L \times L$ matrix, L being the number of experimental points. Moreover, these methods need to sample nuisance parameters ϕ, γ unrelated to the original inference problem. GP-ABC-SMC avoids these extra sampling (and computation) at the cost of no coupling between ϕ and $\boldsymbol{\theta}$.

In [Wang and Barber \(2014\)](#) gradient matching, using the product of experts, is abandoned by conditioning the data Y^d directly on $\dot{\widehat{\mathbf{X}}}$. This is expressed as

$$p(Y^d|\dot{\widehat{\mathbf{X}}}) = \int p(Y^d|\widehat{\mathbf{X}})p(\widehat{\mathbf{X}}|\dot{\widehat{\mathbf{X}}}), \quad (3.42)$$

³refer to [Dondelinger et al. \(2013\)](#) for the derivation of equation 3.41 and 3.40

where an implicit numerical integration is performed through the term $p(\widehat{\mathbf{X}}|\dot{\widehat{\mathbf{X}}})$ ⁴. Although this improves the weak modelling assumptions in [Calderhead et al. \(2008\)](#); [Dondelinger et al. \(2013\)](#), sampling of posterior states and nuisance parameter is still required. Furthermore, this method unlike [Calderhead et al. \(2008\)](#); [Dondelinger et al. \(2013\)](#) and GP-ABC-SMC performs the estimation of initial state X_0 , rendering this algorithm not so useful for DDEs.

3.7 Conclusion

In this chapter we have proposed a method that could significantly speed up the task of parameter inference in comparison with state of the art methods that use ABC and SMC based approaches when applied to dynamical system models that are described by ordinary and delay differential equations. We achieve this speed-up by circumventing the need to numerically integrate the differential equations, a task that is repeatedly required in other ABC methods to generate samples from candidate models for comparison with the observed data. The key idea behind our method lies in building on [Calderhead et al. \(2008\)](#), [Ramsay et al. \(2007\)](#) and [Wang and Barber \(2014\)](#) to work directly with the vector field of the dynamical system, which we model using Gaussian process regression, and thus create a distance function in derivative space for use in the ABC-SMC algorithm as proposed in [Toni et al. \(2009\)](#). Thus we proposed a modified ABC-SMC algorithm for parameter estimation (and can be trivially extended to model selection) in ODEs or DDEs. Furthermore, improvements of ABC-SMC through perturbation kernels as proposed in [Filippi et al. \(2013\)](#) could be integrated with our approach to obtain enhanced performance. Note that Gaussian processes can accommodate other noise models ([Rasmussen and Williams, 2006](#)) which are more complex than the zero mean i.i.d Gaussian noise. Thus for real data more flexibility can be introduced while modelling the state and its derivatives.

Interestingly, saving time in simulation could potentially open up space for exploring other algorithmic settings. For example, a variety of tolerance schedule and perturbation kernels can be explored to reach the best estimation. These sort of explorations remain computationally prohibitive without transforming the problem into the derivative space as in our approach.

Our proposed approach is fundamentally limited by the ability of the Gaussian process regression in smoothing the observed time series data while retaining the essential characteristics that are meant to be captured by the dynamical system model. Thus in those cases where smoothing the experimental data by GP regression introduces artefacts, the GP-ABC-SMC algorithm is more likely to produce poor parameter estimates.

⁴see [Wang and Barber \(2014\)](#) for further details of this implicit integration

In the next chapter we will apply the GP-ABC-SMC algorithm to the task of estimating posterior parameter and model probabilities of various ODE and DDE model systems. Through these experiments we will highlight the benefits of using our proposed GP based ABC-SMC compared to ABC-SMC with explicit integration and other GP based MCMC methods.

Chapter 4

Experimental evaluation of the GP-ABC-SMC algorithm

4.1 Introduction

To evaluate the GP-ABC-SMC algorithm we have chosen five benchmarking differential equations: The Lotka Volterra predator-prey model (Murray, 2002), the Hes1 loop model (Monk, 2003), signal transduction cascade model (Vyshemirsky and Girolami, 2008; Wang and Barber, 2014), a family of SIR models (Anderson et al., 1991) and the Mackay-Glass model (Glass and Mackey, 1979) that we presented in the previous chapter. Each is a set of non-linear differential equations modelling biological systems and show non-trivial dynamical phenomena such as limit cycle oscillations, chaos and non-stationary time evolution. For all these examples we have used the distance function given by (equation 2.8) and have run the ABC-SMC algorithm with explicit integration using a component-wise univariate normal kernel (ABC-SMC-Comp) (Toni et al., 2009) as well as a multivariate normal kernel with the optimal local covariance matrix (ABC-SMC-OLCM) (Filippi et al., 2013). For our proposed GP-ABC-SMC we have also used both the aforementioned perturbation kernels. We refer these as the GP-ABC-SMC and GP-ABC-OLCM respectively. We believe a comparison between these four variants of ABC-SMC is required to capture the difference in speed of execution between the GP based ABC-SMC and the previous approaches reported in (Toni et al., 2009; Filippi et al., 2013), while comparing posterior estimates of the parameters. For all the examples presented here, we ran all these variants of ABC-SMC, including the proposed GP based ones, with $N = 100$ particles using an adaptive tolerance schedule set to the $\alpha = 0.1$ quantile of the distances in the previous populations. Note that we have used such a small number of particles primarily to expedite the simulation time for the Hes1 DDE using explicit integration and have maintained the same population size for other problems as well. Our goal here is the comparison between different variants of ABC-SMC for which

we have found $N = 100$ particles being of adequate size. We arrive at this conclusion after repeating many of the following experiments with higher particle size ($N = 10000$) and obtaining similar estimates.

The ABC-SMC routines are written in MATLAB and for the GP regressions the GPML package (Rasmussen and Nickisch, 2010) for MATLAB is used in the predator-prey, Mackay-Glass, SIR and Hes1 loop example. For the signal transduction cascade model the GPML toolbox for MATLAB <https://github.com/SheffieldML/GPml> is used which has an implementation of the multi-layer perceptron (MLP) covariance kernel (Wang and Barber, 2014), required to handle the non-stationarity of some of the state variables. The explicit integrations are carried out using MATLAB's built in ODE and DDE solver routines.

4.1.1 ODE: The predator prey model

The Lotka Volterra (Murray, 2002) model depicts an ecological system that is used to describe the interaction between a predator and prey species. This ODE given by

$$\begin{aligned}\dot{x} &= \alpha x - xy \\ \dot{y} &= \beta xy - y,\end{aligned}\tag{4.1}$$

shows limit cycle behaviour and has been used for benchmarking in (Toni et al., 2009; Dondelinger et al., 2013). $\theta = (\alpha, \beta)$ is the set of parameters and $\mathbf{X}(t) = (x(t), y(t))$ is the state vector comprising the concentrations of the predator and the prey species respectively. To create a realistic dataset we generated 11 uniformly spaced samples between the time interval ($0 \leq t \leq 10$) from the model with parameters $\theta = (1, 1)$ and added random Gaussian noise with zero mean and standard deviation $\sigma = 0.5$ to each point. The initial values of the ODE for generating the synthetic data are chosen as $\mathbf{X}(t = 0) = (1.0, 0.5)$. In order to inspect the consistency of our proposed algorithm we created two more datasets obtained by adding two other realizations of the random noise to the ODE time courses. Thus we have three sets of artificial data (denoted as Dataset 1, 2 and 3), each of which has been corrupted by Gaussian noise with zero mean and standard deviation $\sigma = 0.5$ and sampled separately. Note that the GP-ABC-SMC algorithm does not require the estimation of additional nuisance parameters related to the initial values. The time evolution of the state and its derivative is predicted through the GP regression as described in section 3.6. We have used the squared exponential covariance function given by (equation 3.6) for the GP regression in this example.

From the synthetic data we perform the task of parameter inference using the four different variants of ABC-SMC discussed in the last section to compare their performance. Both α and β are chosen from uniform prior distributions $\mathcal{U}(-10, 10)$ in all cases. The number of algorithmic iterations, the value of S_{MC} is set to $S_{MC} = 6$ for the

ABC-SMC-Comp and ABC-SMC-OLCM while it is set to $S_{MC} = 5$ for GP-ABC-SMC and GP-ABC-OLCM. The values differ because we have chosen these on the criteria of minimum number of adaptive iteration required for estimating a reliable posterior distribution. As the ABC-SMC with integration and the GP based variants operate on different spaces thus setting same values for S_{MC} does not produce comparable results. The specific values of S_{MC} for this and subsequent examples are chosen on the basis of multiple trials of all the four ABC-SMC algorithm on each of the datasets. The chosen value of S_{MC} produce similar posterior marginal densities, in most cases, if the corresponding variant of ABC-SMC is run multiple times on the same dataset.

The resulting parameter estimates are listed in Table 4.1 and the evaluation of the performance in Table 4.2 (top). We show in Table 4.1 the mean and standard deviation of the last population of parameters, approximating the marginal posterior, for each variants of the ABC-SMC. The approximate marginal posterior distributions of each of the two parameters learnt by the GP-ABC-OLCM and the ABC-SMC-OLCM are shown in Figure 4.1. The histograms in Figure 4.1 are based on the final particle populations generated by the respective algorithms after they are run on Dataset 1. Note that the

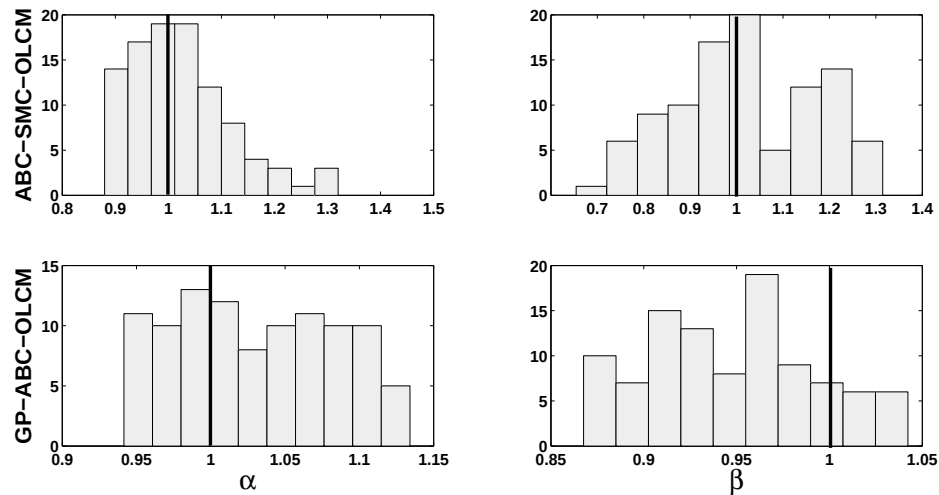


Figure 4.1: Histograms of the final particle populations generated by the GP-ABC-OLCM and the ABC-SMC-OLCM algorithms, approximating the marginal posterior distribution for each of the parameters of Lotka Volterra model. The black line marks the ground truth in each plot.

run-time of the GP-ABC-SMC and the GP-ABC-OLCM algorithms is the sum of the run-time of the ABC and the GP regression (including the estimation of covariance hyperparameters). The value of σ is estimated as part of the GP regression. These estimated values are $\sigma = \{0.4752, 0.8090\}$, $\sigma = \{0.6219, 0.3940\}$ and $\sigma = \{0.6432, 0.4592\}$ for the dataset 1, 2 and 3 respectively. In Figure 4.1 the marginal densities differ between the ABC-SMC-OLCM and the GP-ABC-OLCM variants. We believe the reason behind this being the fact that although both these algorithms are working on the same parameter space the state and the velocity field show different sensitivities to parameter

Table 4.1: Estimated parameters of the Lotka Volterra predator-prey model denoted by the mean and standard deviation of the particles of the final population for datasets 1-3 respectively in each row.

Parameters	True value	ABC-SMC-Comp	ABC-SMC-OLCM	GP-ABC-SMC	GP-ABC-OLCM
α	1	1.0688 ± 0.0108	1.0722 ± 0.0494	1.0373 ± 0.0410	1.0356 ± 0.0359
		1.0389 ± 0.0318	1.0305 ± 0.0277	1.0843 ± 0.0472	1.0718 ± 0.0523
		1.0293 ± 0.0805	1.0421 ± 0.0374	1.0103 ± 0.0882	1.0144 ± 0.0713
β	1	0.9698 ± 0.0128	0.9763 ± 0.0792	0.9567 ± 0.0426	0.9540 ± 0.0344
		0.9749 ± 0.0477	0.9966 ± 0.0421	0.9846 ± 0.0410	0.9760 ± 0.0426
		1.0103 ± 0.1056	0.9924 ± 0.0487	0.9887 ± 0.0792	0.9906 ± 0.0646

Table 4.2: Run-time and the ratio of total number of particles accepted (θ^{**} in Algorithm 1) to that of generated (θ^*) for the four ABC-SMC algorithms when applied to the three artificial datasets pertaining to the Lotka Volterra predator-prey model (left) and Hes1 model (right). The values for run-time are rounded to nearest integers.

Algorithms	Run-time (seconds)	Accept/Generate
ABC-SMC-Comp	397	700/14737
	477	600/12294
	516	500/13381
ABC-SMC-OLCM	184	800/7846
	221	700/6369
	212	600/6086
GP-ABC-SMC	25	500/7650
	26	500/7547
	26	500/7642
GP-ABC-OLCM	21	500/4655
	20	500/4316
	16	500/3193
Algorithms	Run-time (seconds)	Accept/Generate
ABC-SMC-Comp	106980	1300/763045
	840330	1600/8026943
	201710	1600/1656197
ABC-SMC-OLCM	5496	1300/31342
	8399	1500/50968
	5999	1400/36519
GP-ABC-SMC	30	1100/31439
	38	1000/38911
	32	1000/36521
GP-ABC-OLCM	18	1000/7387
	18	1000/8183
	17	1000/7969

changes. Difference in sensitivity results in different adaptive influences on the learning updates: that is how the particles are updated through the filtering steps of SMC. Thus, it is inevitable that the marginal densities (collection of the final particles) will be noticeably different. In [Transtrum et al. \(2011\)](#); [Gutenkunst et al. \(2007\)](#) it is argued that model sensitivity is related to the geometric structure of the likelihood surface. In ABC the likelihood surface is the same as the surface spanned by the distance function for normally distributed observational errors. In ABC-SMC we encode our prior assumption about the geometric structure of the distance surface through a tolerance schedule. Further enquiries about the intrinsic geometric properties of these (involving state and velocity fields) surfaces are necessary for designing better tolerance adaptation schemes that could lead towards estimating the true (and thus similar) posterior densities irrespective of the sensitivities of the state and velocity field. However, in practice no such geometrically aware method for designing tolerance schedules exist. We like to point out that we have also investigated the role of the particle population size N varying it from $N = 100$ to $N = 10000$ without noticing convergence among the state and derivative space ABC-SMC algorithms reinforcing our aforementioned conclusions.

4.1.2 DDE: The Hes1 model

Our proposed algorithm is also able to estimate parameters of delay differential equations. The Hes1 model system is used in systems biology to provide a simplified account of the oscillatory behaviour of the concentrations $(\mu(t), p(t))$ of a species of mRNA and its corresponding protein. The model, introduced in [Monk \(2003\)](#), is described by the following delay differential equations:

$$\begin{aligned}\dot{\mu} &= \frac{1}{1 + (p(t - t_d)/p_0)^n} - \mu_m \mu \\ \dot{p} &= \mu - \mu_p p,\end{aligned}\tag{4.2}$$

where the parameters μ_m and μ_p are decay rates, p_0 is the repression threshold, n is the Hill coefficient and t_d is the time delay. We generated data from the above model with parameters $\mu_m = 0.03$, $\mu_p = 0.03$, $p_0 = 100$ and $t_d = 25$ and initial conditions $\mu(t_0) = 3$ and $p(t_0) = 3$ for the concentrations between the interval $(0 \leq t \leq 300)$ with uniform spacing of $\Delta t = 2$ by numerically solving the DDE. n is fixed at a value of 5 ([Monk, 2003](#)). We estimated the standard deviations $\sigma_\mu = 6.0020$ and $\sigma_p = 121.7670$ of the generated data, for each of the concentrations $\mu(t)$ and $p(t)$. We then added noise, with standard deviation set to 0.1 times these estimated standard deviations σ_μ and σ_p , to the data to create the artificial datasets. As in the previous example we created three datasets in a similar fashion.

For comparison of performance of the four methods in the parameter estimation task, we keep the same algorithmic settings, as well as the same covariance function for the GP

regression as in the previously example. Unlike the ODE case where our algorithm does not need to guess the initial state values, it does need a history function for $\mathbf{X}(t \leq 0)$ for DDEs in order to work. In most practical cases the initial history function is taken as a constant function. Thus in order to make our algorithm work, we shifted the first element of the estimated state evolution backward in time to create the history function. The four variants of ABC-SMC having the same settings as before, are applied to this artificial dataset. We chose uniform priors for each of the parameters: $\mu_m \sim \mathcal{U}(-2, 2)$, $\mu_p \sim \mathcal{U}(-2, 2)$, $p_o \sim \mathcal{U}(0, 200)$ and $t_d \sim \mathcal{U}(0, 50)$. The number of iterations are chosen as $S_{MC} = 14$ while running ABC-SMC-Comp and ABC-SMC-OLCM. For GP-ABC-SMC and GP-ABC-OLCM this is chosen as $S_{MC} = 9$. As in the previous example these S_{MC} values are also found through multiple trials of each of the algorithms on these datasets.

Table 4.3: Estimated parameters of the Hes1 loop model.

Parameters	True value	ABC-SMC-Comp	ABC-SMC-OLCM	GP-ABC-SMC	GP-ABC-OLCM
μ_m	0.03	0.0307 ± 0.0011	0.0305 ± 0.0015	$0.0295 \pm 7.1431 \times 10^{-4}$	$0.0293 \pm 5.7062 \times 10^{-4}$
		$0.0341 \pm 3.3622 \times 10^{-4}$	$0.0342 \pm 2.2271 \times 10^{-4}$	0.0304 ± 0.0014	0.0302 ± 0.0012
		$0.0336 \pm 2.1358 \times 10^{-4}$	$0.0336 \pm 4.3991 \times 10^{-4}$	$0.0291 \pm 9.6005 \times 10^{-4}$	$0.0292 \pm 9.5677 \times 10^{-4}$
μ_p	0.03	0.0294 ± 0.0010	0.0297 ± 0.0015	$0.0300 \pm 1.9791 \times 10^{-5}$	$0.0300 \pm 1.4669 \times 10^{-5}$
		$0.0267 \pm 2.2945 \times 10^{-4}$	$0.0267 \pm 1.5221 \times 10^{-4}$	$0.0300 \pm 2.1089 \times 10^{-5}$	$0.0300 \pm 1.5644 \times 10^{-5}$
		$0.0268 \pm 1.4991 \times 10^{-4}$	$0.0268 \pm 3.0731 \times 10^{-4}$	$0.0297 \pm 1.8138 \times 10^{-5}$	$0.0297 \pm 1.8876 \times 10^{-5}$
p_0	100	99.4130 ± 0.2574	99.4518 ± 0.3574	99.5991 ± 1.5108	99.6997 ± 1.0554
		102.1872 ± 0.1856	102.2306 ± 0.1441	100.8624 ± 1.3447	100.8624 ± 1.1805
		101.2097 ± 0.1431	101.2549 ± 0.2538	100.0403 ± 1.4338	100.0593 ± 1.2610
t_d	100	25.1318 ± 0.0559	25.1580 ± 0.0774	25.0496 ± 0.5481	25.0502 ± 0.4467
		25.2317 ± 0.4154	25.2428 ± 0.0287	25.9357 ± 1.0415	25.6215 ± 0.8149
		25.0730 ± 0.0282	25.0714 ± 0.0549	25.3187 ± 0.7251	25.4469 ± 0.7790

The results are listed in Table 4.3 and Table 4.2 (bottom). The marginal posterior distributions of each of the four parameters learnt by the GP-ABC-OLCM and the ABC-SMC-OLCM are shown in Figure 4.2. As in the previous example the histograms in Figure 4.2 are based on the final particle populations generated by the respective algorithms. In this example we see a huge speedup while using our proposed GP-ABC-

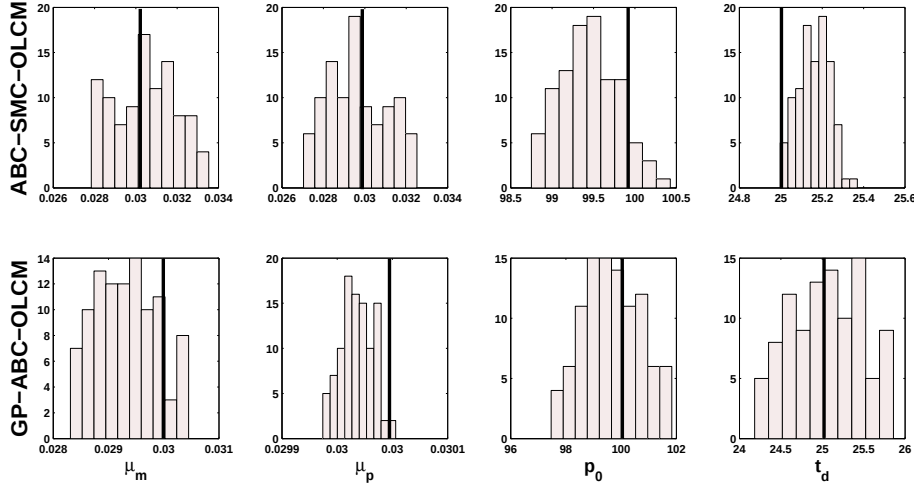


Figure 4.2: Histograms of the final particle populations generated by the GP-ABC-OLCM and the ABC-SMC-OLCM algorithms, approximating the marginal true posterior distribution for each of the parameters of Hes 1 respectively. Inference based on Dataset 1 is shown here. The black line marks the ground truth in each plot.

SMC and GP-ABC-OLCM algorithms, demonstrating the benefits of this approach. As in the previous example we noticed higher acceptance rates (fewer generated particles) for the GP variants of ABC-SMC. The noise is estimated as $\sigma_\mu = 6.8080, \sigma_p = 128.6910, \sigma_\mu = 6.9280, \sigma_p = 123.2920$ and $\sigma_\mu = 6.1220, \sigma_p = 128.1290$ for the dataset 1, 2 and 3 respectively. As in the case of Lotka Volterra model we see differences between the marginal densities for the Hes1 model parameters as shown in Figure 4.2. We believe the same geometric reasons explain these differences. Furthermore, from Table 4.2 it is evident that the acceptance rates increase slightly for the Lotka Volterra (Table 4.2, top) and significantly for the Hes1 (Table 4.2, bottom) model. Also, for the Hes1 model (see Table 4.3) better parameter estimates are produced by the GP based ABC-SMC variants since the estimated means (obtained from the final population) are closer to the true parameter values. Increase in acceptance rates results from an increase in sensitivity of the velocity field, which in turn reinforces our geometric intuitions about the output of ABC-SMC algorithms. For similar reasons an increase in the identifiability of the model (as evident from better estimates) can be attributed to the increase in model sensitivity when the velocity field is used to create the ABC distance function. We also like to point out the fact that for GP-ABC-SMC (with both the perturbation kernels) no knowledge of the initial history function is supplied. Thus, in a practical setting

we believe a GP based ABC-SMC algorithm is the optimal choice among these four methods for parameter estimation in DDEs.

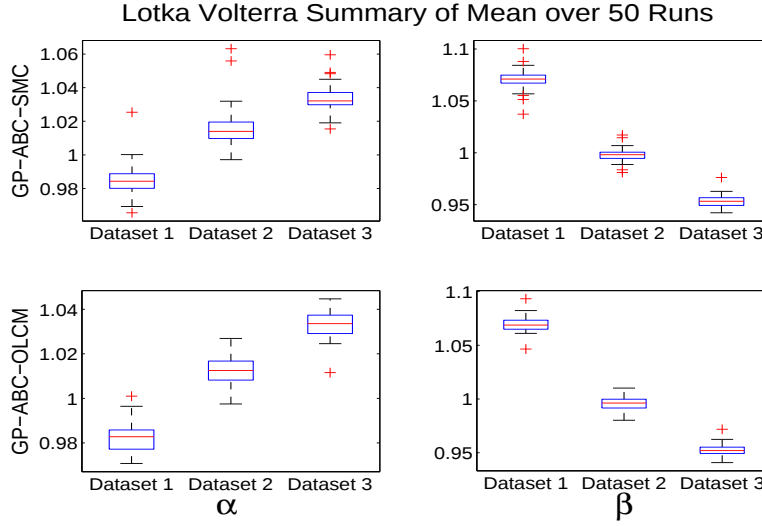
4.1.3 ABC variability

Our proposed method comprises of two levels of approximation, one induced through the GP regression and the other one resulting from the approximate inference scheme. Thus in order to check the robustness of our proposed algorithm we repeated the GP-ABC-SMC and GP-ABC-OLCM parameter inference steps for 50 runs on each of the three artificial datasets for both the Lotka Volterra and Hes1 models. We used the same algorithmic settings and prior distributions as in the previous examples. Figure 4.3 and 4.4 summarize the distributions of the sample mean and variance (corresponding to the final particle population for each run of GP-ABC-SMC and GP-ABC-OLCM on the three artificial datasets) across all the 50 runs on the data from Lotka Volterra and Hes1 respectively. It is evident from Figure 4.3 that the GP-ABC-OLCM algorithm produces fewer outliers compared to the GP-ABC-SMC for both the mean and variance estimates. This can be attributed to the local moves in the parameter space caused by the multivariate (OLCM) perturbation kernel. However, presence of these outliers indicate that the marginal densities for these runs are noticeably different from the rest of the runs.

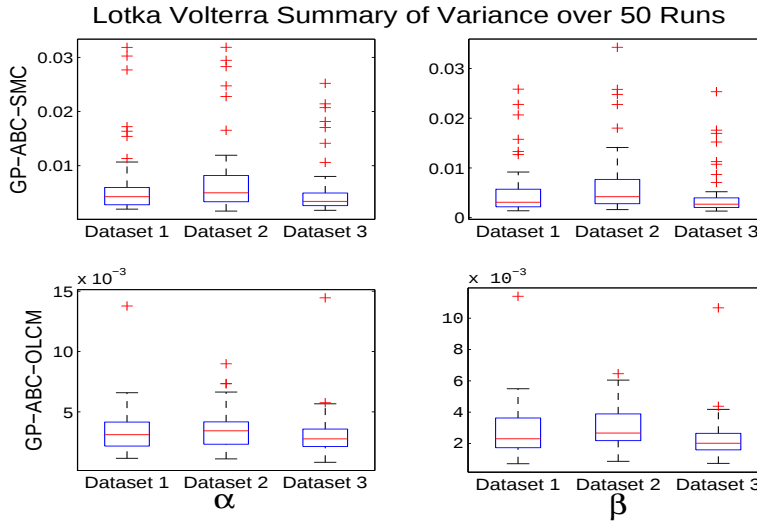
In case of the Hes1 model it is apparent from Figure 4.4 that the distributions are less variable across multiple runs and variants of the algorithms. Moreover, in this case we notice that the distribution of the variances have very few outliers indicating greater accordance among the posteriors learnt after each run of the algorithms.

The outliers present in both Figure 4.3 and 4.4 point to the fact that the final tolerances have a definitive role in the differences among the densities, thus the variances, obtained from many runs of GP-ABC-SMC. The final particle population in each run constitutes those parameter values that produce distances (in the parameter space) that satisfy the corresponding final tolerance values. Now for an adaptive schedule the final tolerances, corresponding to each run, are not constrained to be the same and thus the resulting particle populations are also not constrained to gather around the same region in the parameter spaces. For this reason some of the marginal densities among the 50 runs are noticeably different from others and generate outliers.

Figure 4.5 and 4.6 show the learnt state trajectories of the Lotka Volterra and Hes1 model compared against the true state trajectories for each of the datasets. The true trajectories correspond to the true parameters and the reconstructed trajectories are generated by solving the Lotka Volterra (equation 4.1) and Hes1 (equation 4.2) model

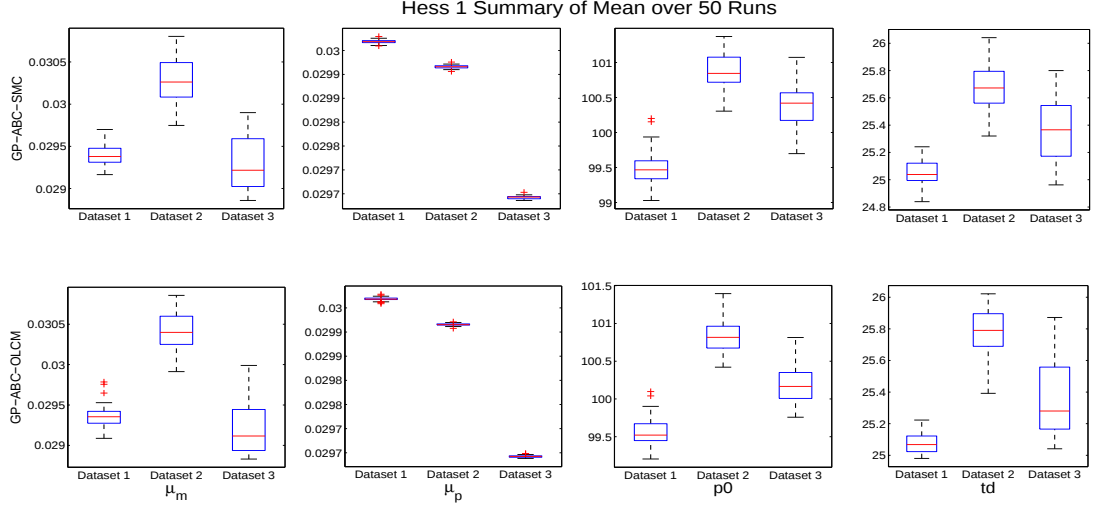


(a) Distribution of the mean for Lotka Volterra across 50 runs of GP-ABC-SMC and GP-ABC-OLCM.

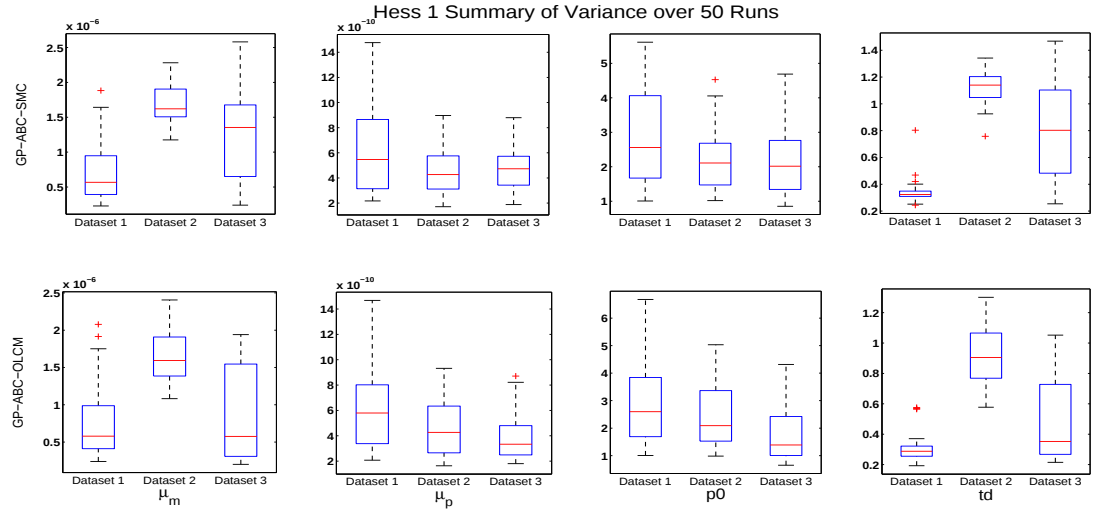


(b) Distribution of the variance for Lotka Volterra across 50 runs of GP-ABC-SMC and GP-ABC-OLCM.

Figure 4.3: The boxplots represent the distributions of the mean and variances (across 50 runs) of the final population representing the marginal approximate posterior parameter distributions learnt by the GP-ABC-SMC and the GP-ABC-OLCM from the three artificial datasets of Lotka Volterra model.



(a) Distribution of the mean for Hes1 across 50 runs of GP-ABC-SMC and GP-ABC-OLCM.



(b) Distribution of the variance for Hes1 across 50 runs of GP-ABC-SMC and GP-ABC-OLCM.

Figure 4.4: Distributions of the mean and variances learnt by the GP-ABC-SMC and the GP-ABC-OLCM from the three artificial datasets of Hes1 model.

equations. While solving (numerically integrating) these differential equations the parameters are taken as the median of the parameters learnt by the GP-ABC-SMC algorithm considering all the 50 runs. The median value is considered here to reflect the effect of variability (in parameter learning by the GP-ABC-SMC) in reconstructing the dynamics of the considered models.

4.1.3.1 ABC variability: Effect of population size

In order to investigate the effect of particle population size we ran the GP-ABC-SMC algorithm using the component-wise perturbation kernel on dataset 1 pertaining to the Lotka Volterra model using different values of N . Furthermore, we used a single fixed

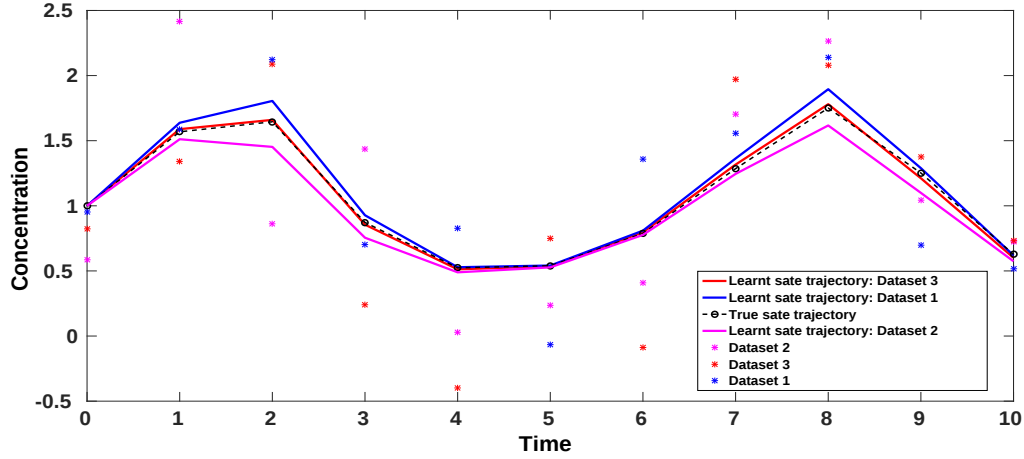
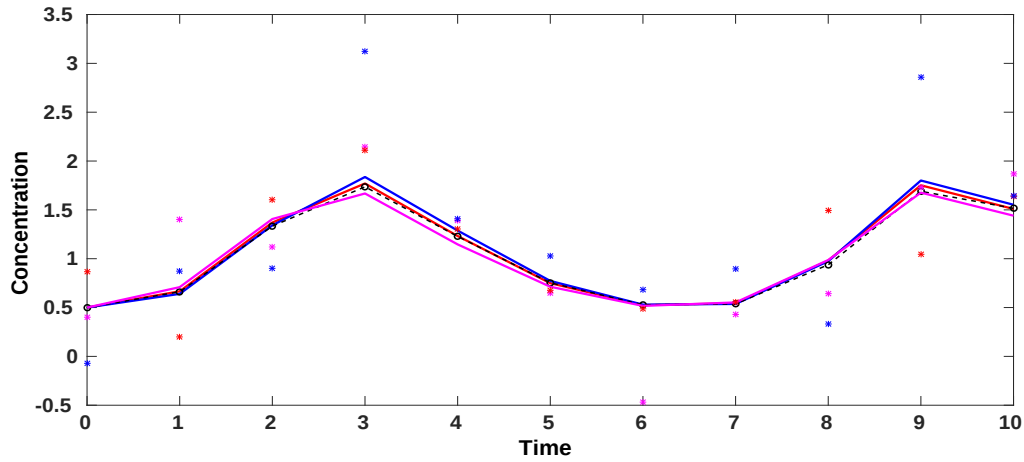
(a) Trajectories of $x(t)$.(b) Trajectories of $y(t)$.

Figure 4.5: Reconstructed and true state trajectories of the Lotka Volterra model. The results corresponding to the three datasets (1, 2, 3) are shown using blue, red and magenta colours respectively. Reconstructed trajectories are represented as curves and observations as stars. The ground truth is the black (dashed and circled) curve.

tolerance schedule in each case to stop the introduction of variability due to differences in final tolerance values. Figure 4.7 shows the different posterior marginal densities corresponding to 5 different population sizes $N = \{100, 500, 1000, 5000, 10000\}$. Differences among the posterior densities corresponding to the population sizes are barely noticeable in Figure 4.7. Thus this analysis confirms the fact that variability can only appear when an adaptive tolerance schedule is used within GP-ABC-SMC.

4.1.4 Signal transduction cascade

We have, so far, used the benchmarking examples to compare our proposed GP-based ABC-SMC approach to others of that ilk that exist in the literature. In this example we will compare the parameter estimation results for the proposed GP based ABC-SMC

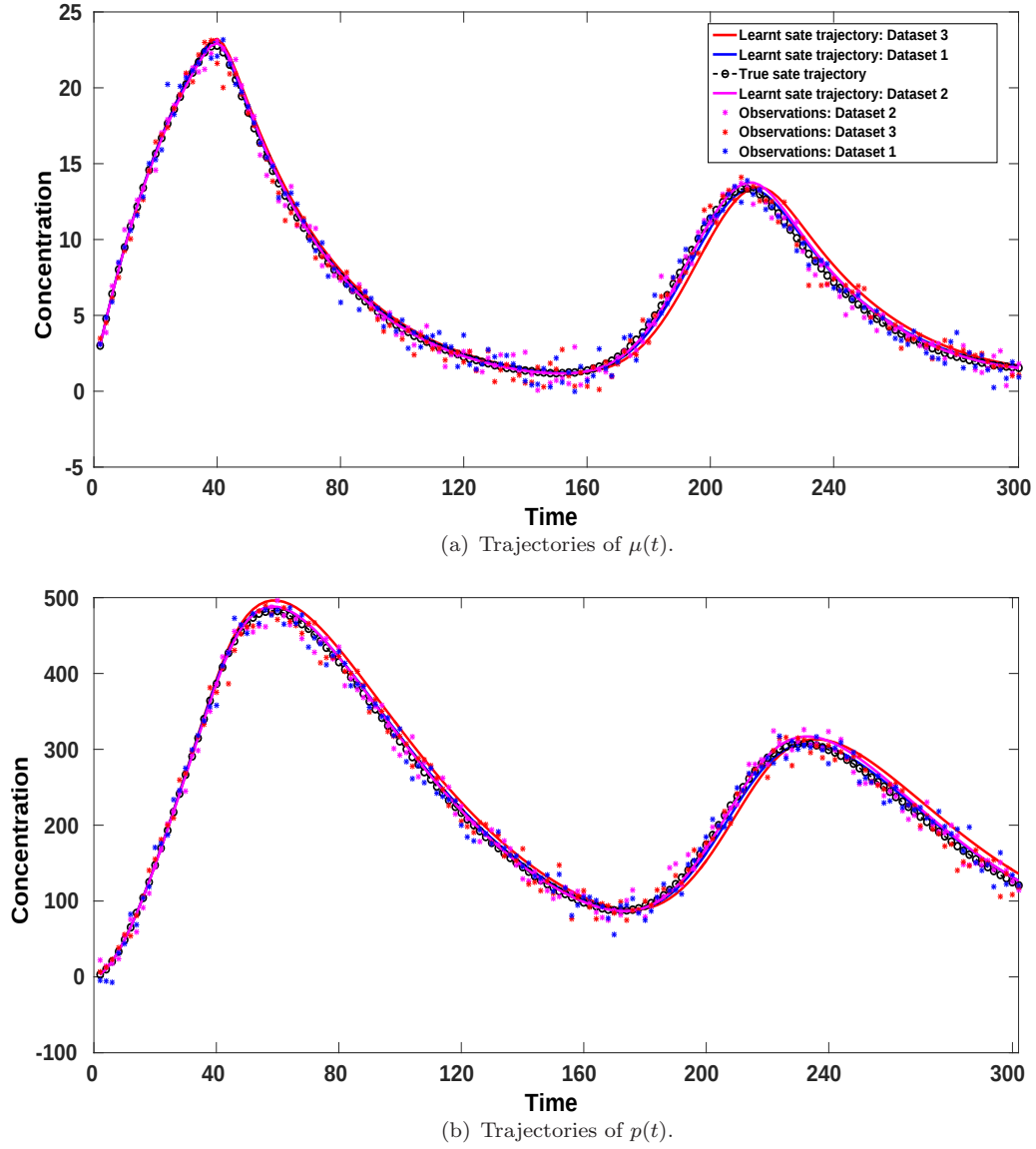


Figure 4.6: Reconstructed and true state trajectories of the Hes1 model.

with other (methods not falling under ABC) recent GP based approximate inference methods for parameter estimation in ODEs. For this purpose we have chosen the signal transduction cascade model (Vyshemirsky and Girolami, 2008). Using this model, a comparison between the competing GP based approaches were reported in Wang and Barber (2014). Thus evaluating the proposed GP based ABC-SMC algorithm on this model (with identical settings to those in (Wang and Barber, 2014)) will enable us to draw comparisons with these other methods. This model is described by a 5-dimensional

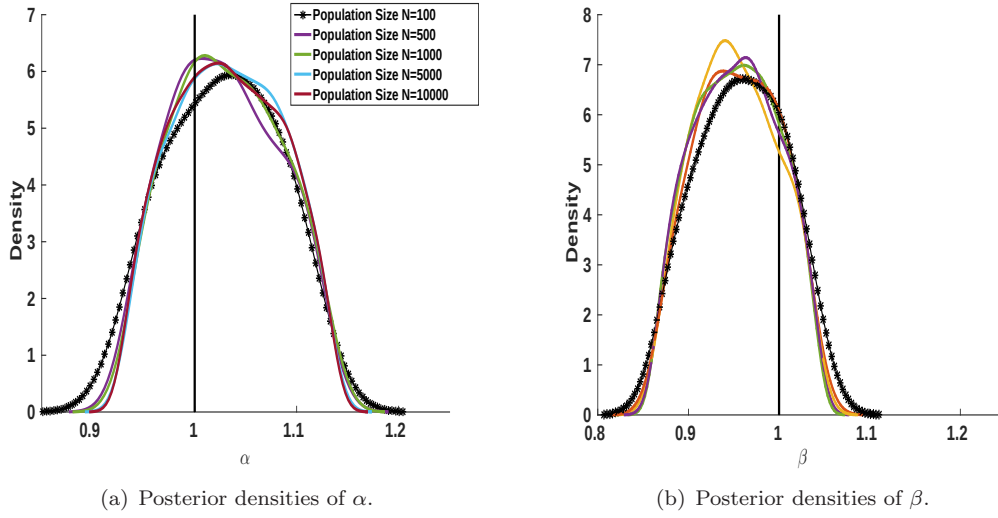


Figure 4.7: Posterior marginal densities for the Lotka Volterra model parameters for different population sizes used in GP-ABC-SMC. We have used kernel density estimation for plotting these densities. The black vertical lines denote the ground truth and the black thick curve in both plots shows the respective posterior densities for $N = 100$.

coupled ODEs given by

$$\begin{aligned}
 \frac{d[S]}{dt} &= -k_1[S] - k_2[S][R] + k_3[RS] \\
 \frac{d[S_d]}{dt} &= k_1[S] \\
 \frac{d[R]}{dt} &= -k_2[S][R] + k_3[RS] + \frac{V[R_{pp}]}{K_m + [R_{pp}]} \\
 \frac{d[RS]}{dt} &= k_2[S][R] - k_3[RS] - k_4[RS] \\
 \frac{d[R_{pp}]}{dt} &= k_4[RS] - \frac{V[R_{pp}]}{K_m + [R_{pp}]}
 \end{aligned} \tag{4.3}$$

where $\theta = (k_1, k_2, k_3, k_4, V, k_m)$ are the parameters of this model and $\mathbf{X}(t) = ([S], [S_d], [R], [RS], [R_{pp}])$ are the concentrations of the state variables. Following Wang and Barber (2014) we generated data from the model between the time interval $(0 \leq t \leq 100)$ with parameters $\theta = (0.07, 0.6, 0.05, 0.3, 0.017, 0.3)$ and initial values of the state variable $[S] = 1$, $[S_d] = 0$, $[R] = 1$, $[RS] = 0$, $[R_{pp}] = 0$. We then sampled the data at time $t^L = \{0, 1, 2, 4, 5, 7, 10, 15, 20, 30, 40, 50, 60, 80, 100\}$ and added random noise with standard deviation $\sigma_{[S]}$, $\sigma_{[S_d]}$, $\sigma_{[R]}$, $\sigma_{[RS]}$, $\sigma_{[R_{pp}]}$ set to 0.1 for generating the synthetic data. For inferring parameters in this example we apply the GP-ABC-OLCM algorithm from our study with multiple runs, where we found this algorithm to provide a stable and fast inference mechanism. The non-stationarity in the time evolution of the state

Table 4.4: Estimated parameters of the signal transduction cascade by all the GP based approaches including the GP-ABC-OLCM. The estimates for **GP-ODE**, **AGM** and **GM** are taken from Wang and Barber (2014).

Parameters	True value	GP-ABC-OLCM	GP-ODE	AGM	GM
k_1	0.070	0.0708 ± 0.0086	0.0747 ± 0.0130	0.0771 ± 0.0130	0.0762 ± 0.0130
k_2	0.6	0.5806 ± 0.0706	0.6230 ± 0.1246	0.5460 ± 0.1259	0.5632 ± 0.1256
k_3	0.05	0.0480 ± 0.0074	0.0530 ± 0.0135	0.0593 ± 0.0111	0.0594 ± 0.0115
k_4	0.3	0.3439 ± 0.0659	0.2960 ± 0.0281	0.3750 ± 0.0999	0.3754 ± 0.1051
V	0.017	0.0170 ± 0.0009	0.0177 ± 0.0014	0.0172 ± 0.0015	0.0173 ± 0.0014
k_m	0.3	0.3110 ± 0.0774	0.4220 ± 0.0690	0.4090 ± 0.0911	0.4186 ± 0.0953

variables is captured by the MLP covariance function given by

$$k(t, t') = \sigma_{kern}^2 \times \frac{2}{\pi} \text{asin} \left(\frac{\sigma_w^2 t^\top t' + \sigma_b^2}{\sqrt{\sigma_w^2 t^\top t + \sigma_b^2 + 1} \sqrt{\sigma_w^2 t'^\top t' + \sigma_b^2 + 1}} \right), \quad (4.4)$$

where the kernel variance σ_{kern}^2 , the neural network weight variance σ_w^2 , and the bias variance σ_b^2 are the hyperparameters of the covariance function. The derivative of this kernel with respect to the input time t is given by

$$\frac{\partial k(t, t')}{\partial t} = \frac{\sigma_{kern}^2}{\sqrt{1 - Z^2}} \frac{\partial Z}{\partial t}, \quad (4.5)$$

where

$$Z = \frac{\sigma_w^2 t^\top t' + \sigma_b^2}{Z_{norm}} \quad (4.6)$$

with $Z_{norm} = \sqrt{\sigma_w^2 t^\top t + \sigma_b^2 + 1} \sqrt{\sigma_w^2 t'^\top t' + \sigma_b^2 + 1}$. All the other algorithmic settings were kept the same. The prior distributions are chosen as $k_1 \sim \mathcal{U}(0.05, 0.09)$, $k_2 \sim \mathcal{U}(0.4, 0.8)$, $k_3 \sim \mathcal{U}(0.03, 0.07)$, $k_4 \sim \mathcal{U}(0.1, 0.5)$, $V \sim \mathcal{U}(0.015, 0.0195)$ and $k_m \sim \mathcal{U}(0.1, 0.5)$. In this example S_{MC} is set to 3.

The resulting parameter estimates are furnished in Table 5.1 along with the parameter estimates obtained from other GP based algorithms run on the same model. These algorithms are the **GP-ODE** method proposed in Wang and Barber (2014), the adaptive gradient matching (**AGM**) proposed in Dondelinger et al. (2013) and the gradient matching (**GM**) proposed in Calderhead et al. (2008). In Table 5.1 we have summarised the GP-ABC-OLCM output using mean and standard deviation where the choice of the latter has been made to facilitate comparison. We have compared the true state trajectories with the reconstructed trajectories in Figure 4.8. We generated the reconstructed trajectories by solving (equation 4.3) using the mean of the final population of GP-ABC-OLCM, representing the marginal posterior densities of the parameters. The estimated values of the standard deviations are $\sigma_{[S]} = 0.0964$, $\sigma_{[S_d]} = 0.0818$, $\sigma_{[R]} = 0.0707$, $\sigma_{[RS]} = 0.0591$ and $\sigma_{[R_{pp}]} = 0.0754$.

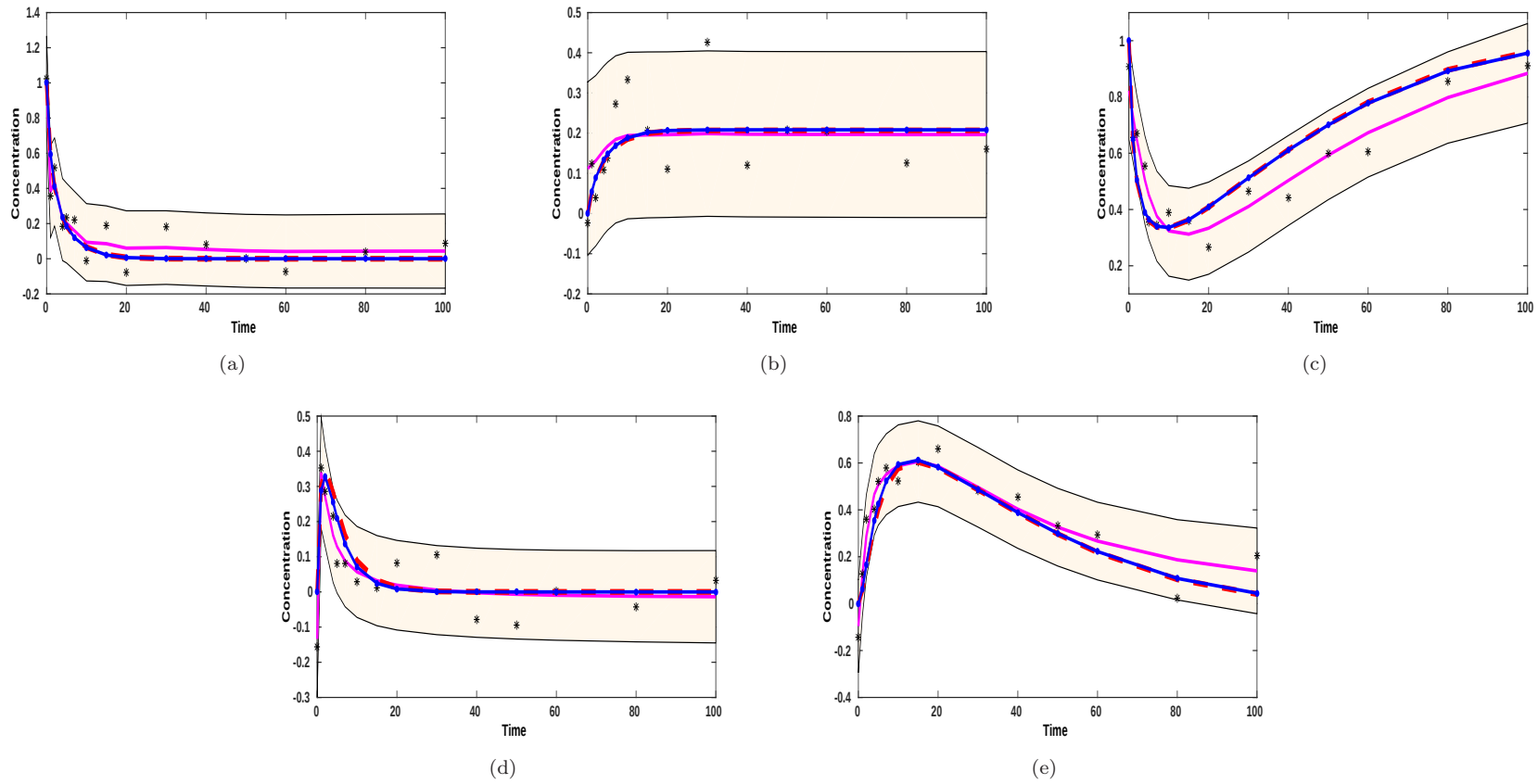


Figure 4.8: Results of GP-ABC-OLCM for the signal transduction cascade ($[S]$ – $[R_{pp}]$ in plots (a)–(e) respectively). In all the plots observations are the black stars, the true state trajectory is the red (dashed) curve and the blue curve shows the reconstructed trajectory. We have also plotted the GP mean function as the magenta curve and the 95% confidence region is shown as the shaded area. The reconstructed trajectory is generated by numerically integrating (equation 4.3) with the parameters set to the mean of the posterior distribution estimated by the GP-ABC-OLCM algorithm.

We avoided the comparison of run-time or acceptance rates as the GP-ABC-OLCM and other GP based algorithms depend on completely different approximate inference scheme. However, GP-ABC-OLCM is significantly faster than the other approaches. The GP-ABC-OLCM finishes the estimation in around 20 seconds while the other methods were run for 30 minutes to obtain a properly mixed Markov chain. The ratio of the last two parameters V/k_m (Dondelinger et al., 2013) is a crucial quantity that determines the reconstruction accuracy. GP-ABC-OLCM is able to infer this quantity with the best (based on the estimated posterior means of V and k_m) accuracy among all the GP based algorithms.

It is interesting to note that the variance in estimates for the parameters are less than other GP based approaches which use Markov chain Monte Carlo (MCMC). Since we use ABC for inference our estimated variance should be bigger compared to MCMC based inference results. One reason for this could be particle degeneracy which drives all but a few weights to near zero values. Particle degeneracy could be monitored through quantifying the effective sample size (ESS) (Del Moral et al., 2012) given at SMC step τ by

$$ESS(w_\tau^{(i)}) = \left(\sum_{i=1}^N (w_\tau^{(i)})^2 \right)^{-1}, \quad (4.7)$$

where $w_\tau^{(i)}$ is the weight of particle i and N is the population size. The value of ESS lies between 1 and N and its interpretation is that inference based on N weighted samples is approximately equivalent inference based on $ESS(w_\tau^{(i)})$ perfect samples from the intermediate distribution at SMC step τ . Although ESS is not a perfect measure, it does provide a probe into the behaviour of the algorithm. Generally when ESS falls below a threshold N' , generally greater than $N/2$ the particles are resampled. In the ABC-SMC the importance sampling is performed with the weights and thus resampling is not required unlike SMC samplers (Del Moral et al., 2006). However, for corner cases such as this example where the priors have very narrow support, monitoring degeneracy through ESS is useful. Thus we calculated the ESS values for each of the SMC steps (plotted in Figure 4.9, blue curve) and from these values, indicating the number of healthy particles, it is clear that the posterior estimates have less than half of the total particles representing a believable sample approximating the true posterior. This specific example highlights a crucial fallacy that the ABC-SMC, for unnatural priors, can potentially fall prey to particle degeneracy. To be absolute sure about the choice of priors leading to this problem we ran the ABC-SMC keeping the same GP estimates of the states and the velocities with increasing the support of the priors. The new priors are thus chosen as $k_1 \sim \mathcal{U}(0, 2)$, $k_2 \sim \mathcal{U}(0, 2)$, $k_3 \sim \mathcal{U}(0, 2)$, $k_4 \sim \mathcal{U}(0, 2)$, $V \sim \mathcal{U}(0, 0.1)$ and $k_m \sim \mathcal{U}(0.1, 2)$. We have kept the prior for V comparatively narrow to ensure reconstruction accuracy. The S_{MC} is set to 4. The resulting posterior distributions are shown as histogram plots of the final population of particles in Figure 4.10. We have also monitored ESS (shown in Figure 4.9 brown curve) which with these choices of prior distributions do

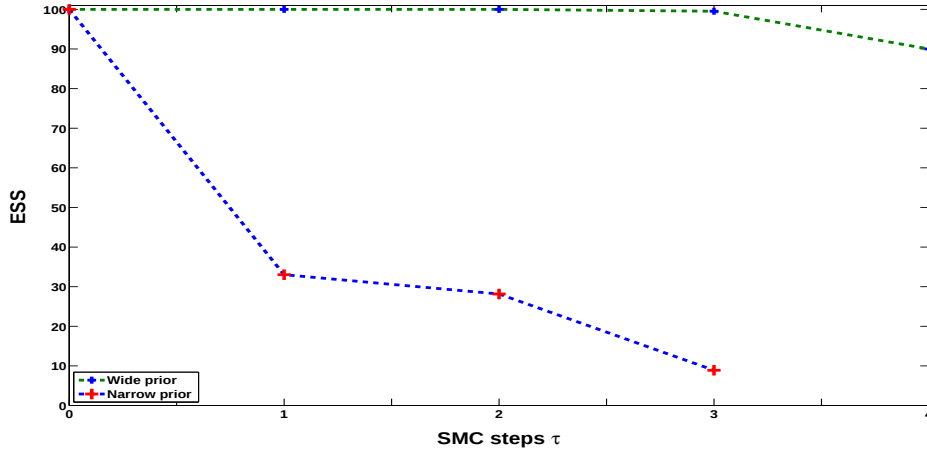


Figure 4.9: The ESS values indicating number of healthy or alive particles at each of the SMC steps. Except the first SMC step the number of alive particles fall well below 50 particles indicating particle degeneracies when using narrow priors (blue dashed curve) centred on the true values of the parameters. While using a wider prior we see a marked improvement in the ESS above 90%(brown dashed curve).

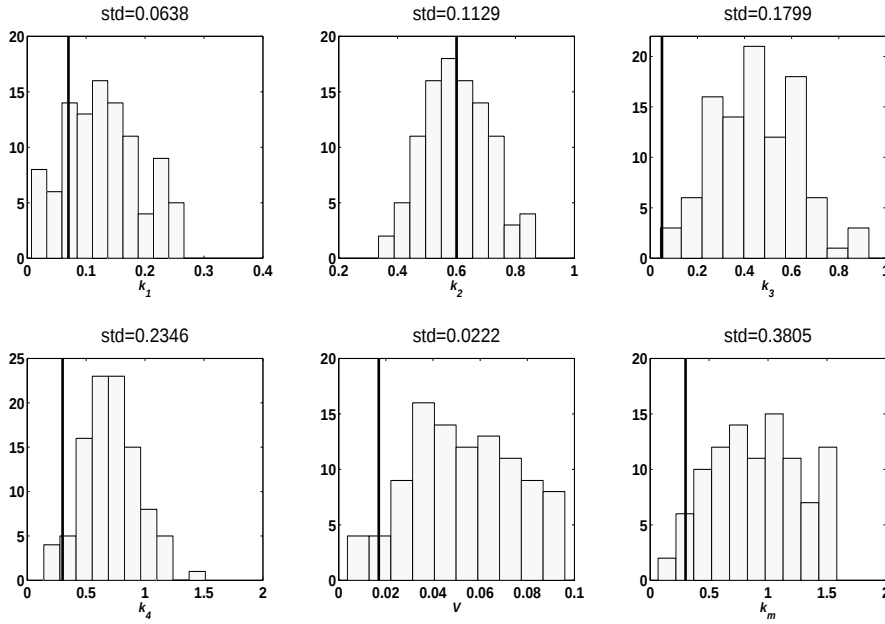


Figure 4.10: Posterior distributions shown as histogram plots. The standard deviations *std* for each parameters' final particle populations are shown on top of each of the plot. Clearly the posterior variances are bigger than those using MCMC in Table 4.

not show particle degeneracy. The reconstruction accuracies of the state time courses is shown in Figure 4.11. Clearly the reconstruction suffers in comparison to Wang and Barber (2014) and Dondelinger et al. (2013). Compared to Wang and Barber (2014)

we use a much less informative priors. Also note that our proposed algorithm lacks a feedback mechanism (Dondelinger et al., 2013) from the ODE system while carrying out GP regression and thus since some of the states are not well estimated by the GP model, the resulting posterior estimates suffer when uninformative priors are chosen. In Dondelinger et al. (2013) the posterior GP estimates are noticeably better than ours because of the feedback mechanism resulting in better estimates than what we achieve in Figure 4.11. However, when we use narrower prior we achieve similar or better estimates than Wang and Barber (2014); Dondelinger et al. (2013) (see Figure 4.8), albeit at the cost of population degeneracy. Thus a critical comparison of GP-ABC-SMC and **GP-ODE** (which use the aforementioned feedback mechanism) is required for these choices of priors to benchmark our two-step against adaptive methods such as **GP-ODE**. However, our primary goal is to resolve the computational issues of ABC-SMC. Thus, we believe this critical comparison is not essential for the development of ABC-SMC.

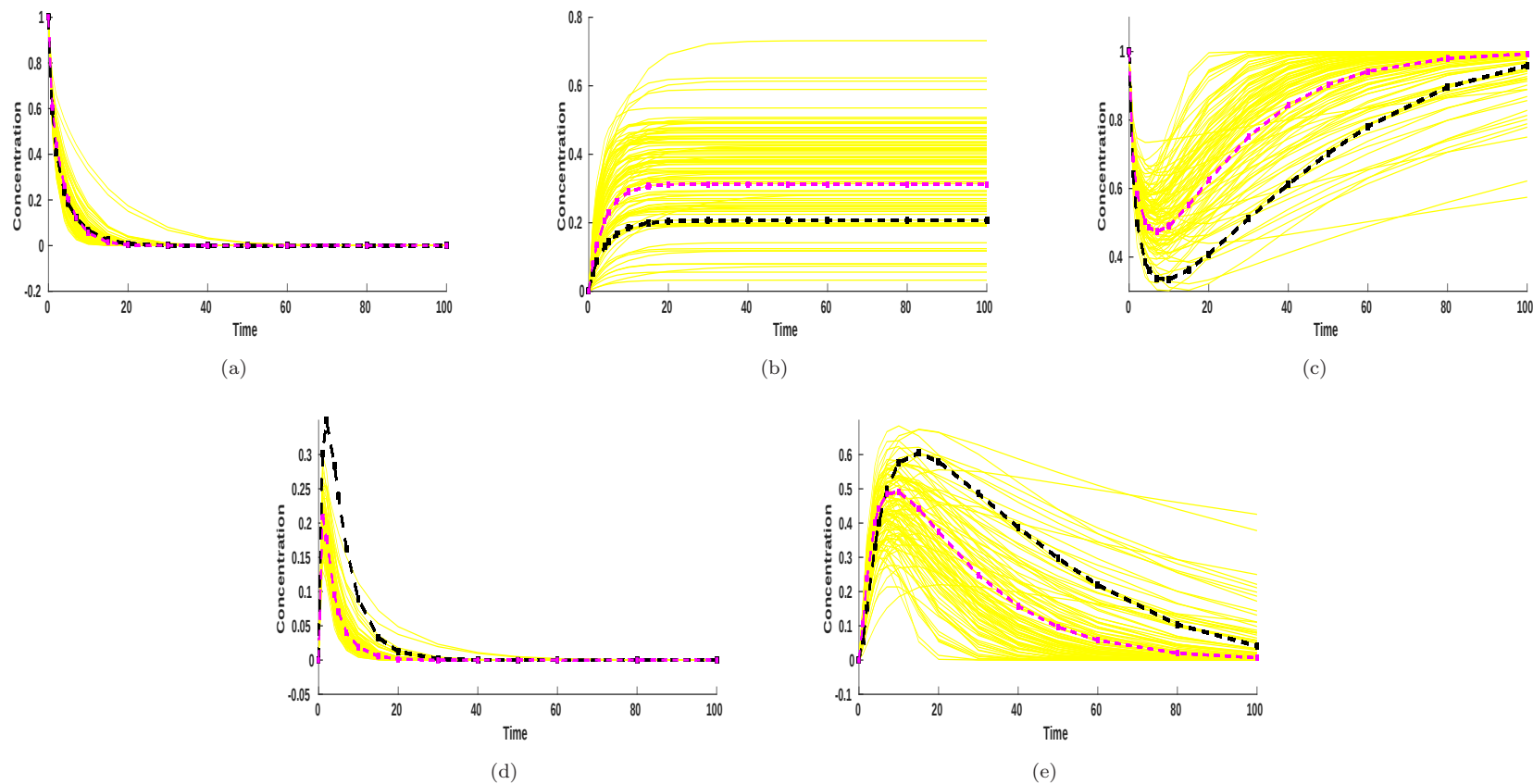


Figure 4.11: Results of GP-ABC-OLCM for the signal transduction cascade using a wide support for the priors. In all the plots the true state trajectory is the black (dashed) curve and the magenta curve shows the reconstructed trajectory generated by numerically integrating (equation 4.3) with the parameters set to the mean of the posterior distribution estimated by the GP-ABC-OLCM algorithm. The yellow thin curves are reconstructions based on each of the $N = 100$ posterior parameter sample from the last step of GP-ABC-OLCM.

4.2 Model selection with GP-ABC-SMC

To test GP-ABC-SMC for model selection we have chosen an example previously presented in [Toni et al. \(2009\)](#). In this example the ABC-SMC model selection algorithm was used to find the most suitable among four epidemic models that describe a 21 day common-cold data from October 1967. This data-set was provided in [Toni et al. \(2009\)](#). The common-cold broke out on an isolated island called *Tristan da Cunha* in the Atlantic ocean with a population of 300 individuals. In this example we will carry out a model selection exercise on three of those epidemic models. These three are ODEs and the one that we have left out is a DDE. All these models are variations of the SIR model ([Anderson et al., 1991](#)). SIR models describe the spread of such disease in a population of susceptible $S(t)$, infected $I(t)$ and recovered $R(t)$ individuals. The simplest model, Model 1, is the basic SIR model given by

$$\begin{aligned}\dot{S} &= \alpha - \gamma SI - dS \\ \dot{I} &= \gamma SI - vI - dI \\ \dot{R} &= vI - dR.\end{aligned}\tag{4.8}$$

where α is the birth rate, d — the death rate, γ —the infection rate and v —the recovery rate. The second model, Model 2, introduces another state $L(t)$ to model individuals with a latent phase of infection. In this phase one gets infected but yet cannot infect others. This model is given by

$$\begin{aligned}\dot{S} &= \alpha - \gamma SI - dS \\ \dot{L} &= -\gamma SI - \delta L - dL \\ \dot{I} &= \delta L - vI - dI \\ \dot{R} &= vI - dR.\end{aligned}\tag{4.9}$$

Here δ is the transition rate from the latent to the infective stage. The final model, Model 3, allows the recovered individuals to become susceptible again.

$$\begin{aligned}\dot{S} &= \alpha - \gamma SI - dS + eR \\ \dot{I} &= \gamma SI - vI - dI \\ \dot{R} &= vI - (d + e)R,\end{aligned}\tag{4.10}$$

where the new parameter e denotes the rate of becoming susceptible again.

This common-cold dataset is presented in Figure 4.12 as time courses of the numbers of infected $I(t)$ and recovered $R(t)$ individuals. Thus while using GP-ABC-SMC we have to consider $S(t)$ as a hidden variable. This hidden variable will be handled within GP-ABC-SMC using explicit integration (using the Euler method) as mentioned in section 3.6.4 in Chapter 3. Thus we need to consider the initial condition $S(t = 0)$ as an

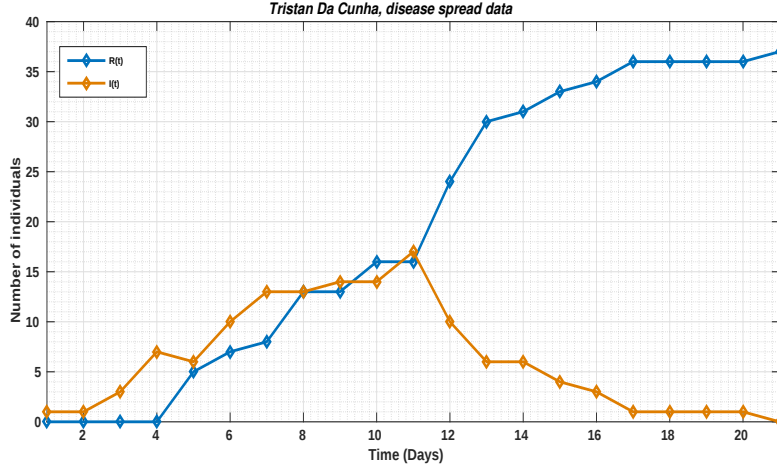


Figure 4.12: Common-cold data from *Tristan da Cunha* collected in October 1967 (Toni et al., 2009).

Table 4.5: Total number of times m indicates the corresponding models in the final population generated in each run of GP-ABC-SMC. Model 1, 2, and 3 correspond to the ODEs in equation 4.8, 4.9 and 4.10.

Run	Model 1	Model 2	Model 3
1	18	79	3
2	42	45	13
3	50	50	0

additional parameter, which we denote as S_o . We define the model indicator $m \in \{1, 2, 3\}$ representing the above models in the same order. Since the data is for a 21 day period we have set, following Toni et al. (2009), α and d to 0 assuming no new birth or death in that period. The model specific parameter vectors are defined as $\theta(m) : \theta(1) = (\gamma, v, S_o); \theta(2) = (\gamma, v, \delta, S_o); \theta(3) = (\gamma, v, e, S_o)$. We ran the GP-ABC-SMC algorithm for model selection on this data-set three times. The model selection part of the GP-ABC-SMC followed Algorithm 2 (see Chapter 2). We chose the same prior distributions as found in Toni et al. (2009): $\gamma \sim \mathcal{U}(0, 3)$, $v \sim \mathcal{U}(0, 3)$, $\delta \sim \mathcal{U}(-0.5, 5)$, $e \sim \mathcal{U}(-0.5, 5)$ and $S_o \sim \mathcal{U}(37, 100)$. A univariate perturbation kernel is chosen and the squared exponential kernel is used for the GP regression. The choices of priors and the perturbation kernel are motivated by the fact that we want to set up the inference task in a similar way to Toni et al. (2009) for a fair comparison. An adaptive tolerance schedule is used where S_{MC} is set to 6. Table 4.5 furnishes the posterior of the model indicator corresponding to the final populations generated in each run of GP-ABC-SMC. Each number in Table 4.5 thus represents the number of times m indicates a specific model in the final population. To choose the most suitable among any two models the ratio of the indicator frequencies has to be evaluated. This ratio gives us the Bayes factor. The third model (Model 3) is found to be the least suitable model in all the three runs as the Bayes factor (the ratio), $BF_{1/3} > 3$ and $BF_{2/3} > 3$. For the first run Model 2 emerges

as the winning model, $BF_{1/2} > 3$. For the subsequent runs both Model 1 and 2 describe the data equally well. In [Toni et al. \(2009\)](#) similar observation was made in terms of models ranking, however Bayes factors were not reported. We have plotted the model indicator allocation for all the SMC steps in the third case where a tie occurs between the first two models. These plots are presented in Figure 4.13. Population number 5 in Figure 4.13 is perhaps an indication of Model 2 getting stuck near a local minima which it manages to escape in the next population.

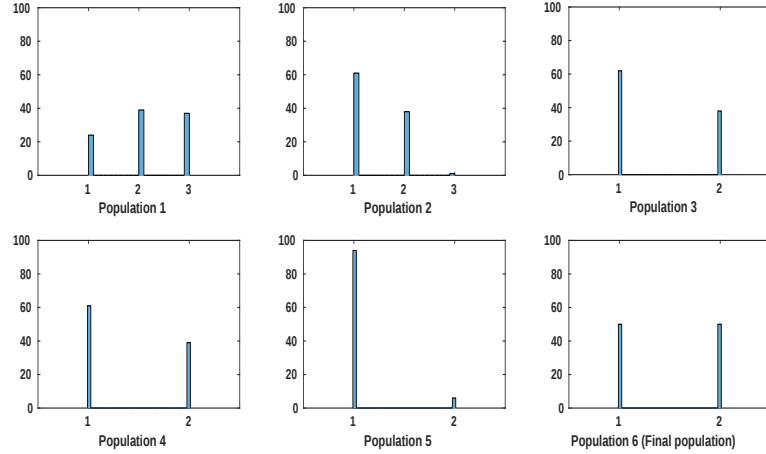


Figure 4.13: Histograms show populations 1 – 6 of the model particles m for the third run of GP-ABC-SMC. Population 6 represents the final posterior marginal estimate of m . The third model get eliminated from population 3 onwards.

We have presented the posterior distributions of each of the model parameters for Model 1 & 2 in Figure 4.14. The posterior distributions for Model 3 are centred on similar values found in [Toni et al. \(2009\)](#). This indicates that the GP-ABC-SMC can recover the desired posteriors while operating in the model selection mode. However, note that the initial value estimates from both the models are quite low considering the total number of inhabitants of the island. This indicates that none of these models are fully capable of explaining the disease spread accurately. We have compared the observations with reconstructed data in Figure 4.15. Reconstruction is done using the mean of the particles (for each model parameter) from the final populations of Model 1 & 2. As indicated through the model selection process it is difficult to distinguish between both these models in terms of reconstruction accuracies.

4.3 Chaotic attractor: The Mackey-Glass equation

So far we have concentrated our efforts on model systems with oscillatory time courses as well as non-stationary ones (see time course of $[RS]$ in the signal transduction example). Thus in this final example we will showcase the GP-ABC-SMC applied to a chaotic

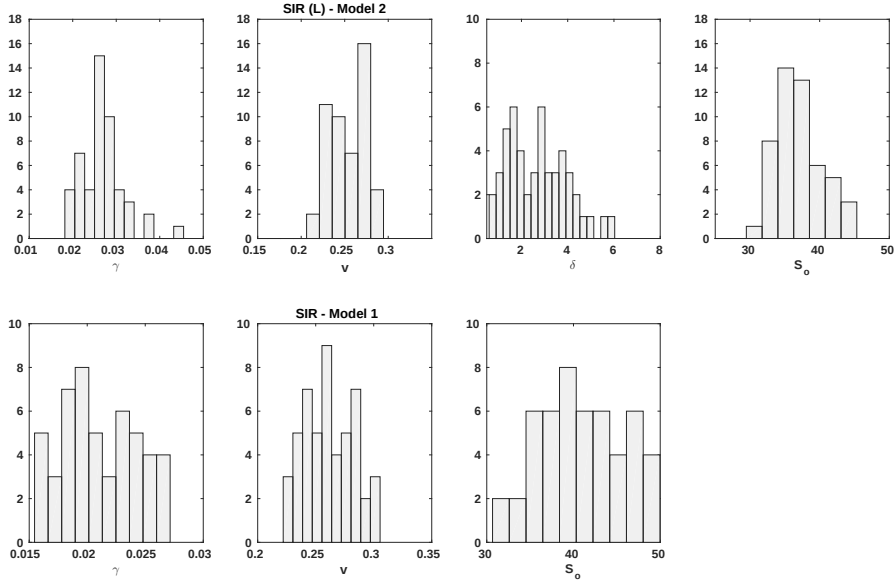


Figure 4.14: Histograms show the posterior marginal densities of the learned parameters. Posteriors in the top & bottom rows correspond to Model 2 and Model 1 parameters respectively.

time series. This problem presents a significant challenge to any inference algorithm as parameters need to be estimated from observations of a chaotic state trajectory corrupted with noise.

In order to test our algorithm we generated 200 samples between the interval $[0 : 0.1 : 20]$, from the chaotic regime of the Mackey-Glass model system (equation 3.33) with the parameter values, $\gamma = 1$, $\beta = 2$, $t_d = 2$ and $n = 9.65$, that can generate chaos. Please see the illustration of the velocity estimation process using this system in section 3.6.1 of Chapter 3. We further added random noise with standard deviation $\sigma_x = 0.1$ to generate synthetic data. As given in section 3.6.1 of Chapter 3, we have chosen the Matérn covariance kernel for the GP. Furthermore, we have chosen OLCM perturbation kernel and set S_{MC} to 5. The priors are chosen as $\beta \sim \mathcal{U}(0, 5)$, $n \sim \mathcal{U}(2, 12)$, $\gamma \sim \mathcal{U}(0.5, 2)$, $t_d \sim \mathcal{U}(0, 5)$. The resulting parameter estimates obtained by the algorithm are listed in Table 4.6. The posterior distributions are plotted in Figure 4.16. The reconstructed

	β	n	γ	t_d
ground truth	2	9.65	1	2
estimated	1.7225	9.6713	0.8504	2.0092

Table 4.6: Estimated parameters of the Mackey-Glass equations.

time course of x is shown in Figure 4.17 along with the actual and corrupted time courses. Reconstruction is carried out using the mean of the final population as in some of the previous examples.

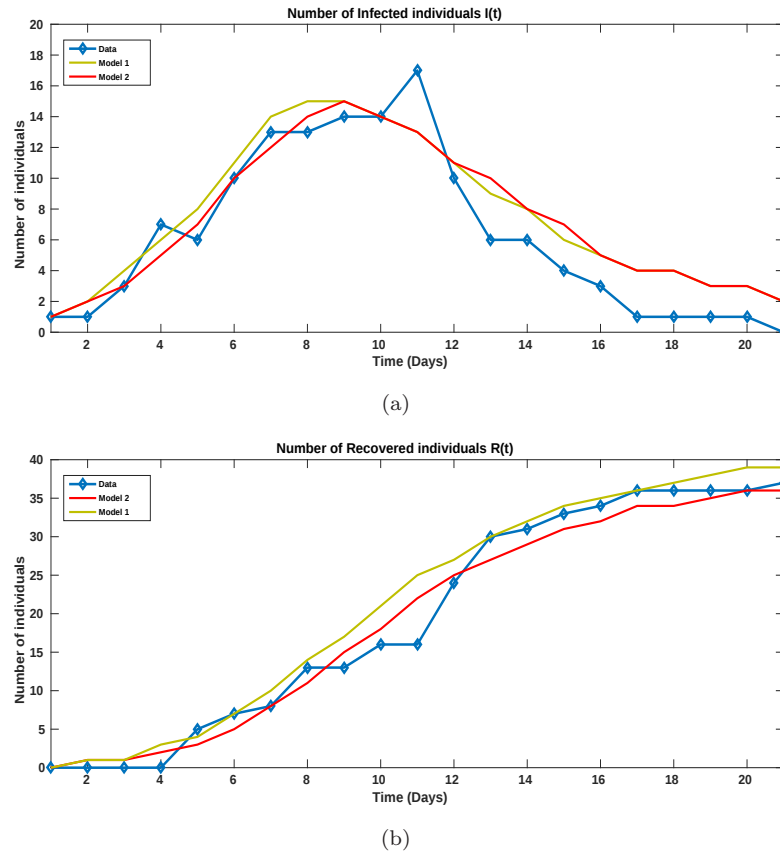


Figure 4.15: Comparison of the actual common-cold data with the reconstructed time courses from Model 1 & 2. The reconstruction is done using mean of the final particle populations generated in the third run of GP-ABC-SMC. In terms of model selection none of these models stand out from the other. Plot (a) shows the time courses of $I(t)$ and (b) shows $R(t)$.

Although it may be evident from Figure 4.17 that a fair degree of accuracy is maintained in the reconstruction, the real challenge presented in this problem is highlighted when we plot a comparison between the original chaotic attractor and the reconstructed one. This comparison is presented in Figure 4.18. Clearly the reconstructed attractor differs significantly from the original. This failure points out that evaluating distances between time series is not optimal for chaotic systems. Thus for these systems distances within ABC should incorporate phase space features commonly found in non-linear dynamics literature. Furthermore, it will be interesting to compare the two step and adaptive methods on a chaotic differential equation model such as this one. We have found that GP regression (using different covariances) achieves a fairly accurate smoothing performance on the trajectory of x , indicating the fact that smoothing is not the problem. Thus, this deceptively simple model can be used as a good benchmark for testing the influence of coupling between the GP and ODE parameters (in an adaptive method) on the attractor reconstruction. Having said this we like to point out that almost no evidence (in literature) can be found about the application of Bayesian inference

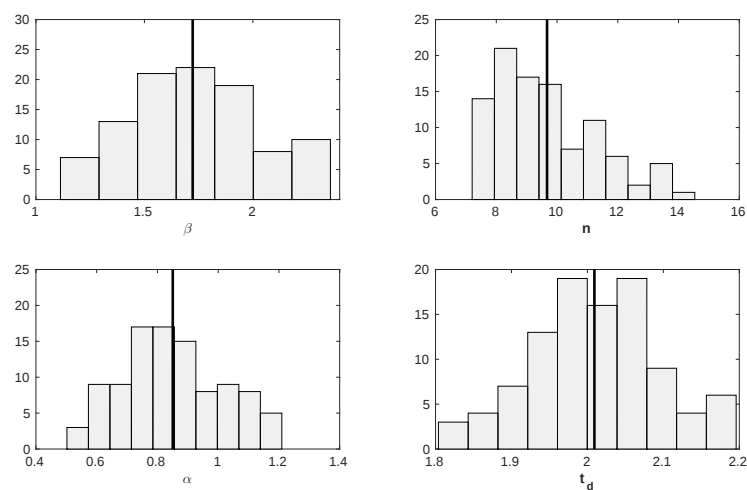


Figure 4.16: Posterior distributions of the parameters of the Mackey-Glass model as histogram of final populations (for each parameter respectively) of GP-ABC-OLCM.

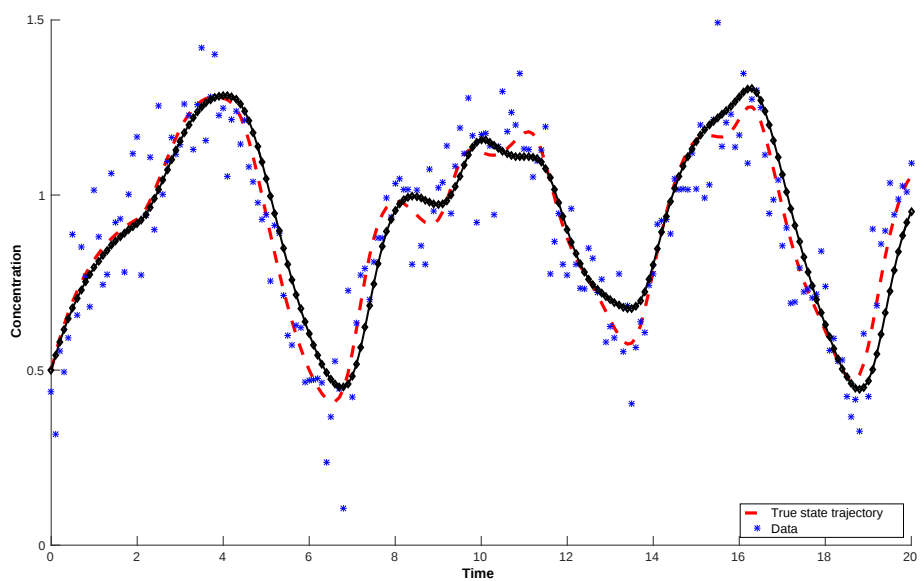


Figure 4.17: Reconstructed time course of the variable x along with the ground truth trajectory and its noisy version.

algorithms on chaotic ODE or DDE systems. This is an open problem that requires novel investigations.

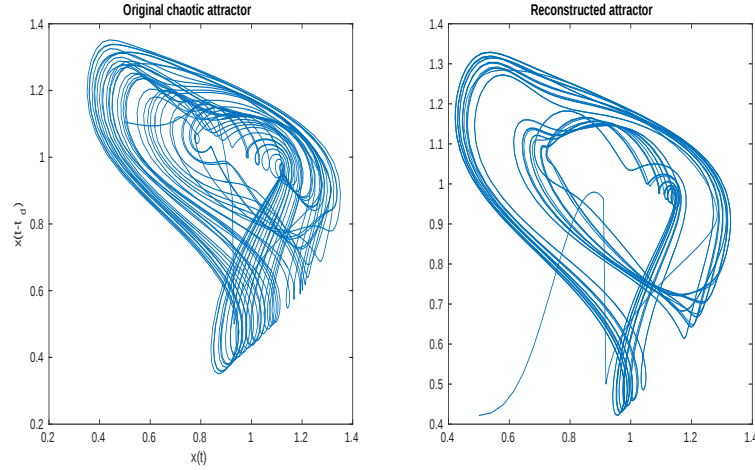


Figure 4.18: Comparison between the original attractor (left) and its reconstruction (right) based on the outputs of GP-ABC-OLCM.

4.4 Computational limitations of GP-ABC-SMC

The number of observations creates a significant problem for any GP method in terms of memory and computational time. Each step of likelihood optimisation and the inference requires inverting a matrix of size $n \times n$ with complexity $\mathcal{O}(n^3)$ and storage $\mathcal{O}(n^2)$. For GP-ABC-SMC we need to carry out GP regression once and thus as long as we are able to complete GP regression within a suitable time (≈ 200 seconds for the Lotka Volterra model as an example) and without running out of memory, the rest of the SMC steps involving simulating $\mathbf{f}(\widehat{\mathbf{X}}(t), \boldsymbol{\theta})$ is always faster than solving the corresponding non-linear ODE. On our machine (Intel(R) Core(TM) i7-2600 CPU 3.40GHz, 16 GB RAM) using the GPML package ([Rasmussen and Nickisch, 2010](#)), running GP regression on one of the species of Lotka Volterra for $n = 3000$ input time points require 305 seconds while using exact inference for the GP. Generation of ≈ 13000 simulated observations as required while running ABC-SMC-OLCM (see Table 2) takes around 2 hours, where we have calculated the time for generating a single simulated trajectory, by solving the Lotka Volterra ODE, to be ≈ 0.08 seconds using MATLAB's ODE45 routine. Hence for a similar model with ≤ 3000 input time points we recommend the use of GP-ABC-SMC over ABC-SMC using exact inference for the GP regression.

4.5 Conclusion

We ran several experiments using toy models to compare the statistical and computational performance of GP based ABC-SMC algorithm(s) to that of ABC-SMC with explicit integration. These experiments suggest that GP based ABC-SMC algorithm(s) can produce similar quality of estimates to that of ABC-SMC with explicit integration

while achieving a significant speed-up in the parameter estimation process. We found similar results for model selection using the GP based approach. We incorporated all the previous developments of ABC-SMC such as multivariate perturbation kernels and adaptive tolerance schedules in our implementation. We also noticed an increase in acceptance rates for the GP based approaches which indicates a smoothing (or annealing) of the distance function surface when the inference is performed on the derivative space.

We also compared our proposed approach with other GP based methods proposed in recent literature and found that our proposed GP-ABC-SMC (with the local multivariate perturbation kernel) performs significantly faster to obtain similar estimates. However, careful choices of prior distributions and other experimental parameters should be made in order to achieve true comparison between the ABC (GP-ABC-SMC) and exact MCMC (GP-ODE, GP-GM, GP-AGM) based gradient matching algorithms. We like to point out that none of the experiments in [Calderhead et al. \(2008\)](#); [Dondelinger et al. \(2013\)](#); [Wang and Barber \(2014\)](#) report the posterior estimates of the noise standard deviations. This parameter is extremely important for gauging the accuracy of the GP regression in estimating the ODE states. Particularly in [Dondelinger et al. \(2013\)](#) the posterior of some of the states of the signal transduction cascade (see Figure 7 in that paper found in the supplementary material) obtained through GP regression point to the fact that the noise standard deviation is wrongly estimated. This is because the GP mean function, for these states ($[R],[RS]$), pass through the data points. This can only be true if the noise amplitude is negligible compared to the signal amplitude, which is not the case in the signal transduction example.

Chapter 5

Mathematical modelling of plantwide electrical responses in higher plants exposed to ozone

5.1 Introduction

Plants respond to their environmental stimuli which are often localised in nature, by producing long distance signals. For example wounding stimuli such as pathogen attacks on the leaves, or salt stresses encountered through the root system evoke localized sensory responses that are coordinated throughout the plant body. Such complex coordination asserts the existence of long distance signalling mechanisms where the sensory information is communicated to distant locations and as a result the plant acts as a whole on the basis of such sensory information. In animals a rapid ionic/membrane potential driven signalling system integrates activities across the organism via the nervous system in addition to long distance hormonal and chemical signals. In recent literature long distance signalling systems for plants have been proposed that transfers information through signals such as electrical/ ion fluxes ([Mousavi et al., 2013](#); [Zimmermann et al., 2009](#); [Felle and Zimmermann, 2007](#)), levels of reactive oxygen species (ROS) ([Capone et al., 2004](#); [Miller et al., 2009](#)) hydraulic waves through xylem ([Christmann et al., 2007](#)). The role of calcium as an important mediator of systematic long distance electrical signals as a response to wounding is elucidated in [Zimmermann et al. \(2009\)](#); [Christmann and Grill \(2013\)](#). Also [Choi et al. \(2014\)](#) reported the existence of long distance calcium wave in response to a variety of stimuli. Thus expanding our understanding of these calcium waves and their roles in plant physiology, and especially its relation to the stimuli, is of great importance. One possible step towards assimilating this understanding may germinate from modelling the mechanism of the calcium wave. Such model may shed light on several aspects of a plant-wide regulatory information exchange and relating that

to the sensory perception. In this chapter we have studied the electrical responses at the surface (tissue) of *Solanum lycopersicum* (tomato) and *Cucumis sativus* (cucumber) plants on application of ozone as a stimulus. In [Mousavi et al. \(2013\)](#) surface electrical signal has been shown to induce jasmonate, triggering defence responses, at local and distant sites. Our primary aim is to build a mathematical/phenomenological model of such responses with a hypothesis of the involvement of calcium wave as a key coordinator of such responses. Such a model should be able to quantify the relationship between the stimulus, the electrical response and the hypothesised calcium wave. Furthermore, we have applied the GP-ABC-SMC methods to fit such model to data and also select the best among candidate models. We have used the time courses of electrical responses recorded from the surface of the plant to validate as well as calibrate our models.

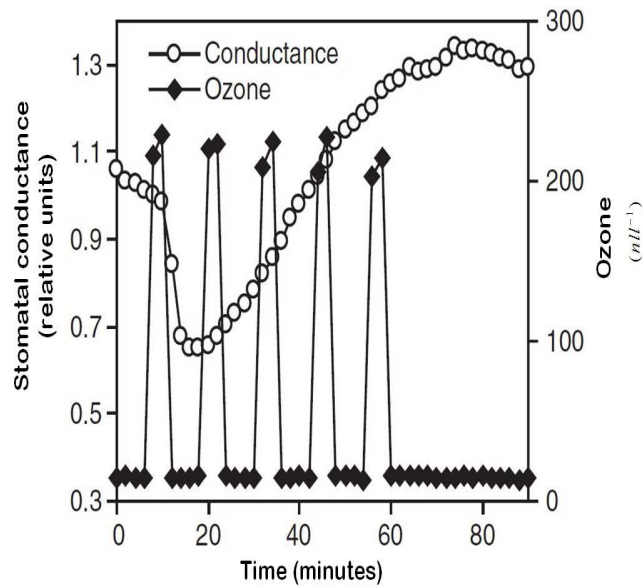
Although we are modelling the surface electrical potentials, much of our mechanistic intuitions are based on cellular electrophysiology and thus our work builds upon previous mathematical models of cellular electrical mechanisms. In regards to cellular electrophysiology mathematical models have been proposed for primarily explaining the action potential generation mechanism ([Gradmann and Mummert \(1991\)](#), [Sukhov and Vodeneev \(2009\)](#), [Beilby \(2007\)](#)). All these models provide a mechanistic explanation based on microscopic ionic currents that exist within a typically excitable plant cell. An analogous model had been proposed by [Sukhov et al. \(2013\)](#) to explain the variation potential. In this model the variation potential is argued to be generated by a turbulent influx of calcium ions in the cell, triggered by some unknown substance following a wounding suffered by the plant. This model describes a complex physiological process involving several species of microscopic ionic currents. There are other simpler models ([Buschmann and Gradmann \(1997\)](#), [Gradmann and Buschmann \(1997\)](#)) that attempted to explain oscillatory time course of the membrane voltage observed at the periphery of a plant cell. The minimal model of membrane voltage proposed in [Buschmann and Gradmann \(1997\)](#) is perhaps the best starting point for understanding the basic building blocks that constitute the electrical activities in a plant cell. Our proposed model borrows the dynamical properties of ion channels as shown in the [Buschmann and Gradmann \(1997\)](#), and extrapolates these dynamics to explain plant-wide electrical activities.

We like to point out that unlike the mechanistic models mentioned in the previous paragraph, our proposed model is phenomenological. Thus to describe the stimulus-response phenomenon succinctly, we solve an inverse problem using experimental observations. To the best of our knowledge, this is a point of departure from the tradition of plant electrophysiological research pertaining to mechanistic modelling. Phenomenological modelling of plant-wide, as in our case, or cellular electrical signals has not been explored yet.

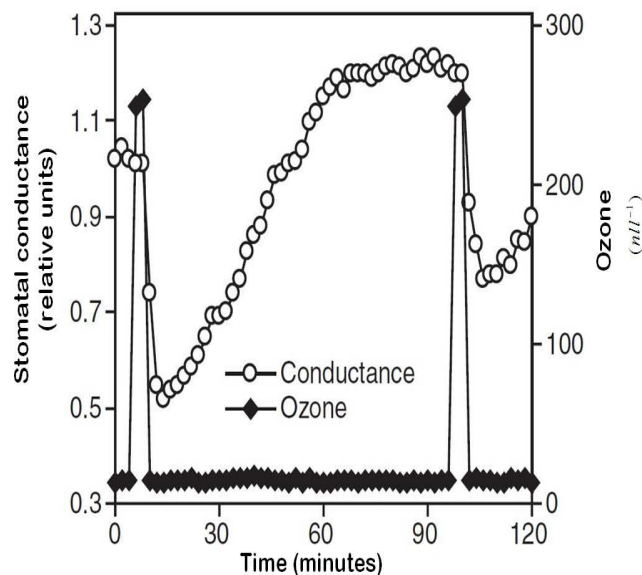
5.2 Ozone as stimulus: Experiment design and analysis

The air pollutant ozone can be used to study in planta processes including stomatal signalling. In [Vahisalu et al. \(2010\)](#) ozone was used to study the reactive-oxygen-species (ROS) dependent rapid transient decrease (RTD) of stomatal conductance¹ in intact *Arabidopsis* plants. It was shown in [Vahisalu et al. \(2010\)](#) that the RTD involves the production of ROS. An important property of the RTD is the existence of a refractory period (a period of insensitivity to further stimulus) between two consecutive RTD episodes. This refractory period is experimentally found to be 90 minutes (see Figure 5.1 for a depiction of this refractory period). Similar experiments in [Clayton et al. \(1999\)](#); [Evans et al. \(2005\)](#) show ozone induced rapid transient increase in systolic calcium. However, the existence of calcium as a part of RTD is not proven yet. It is worth noting that stress signals such as wounding or herbivore attack triggers the stress hormone jasmonic acid through a plant wide electrical wave. Recent studies ([Christmann and Grill, 2013](#)) have put focus on the role of calcium as a mediator of such responses. Thus it is a natural question to ask whether any stressful signal such as ozone would generate a fundamental defence signal - a plant wide electrical signal in addition to that of the ROS signal found in RTD. Also if such electrical responses emerge then, how are they controlled and managed by the calcium channels? To answer these questions it is necessary to study the responses (electrical) generated by the plants due to ozone exposure. This will potentially help us in modelling the quantifiable relationship between fundamental signals related to the innate stimulus-response signalling mechanism in plants. Especially, we want to investigate a plausible association of a calcium wave alongside the reported ROS in the generation of RTD. We thus conjecture the existence of a possible calcium wave that precedes the ROS signal where we interpret the electrical responses as a manifestation of a plant wide calcium current surge. To investigate the electrical response to ozone stimulus 2 sets of experiments were carried out. In the first set of experiments two plants from the same species were exposed to the simultaneous treatment of ozone. This treatment was repeated 10 times after an interval of 2 hours after each spray. 10 replicates of this experiment was done resulting in a total of 100 stimulus-response episodes of 2 hours each. The second experiment was motivated by the study of ozone induced RTDs. To verify the existence of a refractory period one of the experiment given in [Vahisalu et al. \(2010\)](#) was replicated in which one tomato plant was exposed for 6 consecutive ozone treatment at an interval of 15 minutes. For both sets of experiments the ozone exposition duration was of 1 minute. Details of the experimental setup are given in the following section.

¹ Stomata are small pores on the aerial parts of the plant that control carbon dioxide (CO_2) influx for photosynthesis and water vapour loss. Stomatal conductance is the measure of the rate of CO_2 influx and water vapour loss. This is usually measured in $mmol\,m^{-2}s^{-1}$.



(a)



(b)

Figure 5.1: Refractory period of the rapid transient decrease in stomatal conductance as found in experiments reported in [Vahisalu et al. \(2010\)](#). In the first experiment (a) successive ozone expositions have no further effects. However, when successive ozone expositions are separated in time for ~ 90 minutes or more (b) then a second episode of conductance decrease can be observed.

5.2.1 Experimental setup

The experiments were conducted within the framework of the PLEASSED project (PLants Employed As SEnsing Devices, **PLEASSED**) by members (see the list of authors affiliated to the University of Florence in [Chatterjee et al. \(2015\)](#)) of the International Laboratory

of Plant Neurobiology, University of Florence. Two of the major computational goals of this project were: i) exploration of classification strategies for identification of stimulus from plantwide electrical responses and ii) characterisation of such electrical responses using mathematical models.² Within this framework an initial experimental setup for measuring plantwide electrical responses subjected to varying amplitude of light as an external stimulus was proposed in Chatterjee et al. (2014). This setup was further refined and was used to carry out experiments with various stimuli including ozone in Chatterjee et al. (2015) for achieving the first goal of the PLEASED project and also for the inverse problem presented in this chapter. The second set of experiments involving shorter duration of ozone pulses were carried out by Dr. Ilaria Colzi from London Institute of Mathematical Sciences. For both sets of experiments tomato and cucumber plants, grown for 3 weeks, have been used. For the first set of experiments none of the plants are reused for replicate experiments. For each plant, we used three stainless steel needle electrodes - one at the base of the stem as a reference, one in the middle and the other on the top of the stem as shown in Figure 5.2. The electrodes were connected to the instrumentation amplifier (data Acquisition (DAQ) system) in the same way as previously studied in Chatterjee et al. (2014). A detail circuit diagram is shown in Figure 5.3. Plants were then enclosed in a plastic transparent box with proper openings to allow the presence of cables and inlet/outlet tubes, and exposed to artificial light conditions (LED lights responding to plant's photosynthetic needs, mimicking the day/night cycle of 12 hours). Each experiment was conducted in a dark room to avoid external light interferences. The whole setup was then placed inside a Faraday cage to limit the effect of electromagnetic interference as shown in Figure 5.2. After the insertion of the electrodes into the plant, we waited for about 45 minutes to allow the plant(s) to recover before starting the stimulations. Electrical signals acquired by the electrodes were provided as input to a 2-channel high impedance ($10^{15}\Omega$) electrometer (DUO 773) while data recording was carried out through 4-Channel DAQ (LabTrax) and its dedicated software LabScribe (Chatterjee et al., 2015). The sampling frequency was set as 10 samples per second. Ozone produced by a commercial ozone generator (Chatterjee et al., 2015), was injected into the box through a silicone tube, while a second outlet tube removed the ozone from the box to a chemical hood. The concentration of ozone inside the box was monitored using a suitable sensor at an interval of 15 minutes for the first set of experiments (inter-spray interval of 2 hours) and at an interval of 1 minute for the second set of experiment (inter-spray interval of 15 minutes). The plants were exposed to ozone of concentration 16 and 19 part per million for the respective sets of experiments.

²The author of this thesis was responsible for characterisation of electrical responses using mathematical models in the PLEASED project.

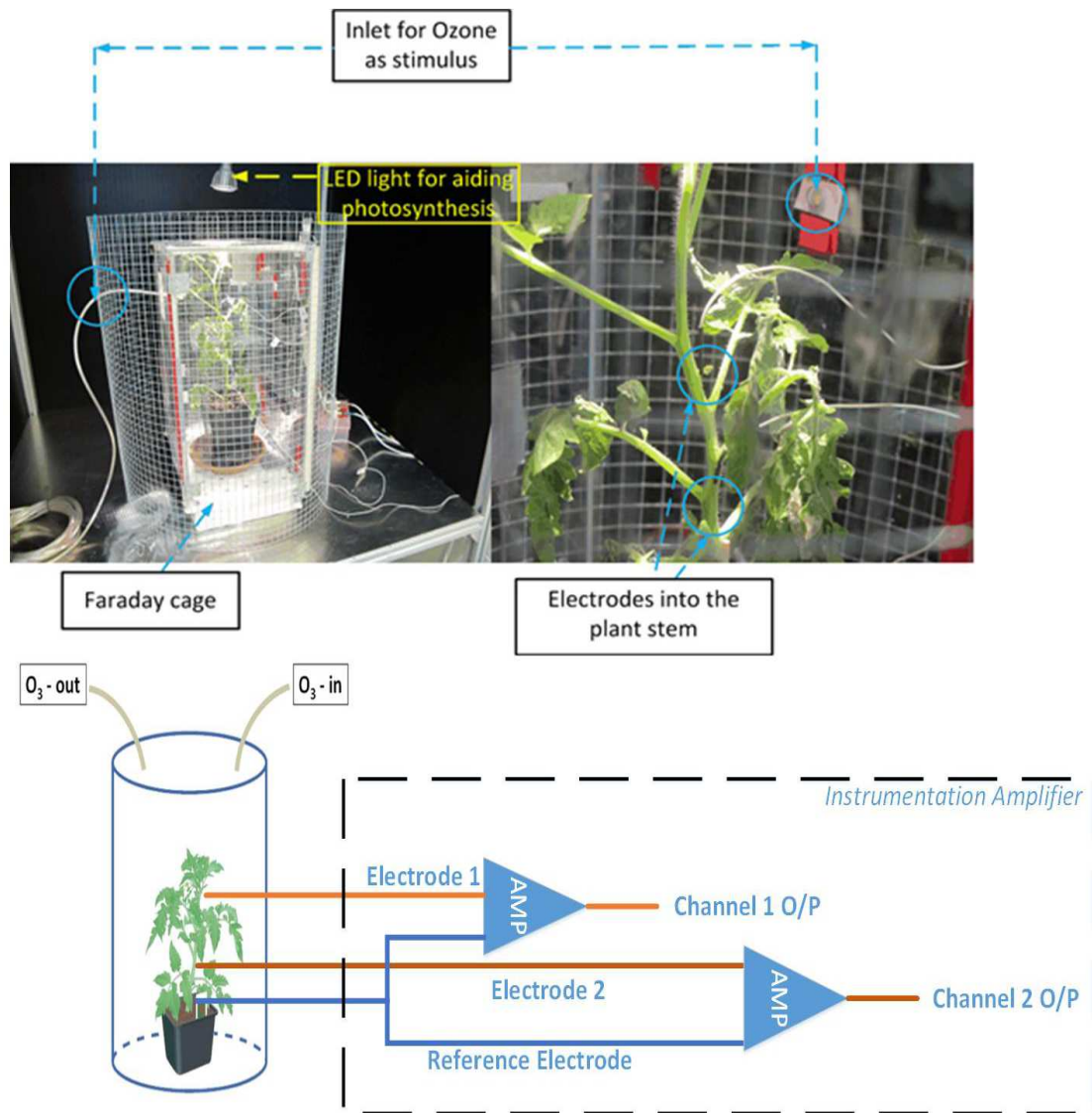


Figure 5.2: Experimental setup showing a tomato plant inside a plastic transparent box, kept inside a Faraday cage. The placement of the electrodes on the stem is also shown.

5.2.2 Electrical responses to ozone as stimulus

Our first experiment consists of an ozone stimulus applied repetitively at an interval of 2 hours on both the tomato and cucumber plants. Figure 5.4 shows plots of some typical electrical responses as contiguous time courses, representing a batch of ozone exposure on a cucumber plant, with the location (in time) of each ozone stimulus marked with a vertical line. The ambient diffusion of the applied ozone is also plotted as a time course of the measured concentration of ozone within the box and is denoted as $U(t)$ (see Figure 5.4(b)).

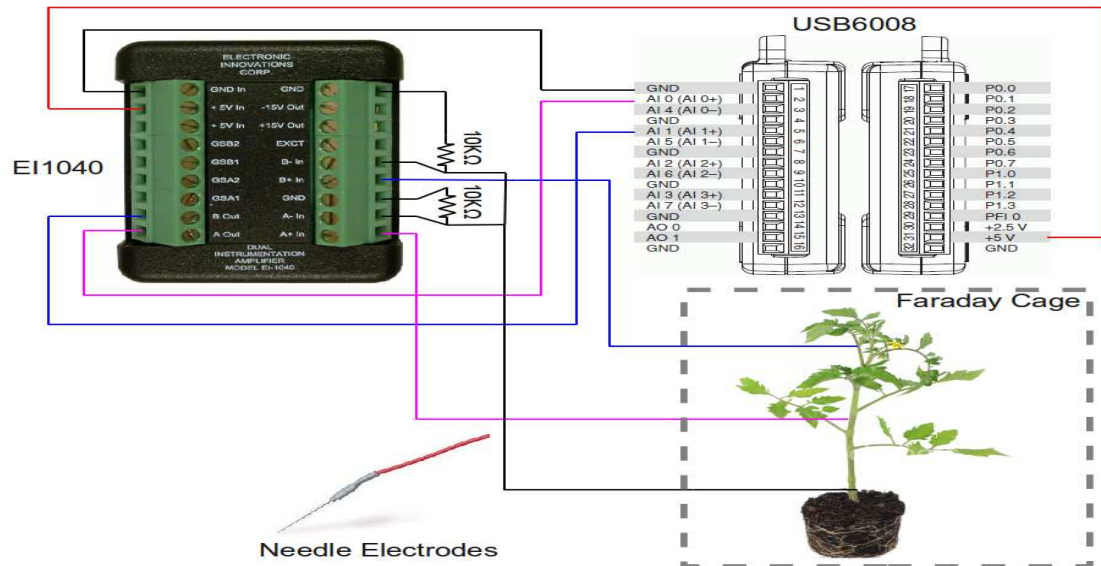
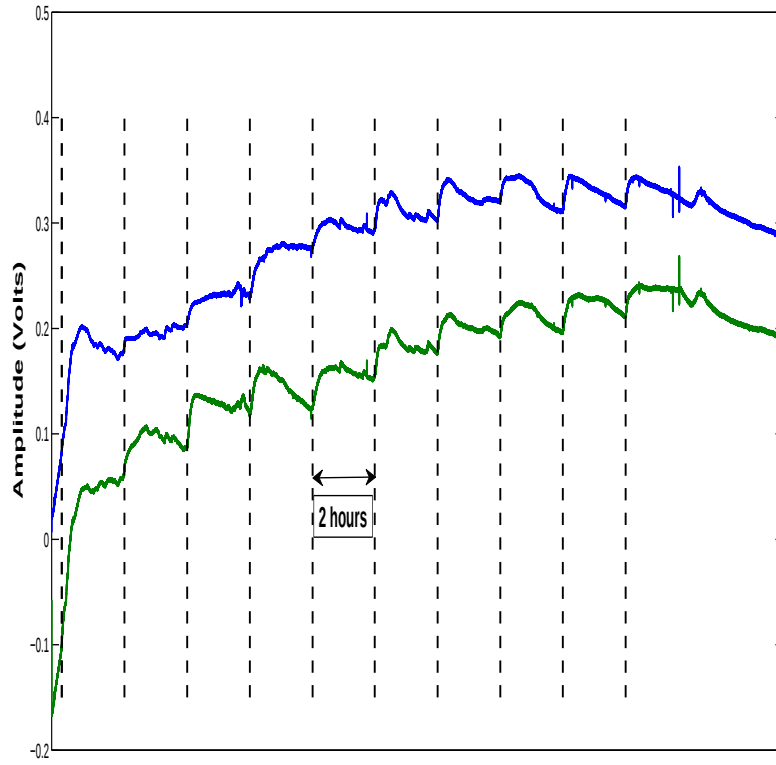
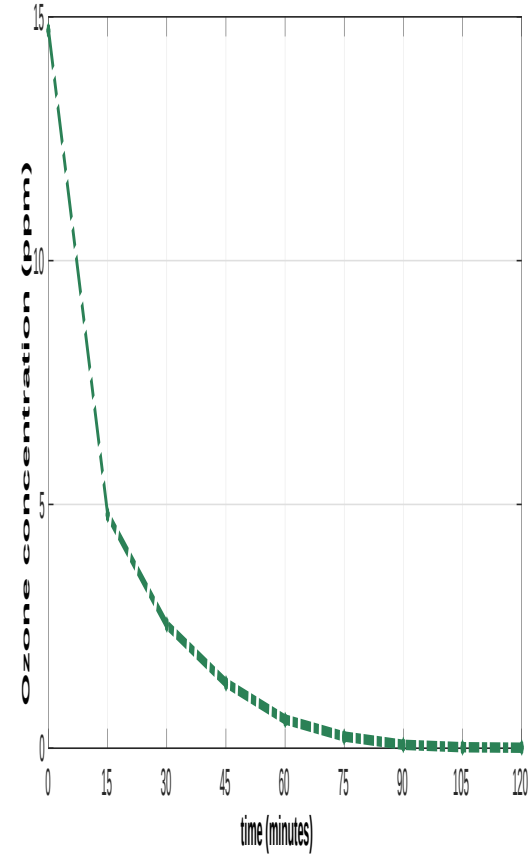


Figure 5.3: Circuit connection diagram for a dual channel instrumentation amplifier-data acquisition system



(a)



(b)

Figure 5.4: Typical electrical responses from the tomato plant. The blue and red curves indicate the amplitudes recorded with the 1st (bottom) and 2nd (top) electrodes on the stem. The black dashed curves (vertical) indicates the time location of the application of ozone stimuli.

The experimentally recorded voltage $V(t)$, at any one electrode location $x = \{1, 2\}$ which we denote as $V_x(t)$, is the difference between the electrical signals $S(t)$ (within the plant) picked up at the chosen electrode x and the reference electrode ref due to the nature of our circuit arrangement. From the circuit diagram (see Figure 5.3) and electrode arrangement (see Figure 5.2, within the dashed line) this fact can be observed. Each channel output is thus the output of a differential amplifier. Hence we have

$$V_x(t) = A(S_x(t) - S_{ref}(t)), \quad (5.1)$$

where A is the amplifier gain.

A fair degree of baseline wandering is present for each batch of experiments as evident from Figure 5.4. These could be attributed to sensor drift as well as an inherent steady state response $V_b(t)$. We also observe the marked ozone driven electrical response $V_d(t)$ as indicated by the upward curve of the amplitudes corresponding to each channel immediately after the ozone stimulus. The experimentally recorded voltage $V(t)$ (dropping electrode specification) is then given by

$$V(t) = V_d(t) + V_b(t). \quad (5.2)$$

We will follow the convention of denoting $V_d(t)$ as the excitation and $V_b(t)$ as the base (spontaneous) component of the electrical response. We are primarily concerned with only the excitation component $V_d(t)$ and thus we applied a moving average filtering to extract such responses as a necessary pre-processing step which we will explain next.

Consider an experimentally observed time series $X(t)$ consisting of T samples. Then the moving average $X_{MA}(t)$ is given by

$$X_{MA}^m(t) = \frac{X(t) + X(t-1) + \dots + X(t-m-1)}{m} \quad (5.3)$$

where $m \leq T$ is the span of the moving average. For our signals substituting the moving average $V_{b_{MA}}(t)$ of $V_b(t)$ in the above equation leads to an estimate of $V_d(t)$ which we denote as $\hat{V}_d(t)$. Figure 5.5 shows the raw as well as the smoothed time courses of the top channel from the same batch of experiment as shown in Figure 5.4. For understanding the nature of the excitation component which is our prime objective we subtract the amplitude as recorded by the top from that of the bottom channel. We denote this differential voltage $\delta\hat{V}_d(t)$ as the difference of amplitudes, after applying moving average filter, $\hat{V}_{d_1}(t)$ and $\hat{V}_{d_2}(t)$ at these electrodes on the stem respectively. This differencing has been done to gauge the pattern of evaluation of the amplitudes at both the electrodes simultaneously as we consider a plant wide current surge moving across the length of the plant as the primary response signal. We will follow this convention throughout this paper and by electrical response we will continue to mean the differential amplitudes considering the top two electrodes. The pattern of the electrical response $\delta\hat{V}_d(t)$ for an

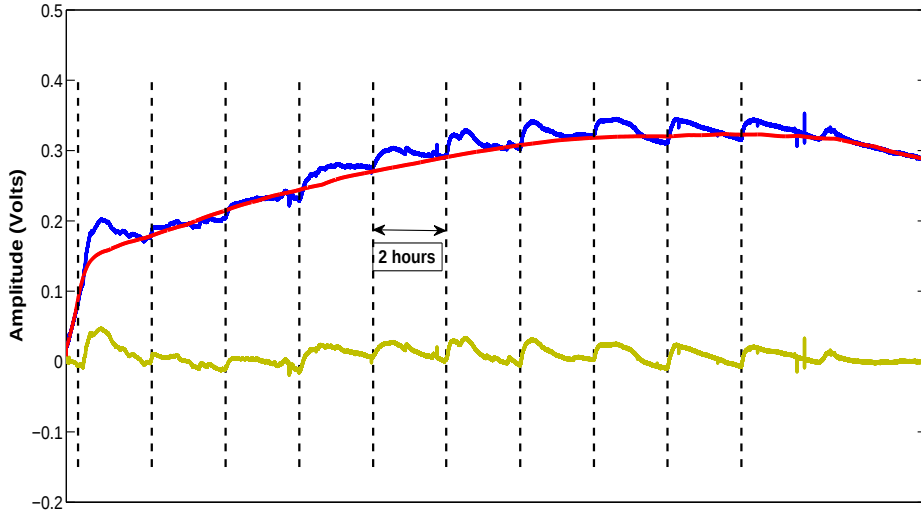


Figure 5.5: Electrical responses pertaining to the (estimated) excitation component $\hat{V}_d(t)$ obtained by smoothing is shown as the green curve for the same batch of experiment as shown previously. The blue curve indicates the net response $V(t)$ and the red curve is the smoothed response representing moving average $V_{b_{MA}}(t)$ of the base amplitude $V_b(t)$. All these amplitudes are based on the recording of the 1st (bottom) electrode.

ozone exposure episode can be identified, for many replicate batch experiments, with a sharp decrease in the differential amplitude followed by a gradual increase back towards the baseline (and beyond it in certain cases). See Figure 5.6 for an example of the differential amplitude for one of the replicate experiments consisting a batch of ozone expositions. For most of the experimental replicates we tend to see similar electrical responses as shown in Figure 5.6. However there are experiments with some episodes where an upward rise of $\delta\hat{V}_d(t)$ following an ozone exposition is observed (not shown here). Although the downward or upward (in some cases) deviance of the differential amplitude is prevalent in most of the stimulus-response episodes, there are a few episodes which do not show a recognizable change in the differential amplitude. Thus we have chosen to analyse and model those experimental episodes that show a downward deviance of the differential amplitude. It is worth mentioning that the direction of change in $\delta\hat{V}_d(t)$ is dependent on the subtraction convention (bottom to top electrode). With the chosen convention it is then necessary to identify the dominant pattern of the time evolution of $\delta\hat{V}_d(t)$.

As mentioned in section 5.2 we carried out a different experiment with shorter (15 minutes) ozone exposure period to enquire about the existence of a refractory period. In contrast to the findings in Vahisalu et al. (2010), in our second experiment the electrical responses (plotted as the differential amplitude) clearly depict the absence of any refractory period (see Figure 5.7). The ozone diffusion in this experiment is sped up

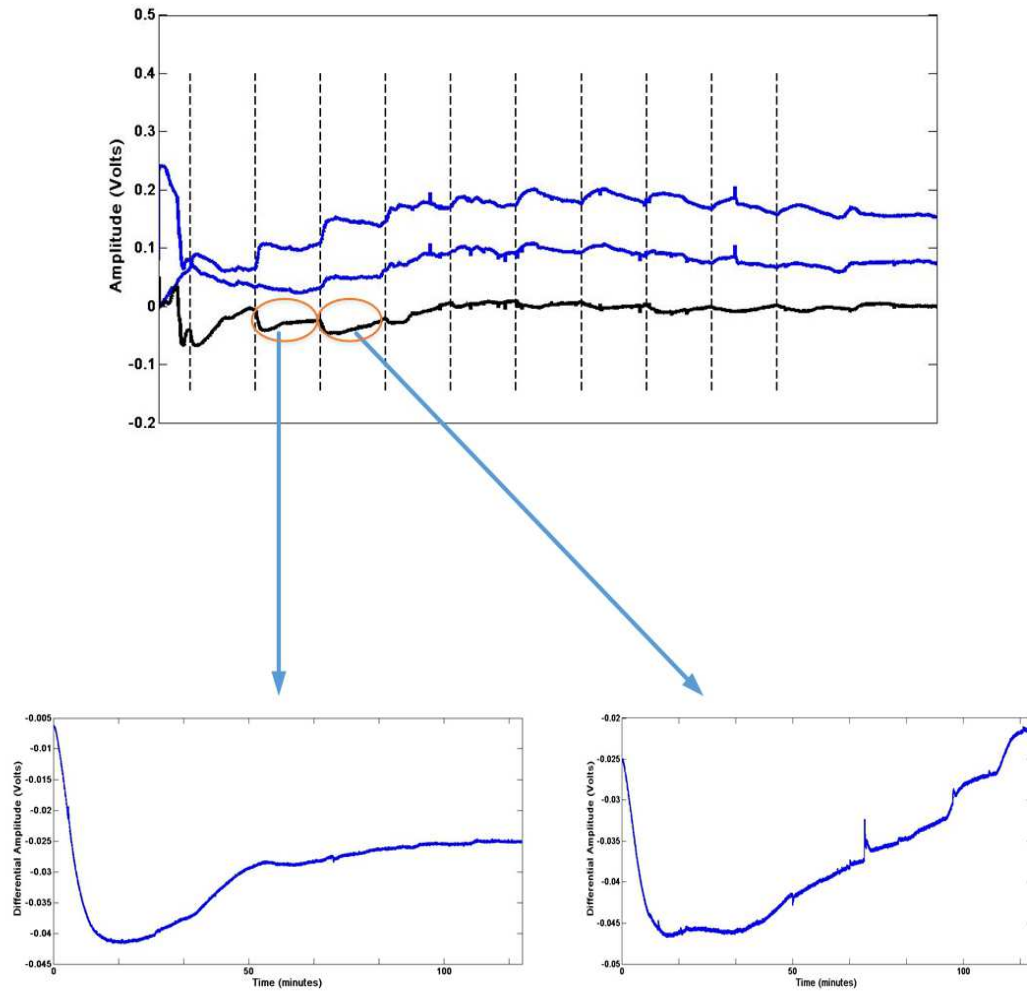


Figure 5.6: Differential response time course for a batch ozone exposures on a cucumber plant. The black curve shows differential excitation component while the blue curves show the corresponding amplitudes recorded at the two electrodes on the stem. The blue curves here show the net amplitude $\hat{V}_{d_1}(t)$ and $\hat{V}_{d_2}(t)$. The blown up sections show the differential amplitudes after the second and third ozone exposures within that batch.

considerably than the previous experiment using a faster outlet exhaust in the glass box thus producing a (almost) pulse type stimulus, in order to reproduce the stimuli duration in [Vahisalu et al. \(2010\)](#).

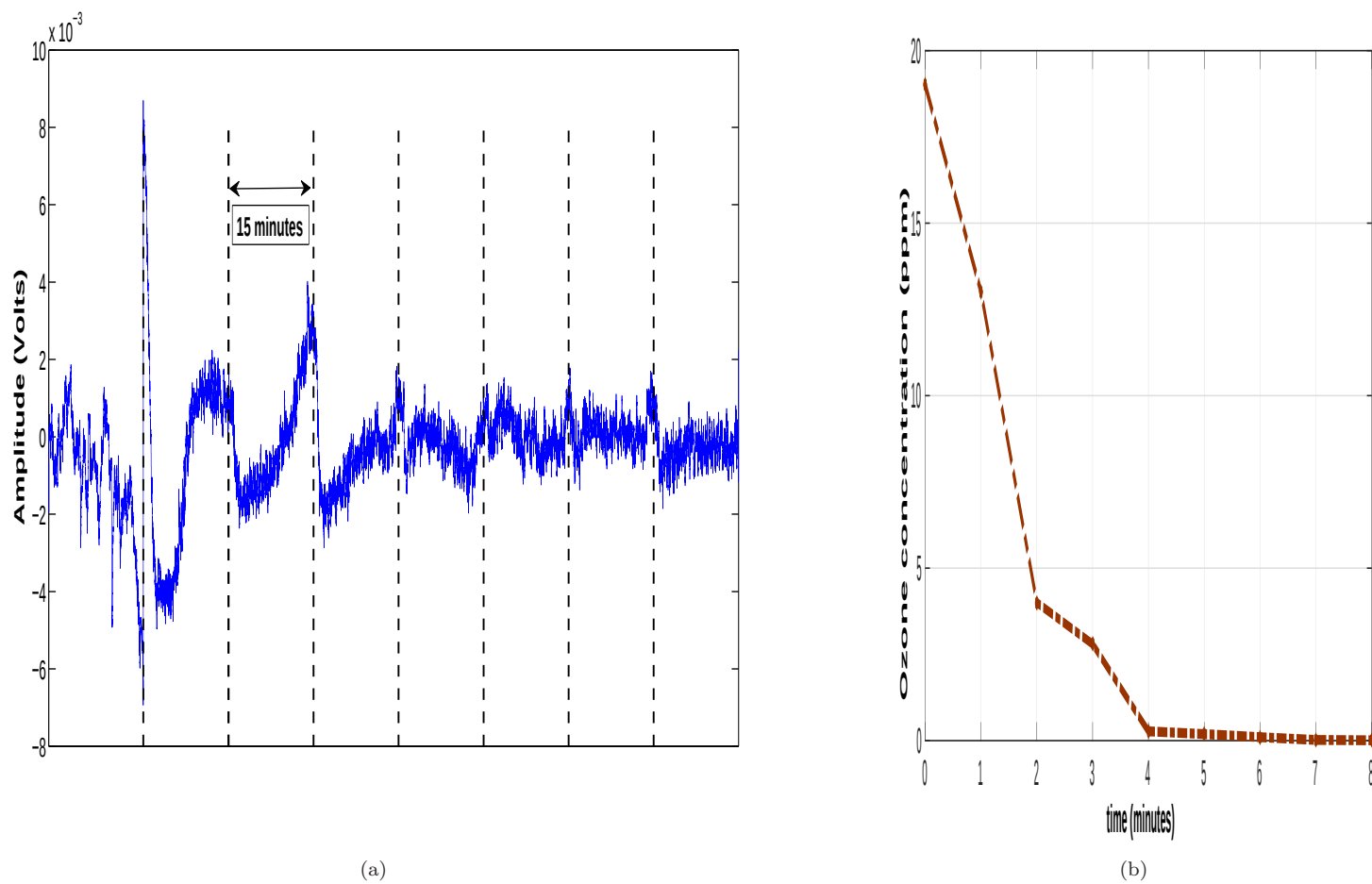


Figure 5.7: Typical electrical responses from the tomato plant using a shorter inter exposition time. The blue curve shows the differential amplitude obtained after smoothing. Here the stimulus is repeated after every 15 minutes.

From both sets of experiments (stimulus at 2 hours and 15 minutes intervals respectively) it is evident that the ozone stimulus produces a discernible electrical response having no refractory period. However, a gradual adaptation to the stimulus can be observed, notice the responses corresponding to the 4th and onwards ozone expositions in Figure 5.7 (experiment 2). Thus we hypothesise here that the electrical responses to ozone possibly have no ROS signal associated with it. This reinforces our assumption that the possible signal that is encoded in the electrical response is a calcium wave. We would now build a mathematical model around this assumption in the next sections and validate such models using these experimental observations.

5.3 Phenomenological model of calcium current

In order to establish a phenomenological characterisation of the electrical activities as evident from our experiments, we consider a hypothetical calcium current to be the fundamental signal and messenger of a plant wide response to the ozone stimulus, which we measure in the form of electrical signals. Thus we want to quantify, through a mathematical model, the dynamics of this current in relation to the stimulus. We consider the observable quantity to be the surface electrical potentials excitation component, the differential amplitude $\delta\hat{V}_d(t)$ mentioned in the previous section. Furthermore, we consider $\delta\hat{V}_d(t)$ to be the manifestation of the calcium current at the tissue (phloem in contact with the electrodes) level which is an agglomeration of the microscopic ionic currents. Thus from the modelling perspective the calcium current picked up by the electrodes as $\delta\hat{V}_d(t)$ is assumed to be a macroscopic average of the cellular current response. The source sites of these emissions include the guard cells which play a quintessential role in gas exchange and therefore stomatal modulations.

5.3.1 Cellular current model

The ion channels and associated currents that are directly affected by the ozone and the ROS compounds is yet to be understood. In Vahisalu et al. (2010) the SLAC1 (slow anion channel 1) and K^+ channels are hypothesized to be acting as the mediator of the rapid stomatal aperture modulations with substantial experimental support. However, through our experimental protocol it is impossible to probe and infer the exact signalling pathway for the generation of calcium current within the cell. Thus, we consider the following currents to be crucial in setting up a hypothetical mechanism leading to the generation of the calcium current.

We like to mention here that we lack a mechanistic understanding of how some of the molecular compounds produced upon ozone entering the plant interact with several ionic species found in plants. We first assume that such interactions disturb the electrical equilibrium within a plant cell. We then assume that the rate at which these interactions

disturb the equilibrium is proportional to the availability of ozone in the plant's environment, which is experimentally measured as $U(t)$. Although this availability of ozone is controlled by regulating the flow between inlet and outlet of the glass box, the resulting (molecular) interactions can vary between episodes giving rise to varied effects on all the other ionic species. With these assumptions we model these ozone induced interactions by considering a current $I_{oz}(t)$, in addition to other currents (corresponding to the ionic species found within the cell), that can potentially disturb the electrical equilibrium.

For all the ionic species, we treat the net contribution rather than individual ion channel effects as we lack sufficient information of individual channels role. This net current is denoted as $I_{ion}(t)$. Furthermore we consider a calcium current $I_{ca}(t)$ which is related to the change in the membrane potential $V_m(t)$. When the plant is exposed to ozone, $I_{oz}(t)$ is generated which drives $I_{ion}(t)$ affecting the charge balance within the cell resulting in the deviation of the membrane potential $V_m(t)$ from its equilibrium potential V_0 . This deviation affects the voltage gated (Hille, 2001) calcium channel's probability of being in an open state, generating a spurt of Ca^{2+} ion efflux resulting in the calcium current $I_{ca}(t)$. This mechanism is shown in Figure 5.8 where the cell is considered to be a Guard cell. The steady state open probability of the calcium channel under the voltage gating,

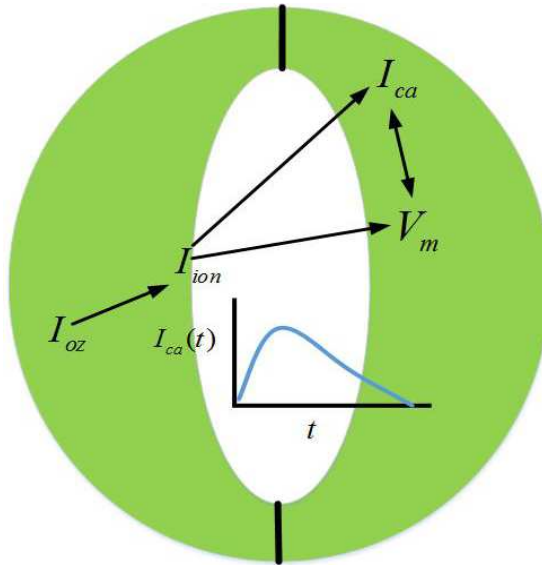


Figure 5.8: Interactions of ionic currents within a Guard cell after ozone exposition. A sample time course of the cellular calcium current is also shown as a magenta curve.

assuming a Markov switching (Buschmann and Gradmann, 1997) between its two states – open $O(t)$ and closed $C(t)$ is a reversible process $O \rightleftharpoons C$ given by

$$O(t) = \frac{1}{1 + e^{B(V_m - V_0)}} \quad (5.4)$$

where $r_{O/C} = e^{B(V_m - V_0)}$ is a voltage dependent rate constant of the reversible process. The calcium current is then given by

$$I_{ca}(t) = gO(t)(V_m - V_0) \quad (5.5)$$

where g is the calcium channel conductivity.

5.3.2 Macroscopic current model

We assume that the electrodes, when inserted in the stem, come in contact with an electrically conducting patch in the phloem, pick up the cellular calcium current $I_{ca}(t)$ from many cells within that patch. Our assumption here is that among many other ionic imbalances (leading to the generation of different ionic currents) that take place due to the ozone exposure, the significant current that the electrode picks up is the calcium current. This assumption is based on prior empirical observations found in [Schroeder and Hagiwara \(1989\)](#). The resulting current as picked up by the electrodes (along with the base potential $V_b(t)$) can be represented as the average calcium current $\mathbf{I}_{ca}(\mathbf{t}) = 1/N \sum_{k=1}^N I_{ca}(t)$, considering N cells generating individual calcium currents. We thus observe $\mathbf{I}_{ca}(\mathbf{t})$ experimentally as $\hat{V}_{d_i}(t)$, where i denotes the corresponding electrode number. Thus the net response (excitation) quantity $\delta\hat{V}_d(t)$ is the difference $\delta\mathbf{I}_{ca}(\mathbf{t}) = \mathbf{I}_{ca}(\mathbf{t})_1 - \mathbf{I}_{ca}(\mathbf{t})_2$, observed experimentally, between two locations on the phloem where the electrodes are inserted. The resulting scenario is depicted in [Figure 5.9](#).

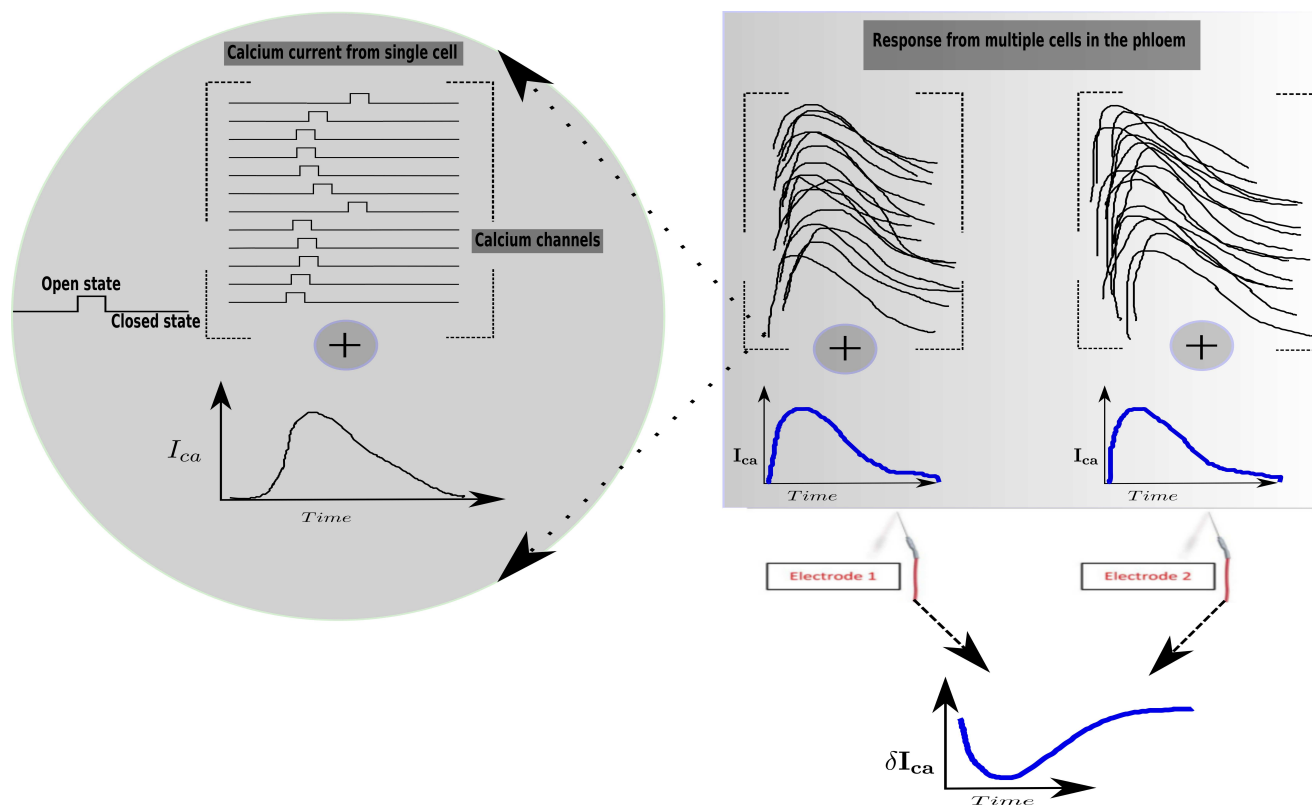


Figure 5.9: Hypothetical mechanism behind the net effect of cellular microscopic currents generated at two different locations (where the top two electrodes are inserted) resulting in a macroscopic calcium current along the phloem.

Along with $\mathbf{I}_{ca}(t)$ we can represent the other currents, established in the cellular model, at the tissue level using the same averaging and differencing principle. Note that the exact number of cells N cannot be determined or speculated. Thus we would now work with these average currents while describing a macroscopic model for the observed (as $\delta\hat{V}_d(t)$) calcium current $\mathbf{I}_{ca}(t)$. Now assuming that ozone affects a large number of guard cells that affect and results in generating individual calcium currents $I_{ca}(t)$ separated at the time instant of generation within the cells (see Figure 5.6), we can abstract the hypothetical mechanism behind the generation of $\delta\mathbf{I}_{ca}(t)$ within the cell to that at the tissue level where the patches of phloem are in contact of the electrodes. Thus we can now quantify the dynamics of $\mathbf{I}_{ca}(t)$ as coupled differential equations using the same cellular mechanism behind $I_{ca}(t)$. This coupled differential equation is given by

$$\begin{aligned} \frac{d(\delta\mathbf{I}_{oz})}{dt} &= -\mu\delta\mathbf{I}_{oz} \\ \frac{d(\delta\mathbf{I}_{ion})}{dt} &= k_1\delta\mathbf{I}_{oz} - k_{11}\delta\mathbf{I}_{ion} \\ \frac{d(\delta\mathbf{V}_m)}{dt} &= k_2\delta\mathbf{I}_{ion} + k_{22}\delta\mathbf{I}_{ca} - k_d\delta\mathbf{V}_m \\ \frac{d(\delta\mathbf{I}_{ca})}{dt} &= k_3\delta\mathbf{I}_{ion} + g\frac{\delta\mathbf{V}_m - V_0}{1 + e^{-k_B(\delta\mathbf{V}_m - V_0)}} - k_{ca}\left(\frac{\delta\mathbf{I}_{ca}^2}{k_m^2 + \delta\mathbf{I}_{ca}^2}\right) \end{aligned} \quad (5.6)$$

where each dynamical variable $\delta\mathbf{X} = \mathbf{X}_1 - \mathbf{X}_2$ is the difference of the average (over a collection of cells) currents and voltages at the tissue level. As we are interested in the characterisation of the phenomenon at a macroscopic level we will use these variables and the proposed equation 5.6 as the model of the ozone driven electrical excitation.

The first term in the expression of $\frac{d\delta\mathbf{I}_{ca}}{dt}$ quantifies the charge imbalance in each cell through the actuation of the differential ionic current $\delta\mathbf{I}_{ion}$ and the second term is an approximation of the open probability over a collection of cells at the electrode contact locations. We have introduced another term that models the homeostasis of the dynamics of the differential response $\delta\mathbf{I}_{ca}$ through a Hill function. This term accounts for the pumping of Ca^{2+} ions into the interior of the cells out of the cytoplasm following a calcium current generation. The model has 12 parameters $\theta = (\mu, k_1, k_{11}, k_2, k_{22}, k_d, k_3, k_B, k_{ca}, k_m, g, V_0)$ which define the rate constants and channel conductivities with parameters g, V_0 being approximations of the channel conductivity and equilibrium potential for a collection of cells. These phenomenological parameters define the rate of interactions over a collection of cells and macroscopic terms controlling the dynamics of $\delta\mathbf{I}_{ca}(t)$. We seek a probability distribution over each of these parameters by fitting this model to the experimental time courses of electrical responses $\delta\hat{V}_d(t)$. These probability distributions of the model parameter would then enable us to predict a range of behaviour of the model under each experimental outcome as a direct method of the validation of this model. Furthermore, such models will propagate the uncertainty in the parameters to the model behaviour as predicted time courses of $\delta\mathbf{I}_{ca}(t)$.

5.3.3 Ligand gated channel

For the cellular current model described in section 5.3.1, there can be an alternative description of the calcium current generation mechanism. Such an alternative mechanism consists of a ligand gated calcium channel (Hille, 2001) at its heart. A ligand gated calcium channel can be activated by initial efflux of calcium ions rather than a voltage gradient. Thus, for our model we can consider the charge imbalance within the cell caused due to the stimulus current $\delta \mathbf{I}_{\text{oz}}(t)$ leading to the efflux of some amount of Ca^{2+} ions from the cell. These ions activate the ligand gated calcium channels resulting in more calcium ions getting out of the cell. This mechanism can be interpreted as a calcium driven calcium current. The steady state open probability of the calcium channel under the ligand gating, considering the same two states – open $O(t)$ and closed $C(t)$ is now given by

$$O(t) = \frac{I_{\text{ca}}^2}{k_l^2 + I_{\text{ca}}^2}, \quad (5.7)$$

where k_l is the rate constant of the reversible process between the two states. Considering the above probability term we modify the differential equation for the differential calcium current, within the macroscopic model, in equation 5.6 as follows:

$$\frac{d(\delta \mathbf{I}_{\text{ca}})}{dt} = k_3 \delta \mathbf{I}_{\text{ion}} + g \frac{(\delta \mathbf{V}_{\text{m}} - V_0) \delta \mathbf{I}_{\text{ca}}^2}{k_l^2 + \delta \mathbf{I}_{\text{ca}}^2} - K_{\text{ca}} \delta \mathbf{I}_{\text{ca}}. \quad (5.8)$$

Note that for this model we have assumed a linear homeostasis term $K_{\text{ca}} \delta \mathbf{I}_{\text{ca}}$ attributing the non-linearity of the resultant time course of $\delta \mathbf{I}_{\text{ca}}(t)$ to the probability term only.

5.4 Model fitting to experimental data using GP-ABC-SMC

Having proposed a set of models our next endeavour is to estimate the parameters of the model (model fitting) using the GP-ABC-SMC algorithm. Before venturing on the said objective it is worth mentioning that some form of preprocessing is required to be able to run the GP-ABC-SMC algorithm. Also suitable choices about the algorithmic settings such as tolerance schedule, perturbation kernel and GP covariances have to be made alongside choosing appropriate prior distributions of the model parameters. In the next sections we would establish these setting and explain our choices.

5.4.1 Preprocessing

For the purpose of model fitting we considered many replicates of the differential excitation component $\delta \hat{V}_d(t)$ of individual stimulus-response episodes lasting for 2 hours as the observed experimental data, Y^d in ABC context. With the sampling frequency of

10 Hz we have 72000 samples for a single 2 hour episode. However, we have noticed that the time evolution of the excitation component of electrical responses occurs at a much slower time scale than the sampling frequency. Considering the differential amplitude of the two sample episodes shown in Figure 5.6, the peak amplitude is reached between 30–60 minutes. Hence we have downsampled the time series retaining the mean of every 500 samples resulting in only 143 samples for a 2 hour episode. We have purposefully ignored the details of amplitude fluctuations, such as an action potential, at a much higher time scale as we have modelled the evolution of the excitation component for the entire time duration between two successive ozone exposures under a batch experiment. We considered here the time evolution of stomatal responses to ozone, which also occurs at similar time scales (Vahisalu et al., 2010), as an affirmation of our choice. Figure 5.10 compares the voltage traces of the electrical potential, for one episode of ozone exposures, to its downsampled version.

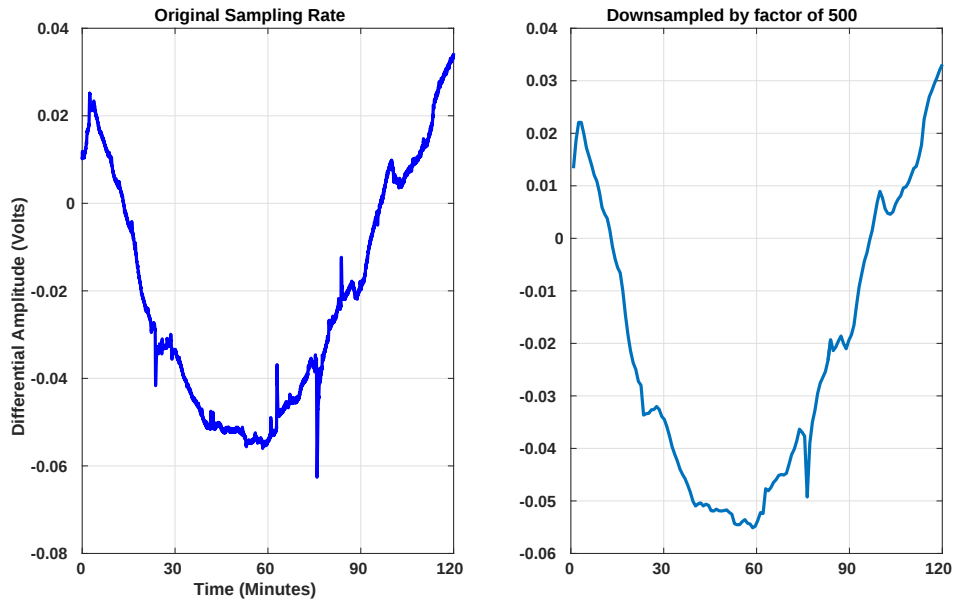


Figure 5.10: The time series on the left portrays one episode represented using the original sampling rate (10 samples / second) and the plot on the right is its downsampled version.

Many of the parameters of our model are correlated and many separate parameter combinations could result in similar time courses of the dynamic variables. Thus a multivariate kernel with a local covariance (Filippi et al., 2013) (OLCM perturbation kernel) which is well suited for tackling parameter correlations is chosen as the perturbation kernel, generating samples in ABC-SMC, to efficiently navigate the parameter space. Furthermore, we have chosen an adaptive tolerance schedule where tolerance ϵ_τ for an intermediate SMC step τ is set to the $\alpha = 0.1$ quantile of the distances $\Delta(Y^d, Y^s) \leq \epsilon_{\tau-1}$ in the previous (particle) population corresponding to the step $\tau - 1$.

5.4.2 Prior distributions

The choice of prior distributions is one of the most crucial aspects of Bayesian inference, indeed considering the fact that we have already constructed the best model that explains the regularities in the data sufficiently. These priors can be constructed from a first principle understanding of the physical properties of a system that these parameters represent or control. However, in case of inverse problems, involving a phenomenological model such as ours the parameters do not capture precise physical properties. Rather in our model the parameters control the collective effects of the dynamical variables to recreate the phenomenon of interest. Thus we need to find alternatives to first principle approaches towards engineering a particular prior distribution for the parameter(s).

This alternative approach consists of using point estimates of the parameters to build the prior distributions. To explain this prior construction let us first introduce the variables of interest in the context of a voltage gated model (same technique can be used for the ligand gated variant). We consider the observed data $\delta\hat{V}_d(t)$ (the differential amplitude) to be given by

$$\delta\hat{V}_d(t) \propto \delta\mathbf{I}_{\text{ca}}(t) + \eta(t) \quad (5.9)$$

where $\eta(t)$ is i.i.d Gaussian measurement noise. In the above equation we are explicitly assuming that the differential current response is measured with noise as the differential voltage excitation component, obtained as the difference between the output of the instrumentation amplifiers. Now if we assign a GP prior on the state of the differential calcium current then we have

$$p(\delta\mathbf{I}_{\text{ca}}(t^L)|t^L) = \mathcal{N}(\delta\mathbf{I}_{\text{ca}}(t^L)|0, K(t^L, t^L)), \quad (5.10)$$

considering $L = 144$ downsampled ³ experimental time points. Using the GP machinery we can thus evaluate the smoothed derivative of the state, which we denote as $\delta\hat{\mathbf{I}}_{\text{ca}}(t) = E[\frac{d}{dt}\delta\mathbf{I}_{\text{ca}}]$. Also from equation 5.6 we have the functional form for the right hand side $\mathbf{f}(\delta\hat{\mathbf{I}}_{\text{ca}}(t), \phi)$ of the derivative of $\delta\mathbf{I}_{\text{ca}}(t)$. Since only the state evolution of $\delta\mathbf{I}_{\text{ca}}(t)$ is observed as the differential voltage measurements we used the technique to construct the hidden variables $\delta\mathbf{I}_{\text{oz}}(t)$, $\delta\mathbf{I}_{\text{ion}}(t)$ and $\delta\mathbf{V}_{\text{m}}(t)$ as in section 3.6.4 to represent the right hand side $\mathbf{f}(\delta\hat{\mathbf{I}}_{\text{ca}}(t), \phi)$ of $\delta\hat{\mathbf{I}}_{\text{ca}}$ which is coupled to these hidden variables. Furthermore, we have augmented the parameter vector ϕ by adding three more parameters which are the initial values $\delta\mathbf{I}_{\text{oz}}(t=0)$, $\delta\mathbf{I}_{\text{ion}}(t=0)$ and $\delta\mathbf{V}_{\text{m}}(t=0)$ of the hidden states.

We have then used non-linear least squares for minimizing the Euclidean distance $\Delta(\delta\hat{\mathbf{I}}_{\text{ca}}(t), \mathbf{f}(\delta\hat{\mathbf{I}}_{\text{ca}}(t), \phi))$ between the velocity field $\delta\hat{\mathbf{I}}_{\text{ca}}(t)$ obtained by GP smoothing and $\mathbf{f}(\delta\hat{\mathbf{I}}_{\text{ca}}(t), \phi)$. This results in the point estimate of the optimal parameter vector $\hat{\phi}$

³For a two hour episode we have 72000 samples. Now downsampling by keeping the average of every 500 samples result in $L = 72000/500 = 144$ time points.

given by

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \sum_{i=1}^L (\delta \widehat{\mathbf{I}}_{\text{ca}}(t^L) - \mathbf{f}(\delta \widehat{\mathbf{I}}_{\text{ca}}(t^L), \phi))^2. \quad (5.11)$$

We have used the Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) algorithm with many initial guesses and chose those parameters which generated the minimum distance following Transtrum and Qiu (2012). Having obtained the point estimate for the parameter vector

$$\hat{\phi} = (\hat{\mu}, \hat{k}_1, \hat{k}_{11}, \hat{k}_2, \hat{k}_{22}, \hat{k}_d, \hat{k}_3, \hat{k}_B, \hat{k}_{ca}, \hat{k}_m, \hat{g}, \hat{V}_0, \delta \widehat{\mathbf{I}}_{\text{oz}}(t=0), \delta \widehat{\mathbf{I}}_{\text{ion}}(t=0), \delta \widehat{\mathbf{V}}_{\text{m}}(t=0)) \quad (5.12)$$

including the initial values, we build a prior distribution for each of its elements by constructing a uniform distribution bounded between $\pm K \times \hat{\phi}_j$ where j indexes each element of $\hat{\phi}$ and K is a multiplication factor controlling the spread of the support of these uniform priors. So for example we have the prior distribution for μ as $\mathcal{U}(\hat{\mu} - K \times \hat{\mu}, \hat{\mu} + K \times \hat{\mu})$. We have constructed the prior distributions for rest of the phenomenological parameters in a similar way. Figure 5.11 shows the histogram plot of the parameter vector obtained through this optimisation process considering 54 episodes from all the batches that show a discernible of response to ozone exposure. We have tried to summarize their

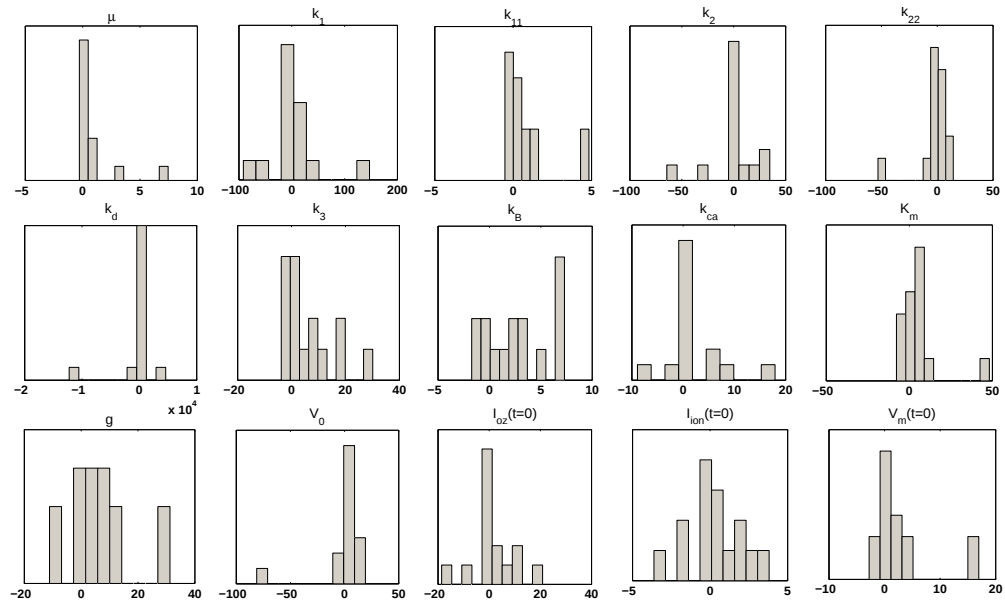


Figure 5.11: Histogram plots for each of the model parameters. Histograms are constructed from the collection of point estimates found after minimizing the r.h.s of equation 5.11. In total 54 episodes from different batches of experiments are used.

dynamics by reconstructing (by numerically solving the ODE system in equation 5.6) the dynamical variables of the model using the mean points of the histogram as parameter values. The said reconstructions are shown in Figure 5.12. The point estimates of the initial values for hidden variables are used for solving the model ODE. The initial

value $\widehat{\delta \mathbf{I}_{ca}}(t=0)$ is chosen as the mean of the first sample of the smoothed time course $\widehat{\delta \mathbf{I}_{ca}}(t)$ obtained using GP regression.

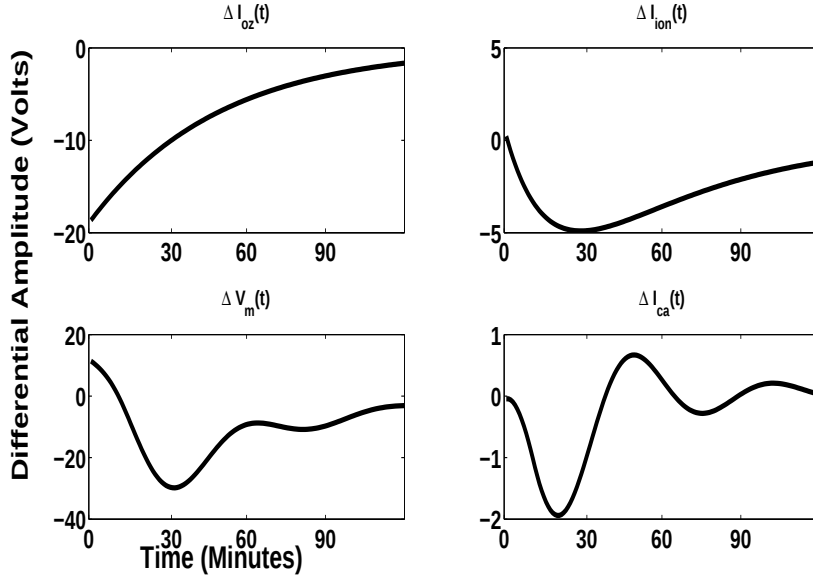


Figure 5.12: Reconstructed trajectories of the dynamical variables in equation 5.6.

The choice of the GP covariance is much more tricky as we need to have a working idea about the smoothness of the actual time evolution of $\delta \mathbf{I}_{ca}(t)$. This becomes extremely difficult to assess a priori due to the amount of variability in the experimental dataset. Taking a closer look at some of the episodes it is evident that the time evolution of $\delta \mathbf{I}_{ca}(t)$ appears as the sum of exponentials. So, for example if we consider a particular episode as shown in Figure 5.13 to which we want to fit a GP, our initial choices of covariance functions can be either a Matern kernel:

$$K_{Matern}(t, t') = \left(1 + \frac{\sqrt{3}(t - t')}{l}\right) \exp\left(-\frac{\sqrt{3}(t - t')}{l}\right), \quad (5.13)$$

or a squared exponential kernel:

$$K_{SE}(t, t') = \sigma_{kernel}^2 \exp\left(-\frac{1}{2} \frac{(t - t')^2}{l^2}\right), \quad (5.14)$$

both of which are well suited for modelling smooth exponential time series. Now, if we compare the negative log marginal likelihoods of GPs comprising each of these kernels then clearly the Matern kernel describes the data best. Consequently the smoothed time series of $\widehat{\delta \mathbf{I}_{ca}}(t)$ (black curve in Figure 5.13(b)) obtained using a Matern kernel tracks the data $\delta \hat{V}_a(t)$ (red curve) more closely than the exponential covariance (see Figure 5.13(a)) at around 30-th and 60-th minute time points. However, in lieu of our modelling assumptions we consider the calcium current $\delta \mathbf{I}_{ca}(t)$ trajectory as an

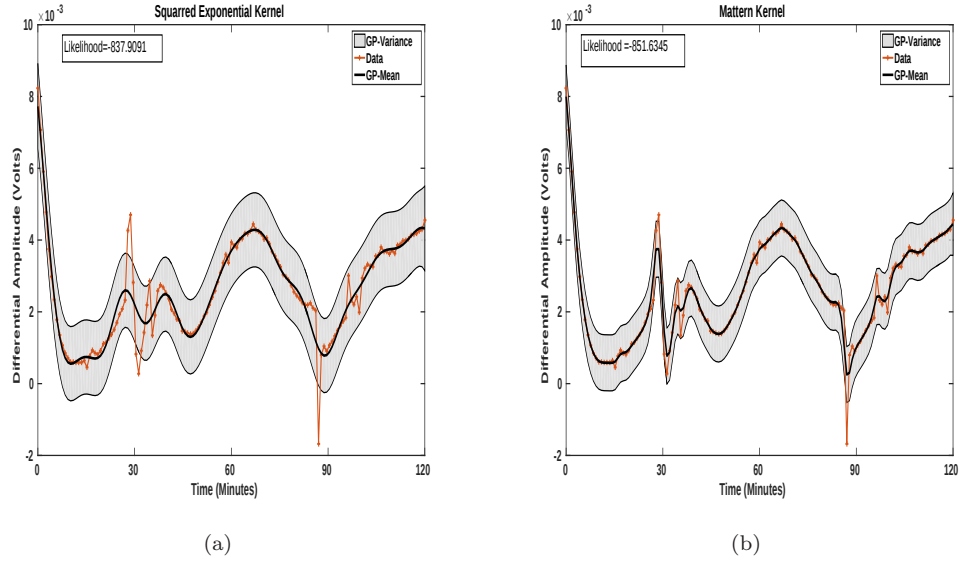


Figure 5.13: Comparison of smoothed time course $\widehat{\delta\mathbf{I}_{ca}}(t)$ (black curve) achieved using two different GP covariance kernels, the squared exponential kernel in (a) and the Matern kernel in (b) on a representative episode (red curve).

excitation that reaches its maximum between 30 to 60 minutes since its inception and any other excitation on top of that at a faster time scale (as seen in Figure 5.13 around 30-th and 60-th minute time points) cannot be related to the stimulus and thus need not be accurately modelled. Thus, we argue that the exponential kernel is a better choice to obtain a smoother representation of the time course of $\delta\mathbf{I}_{ca}$ than what is achieved using the Matern. Hence, we have consistently used the squared exponential covariance throughout the model fitting process.

5.4.3 Results

We used the algorithmic settings (GP covariance, tolerance schedule, perturbation kernel) and prior distributions as they were introduced in the previous sections while applying the GP-ABC-SMC on the experimental data consisting of consecutive replicate episodes from two batches of experiments, one from each species of plants. We like to point out that with each run of the algorithm we fit our model to one single episode among the consecutive ones and then repeat the algorithm for the next episode. Within the context of GP-ABC-SMC we have used the Euclidean distance $\Delta(\widehat{\delta\mathbf{I}_{ca}}(t), \mathbf{f}(\widehat{\delta\mathbf{I}_{ca}}(t), \phi))$ given in equation 5.11 at every SMC iteration to be compared against the tolerances. We have not used a fixed S_{MC} value, the number of SMC steps. While working with real data and a phenomenological model such as ours, it is not possible to correctly guess the ideal number of SMC steps (or the final tolerance value) that might produce a refined posterior distribution, since we do not have the slightest hint on the ideal parameter ranges. Rather, we keep on running the algorithm until

$|\epsilon_{\tau-1} - \epsilon_{\tau}| \leq Cstop$, where $Cstop$ is a constant chosen according to the peak-to-peak amplitude range for individual episodes. In general we have chosen $Cstop = 0.001$ for those episodes where the peak-to-peak amplitude is less than 0.005 Volts and $Cstop = 0.1$ otherwise. We have tried using other values for $Cstop$ in the same range and similar fit is obtained. While choosing the value of the multiplication factor K in the prior distribution, we have to simultaneously monitor its effect on population degeneracy. Thus we ran the GP-ABC-SMC with multiple choices of K and chose the maximum out of those which ensures $ESS \geq 60$ for the final population of particle. Here we have considered $N = 100$ particles. We again point out that this is a small population size. However, this is a first study of this inverse problem where we are primarily trying to show the applicability of the GP-ABC-SMC algorithm for carrying out a complex inference task. Thus we believe the chosen population size is adequate for this first study.

Voltage gated channel model

The model fitting results, for the voltage gated model (see equation 5.6), are presented in Figure 5.15 & 5.16 where we have included 8 (tomato) and 8 (cucumber) consecutive episodes from each of the plant species respectively. We have compared the time course of $\delta\hat{V}_d(t)$ with the reconstructed time courses of $\delta\mathbf{I}_{ca}(\phi, t)$ for each episode under consideration. This reconstruction is done by solving the model differential equation 5.6 to obtain the time course of $\delta\mathbf{I}_{ca}(\phi_{SMC}^{(i)}, t)$ using the particles $\{\phi_{SMC}^{(i)}\}_{i=1, \dots, N}$ from the final population of GP-ABC-SMC which (as a collection) represent the (approximate) posterior distribution $p(\phi | \Delta(\hat{\delta\mathbf{I}_{ca}}(t), \mathbf{f}(\hat{\delta\mathbf{I}_{ca}}(t), \phi)) \leq \epsilon_{SMC})$. Through this reconstruction the uncertainty in the parameters propagate to the uncertainty in the model predictions. Thus for those episodes where the posterior estimates of the parameters have higher levels of uncertainty, the reconstructions show much higher range of variability. The episodes shown in Figure 5.15(a) and 5.15(f), both from batch 1 (tomato), exemplifies such higher uncertainties in model predictions. The time courses of the episodes from the second batch in Figure 5.16, for cucumber plant, clearly show evidence of adaptation to the stimulus. The last few episodes, Figure 5.16(e)-5.16(h), has much lesser amplitude than the preceding ones. This shrinking in the net response is accurately picked up by the inference algorithm and reflected in the increased uncertainty of the model predictions. By the prediction we mean reconstruction in this context. Notice the reconstructed time courses (yellow curves in Figure 5.16) which are much widely spread in plots 5.16(e)-5.16(h) than the preceding plots. The change in uncertainty in model predictions, indicating a regime change, is firstly reflected in the posterior distributions of the parameters. We have plotted the posterior parameter distributions of two episodes from batch 2 in Figure 5.14 corresponding to Figure 5.16(d) & 5.16(e) which clearly show the change in levels of response to stimulus. 10 out of the 15 parameters show an increase of the support of the posterior (marginal) densities. The corresponding densities of these 10 parameters are plotted using a red curve whereas the remaining parameter's densities are plotted using a magenta curve. Clearly such increase in the

support accounts for increased variance of the posterior distributions thus indicating the said drop in amplitude as a result of adaptation to stimulus. However, the dynamics

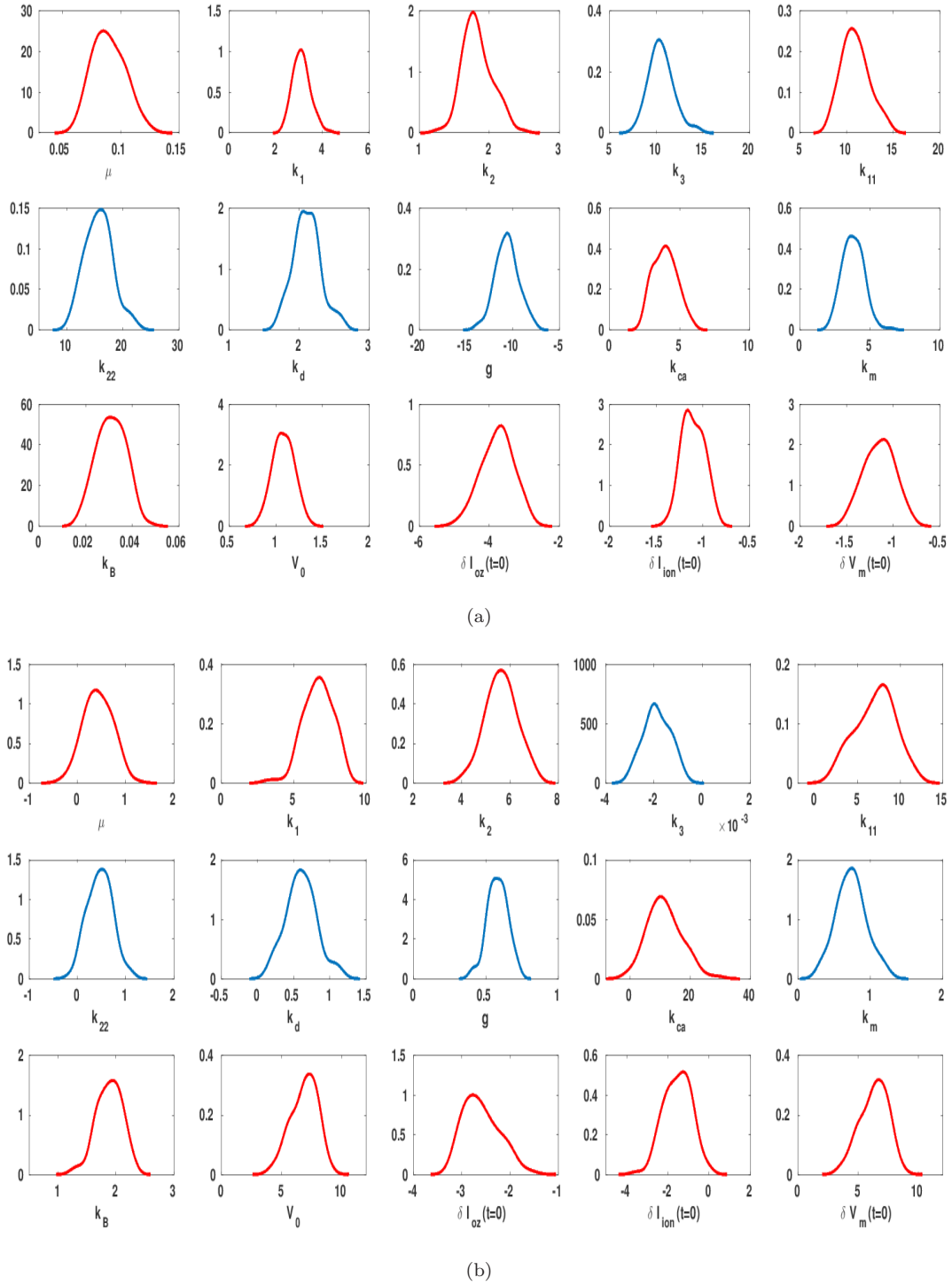


Figure 5.14: The posterior densities of each of the model parameters as obtained after applying the GP-ABC-SMC on episodes 4 (see Figure 5.16(d)) and 5 (see Figure 5.16(e)) from batch 2 corresponding to cucumber plant. We have used kernel density estimates rather than histogram to ensure that the spread of the support of these densities are clearly visible.

of the other variables $\delta \mathbf{I}_{\text{ion}}(t)$, $\delta \mathbf{V}_{\text{m}}(t)$ and most importantly $\delta \mathbf{I}_{\text{oz}}(t)$ (plots not shown) does not show a regime change, indicating adaptation in their own dynamics for the last few episodes. Rather, these quantities show more variability in their time evolution than that of $\delta \mathbf{I}_{\text{ca}}(t)$, making it difficult to interpret any underlying regularities that underpin the observations of the calcium wave through $\delta \mathbf{I}_{\text{ca}}(t)$.

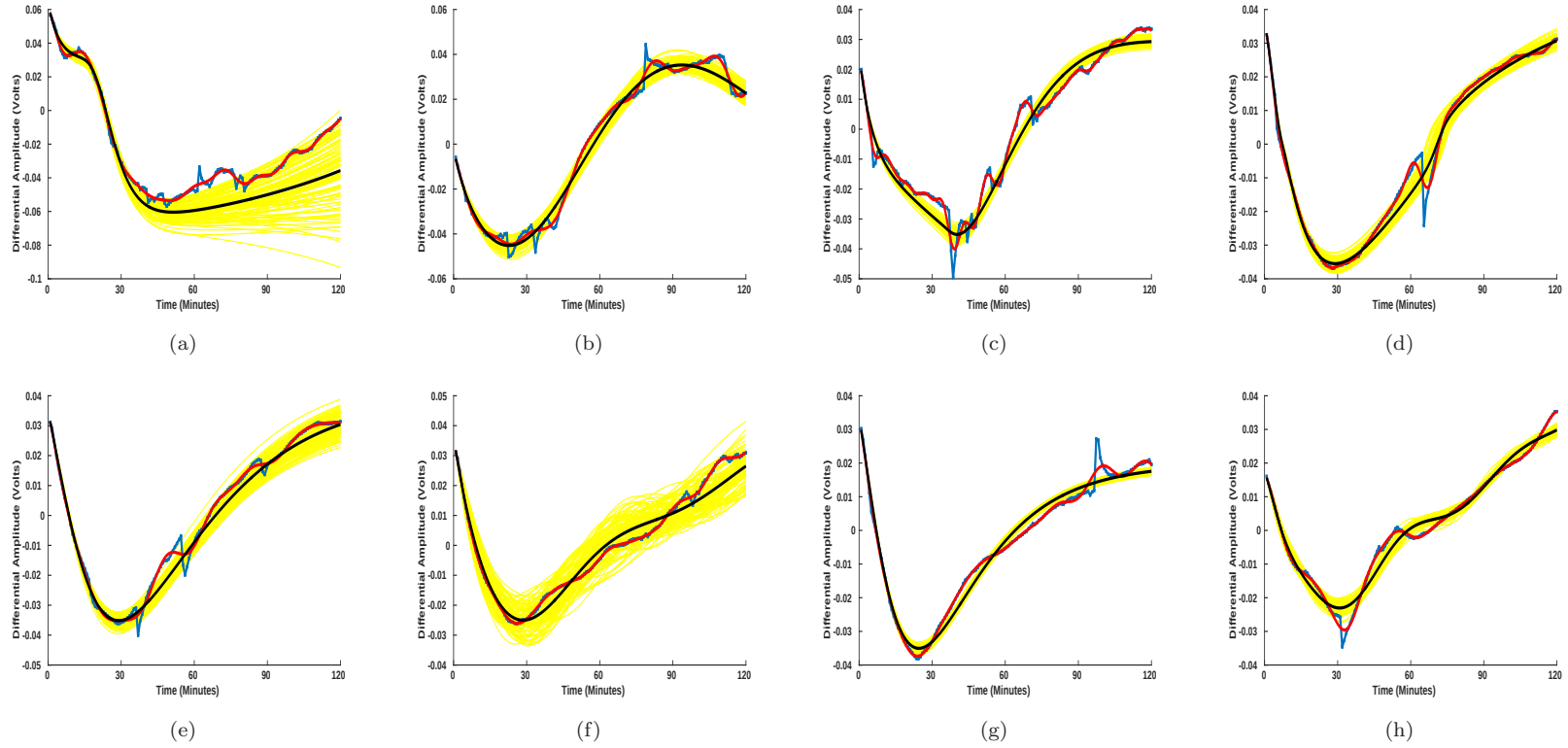


Figure 5.15: Model fitting results after applying the GP-ABC-SMC algorithm on episodes from batch 1, using the voltage gated model. Experimental data $\delta\hat{V}_d(t)$ is shown as blue curves. The reconstructed time courses of $\delta\mathbf{I}_{ca}(\phi, t)$ are plotted as yellow curves. The smoothed time courses of $\widehat{\delta\mathbf{I}_{ca}}(t)$ as obtained through the GP regression are plotted as the red curves.

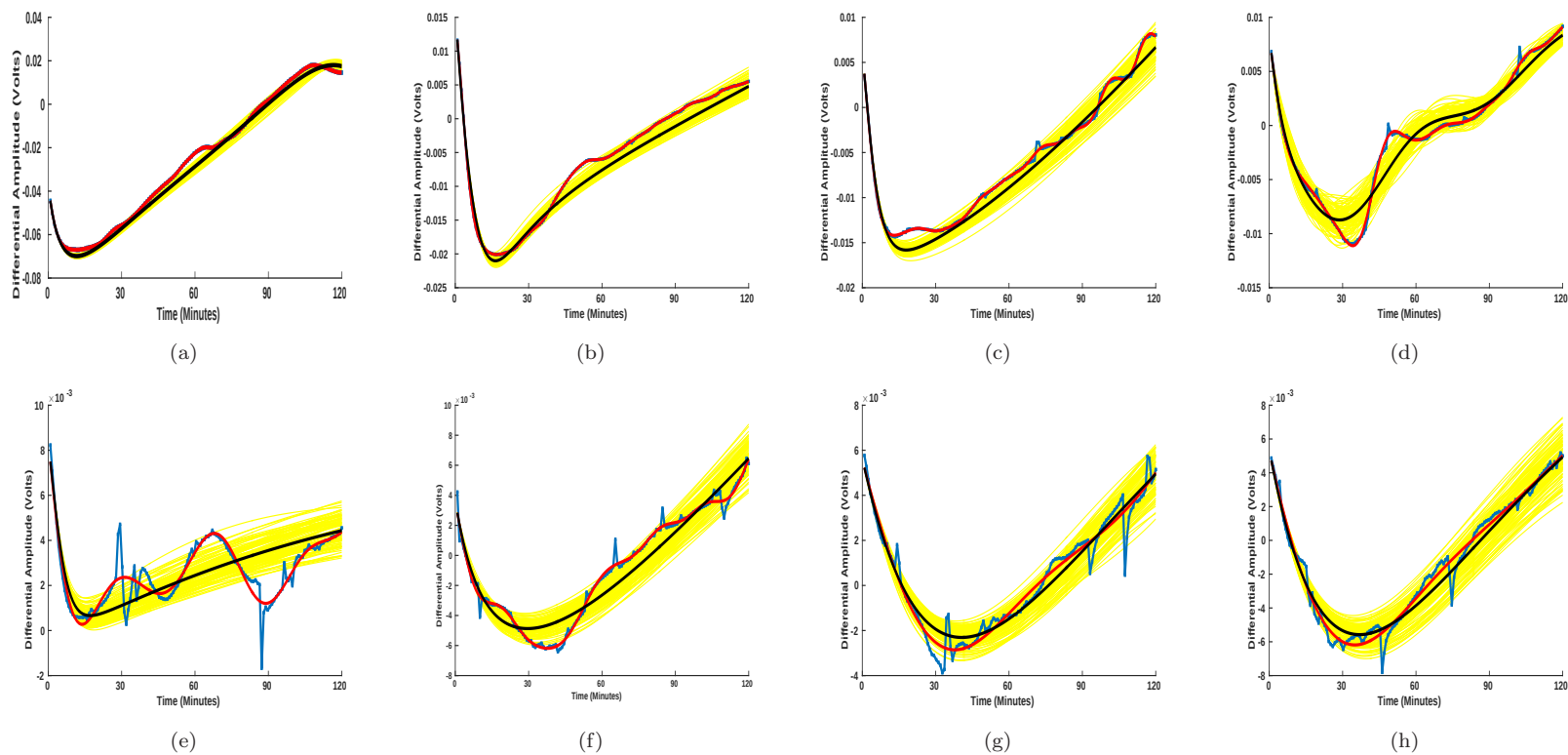


Figure 5.16: Model fitting results after running the GP-ABC-SMC algorithm based on experiments on the cucumber plant, that is batch 2. Notice how the reconstructed curves are more spread around the data in the plots (e-h) indicating a regime change or shrinkage of the electrical responses.

Ligand gated model results

We also ran GP-ABC-SMC for the ligand gated model (see equation 5.8). For this model we have constructed the priors using point estimates following the methods presented in section 5.4.2. We have retained all other algorithmic settings that are used for the voltage gated variant.

The model predictions and fit for the most of the episodes from both the batches (except episode 2 from batch 1 and episodes 1 and 20 from batch 2) are presented in Figure 5.17 & 5.18. It is hard to differentiate the results between the two models, both in terms of fit and predictive uncertainties. However, just like the voltage gated model, we can clearly see the reflection of adaptation to stimulus in the predictive distributions while moving from episode 4 (Figure 5.18(c)) to 5 (5.18(d)) in batch 2. Interestingly, for episode 4 we can see changes in the predictive uncertainty at the points of inflections around the 30-th and after the 90-th minute (see Figure 5.18(c)). Moreover, for both these episodes we can notice a slight improvement in the model fit. However, to choose between these two models in their ability to describe the data succinctly, we need to carry out Bayesian model selection.

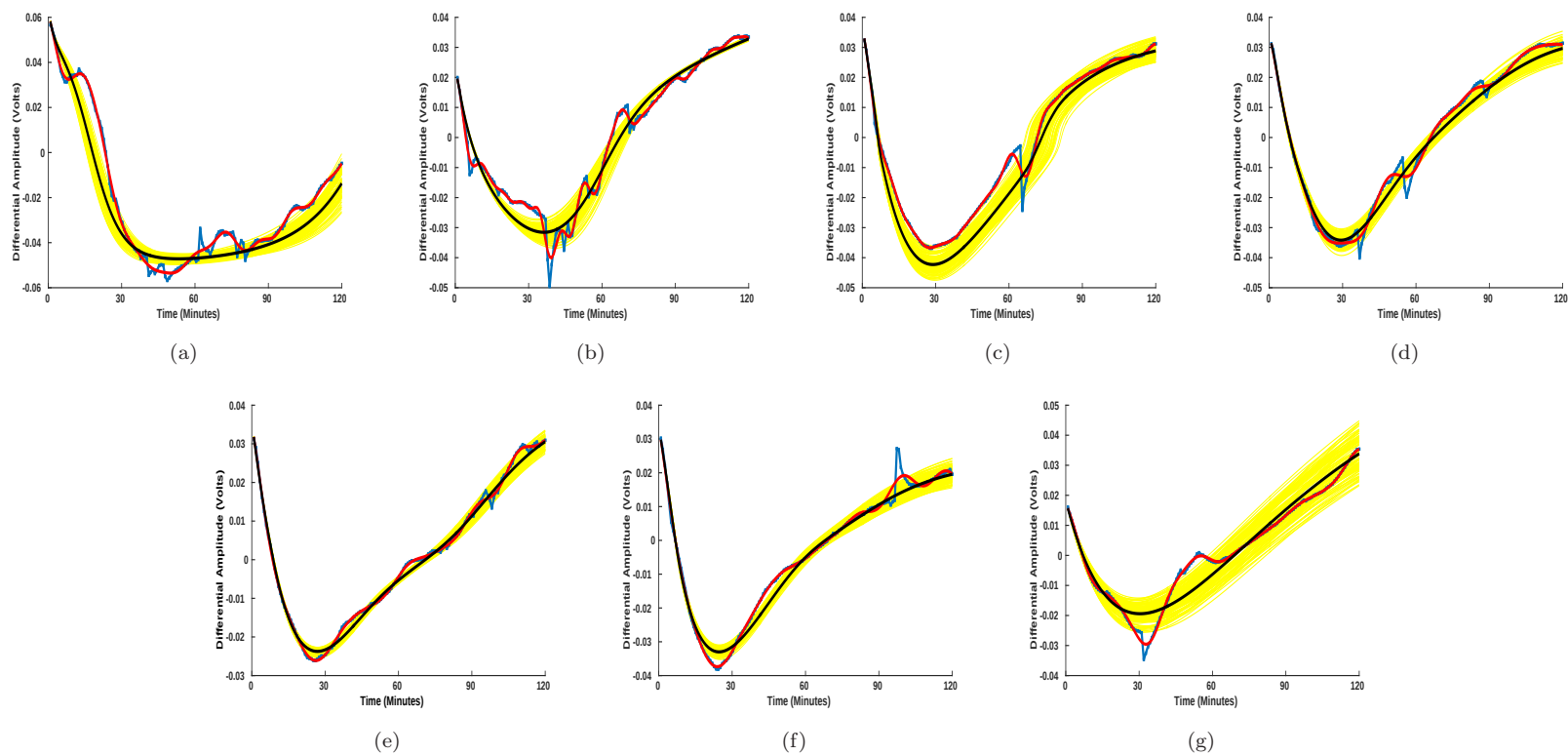


Figure 5.17: Model fitting results for the ligand gated model based on experiments on the tomato plant, that is batch 1.

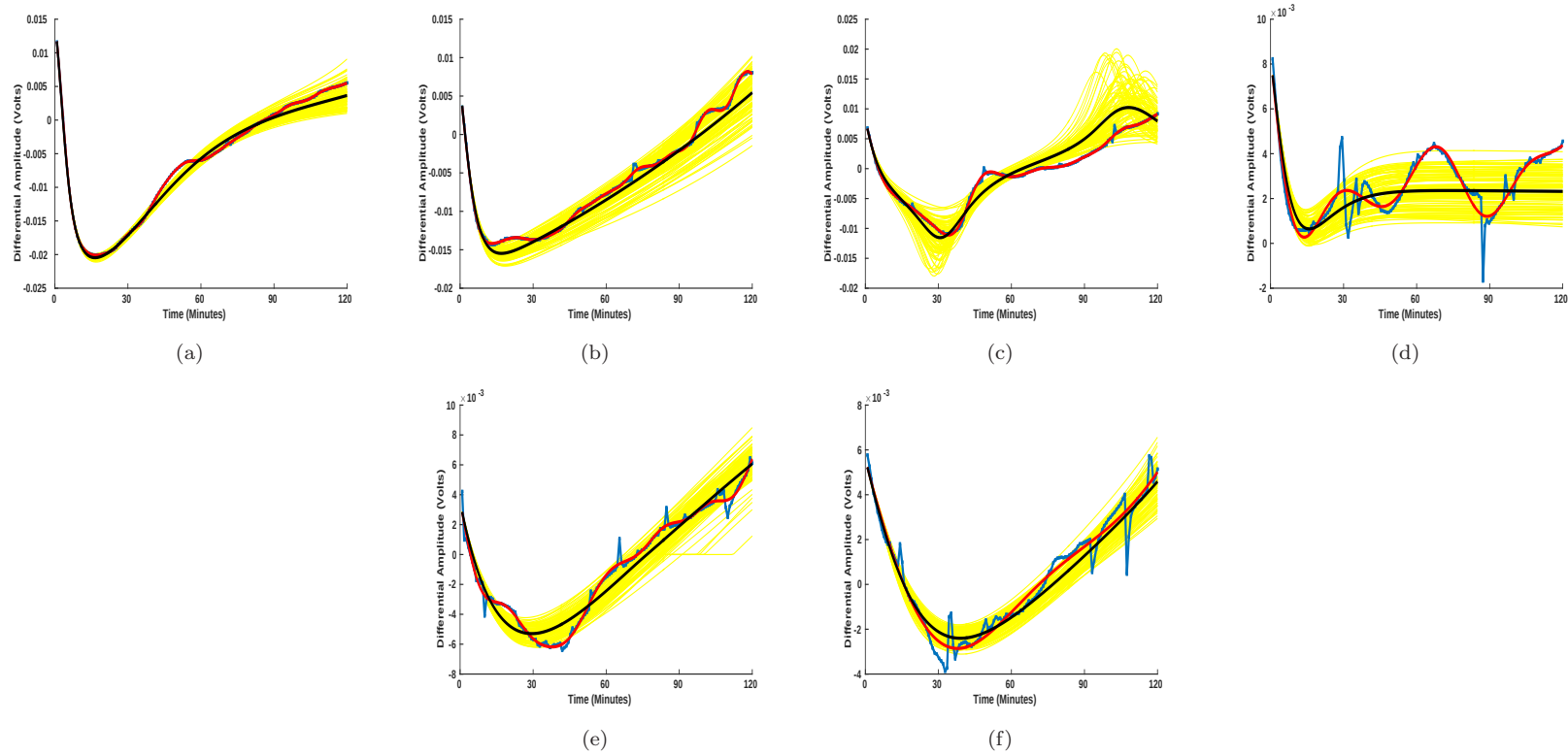


Figure 5.18: Model fitting results for the ligand gated model based on experiments on the cucumber plant, that is batch 2.

Table 5.1: Total number of particles assigned to each model as obtained after running the GP-ABC-SMC for model selection.

Episodes	Voltage gated	Ligand gated	Bayes factor
1 tomato	20	80	4
7 tomato	72	28	2.5714
2 cucumber	73	27	2.7037
3 cucumber	67	33	2.0303
7 cucumber	27	73	2.7037

5.5 Choosing the best model: ABC model selection

Having established two models, the voltage gated and the ligand gated, it is important to select or rank among them. We have thus used the ABC model selection procedure as mentioned in Chapter 2 in conjunction with the GP-ABC-SMC algorithm for this purpose. Furthermore, we have selected 2 episodes, number 1 and 7, from batch 1 (see Figure 5.15(a) & 5.15(g) respectively for tomato plant) and 3 episodes, number 2 ,3 and 7 from batch 2 (see Figure 5.16(b), 5.16(c) and 5.16(g) respectively for cucumber plant). Although we could have chosen all the episodes from both batches for this purpose we specifically selected the aforementioned episodes as they show some specific traits. For example, except episode 1 from batch 1 all other chosen episodes show equally good fit for both models and it is thus difficult to rank the models by simple visual inspection of these time courses. Episode 1 from batch 1 is however used here for a sanity check of the ABC-SMC model selection algorithm as for this episode the ligand gated model fit appears more convincing. Furthermore, episode 7 from batch 2 represents a shrunken response after the plant's adaptation to stimuli. Hence we believe that a Bayesian ranking of the two models for these specific episodes is essential in uncovering the underlying dynamics of the calcium channel gating. We have used the same prior distributions as before, for the respective models. Rest of the algorithmic settings are kept the same. The model selection results are furnished in Table 5.1 which essentially furnishes the number of model indicators (out of 100), from the final step of the SMC, that remain assigned to each model for each corresponding episode. The winning model in each case has higher number of indicators assigned to it. The difficulty of choosing a better model is apparent in the estimated Bayes factors. Except the first episode from batch 1, where clearly the ligand gated model produces a better fit, none of the other episodes generate a Bayes factor greater than 3. Thus from Table 5.1 it is evident that statistically both the models are able to describe the data equally well for all the chosen episodes (except episode 1 where we already know that the ligand gated model provides a better fit). Figure 5.19 illustrates the intermediate SMC stages and their corresponding particle assignments for episode 1 from batch 1 (tomato plant). Although statistically both the models are equally good descriptions of the underlying dynamics, we will use the voltage gated variant for further studies as it has more model

indicators associated with it for 3 out of 5 episodes. However, our choice does not deem this model more suitable by any means.

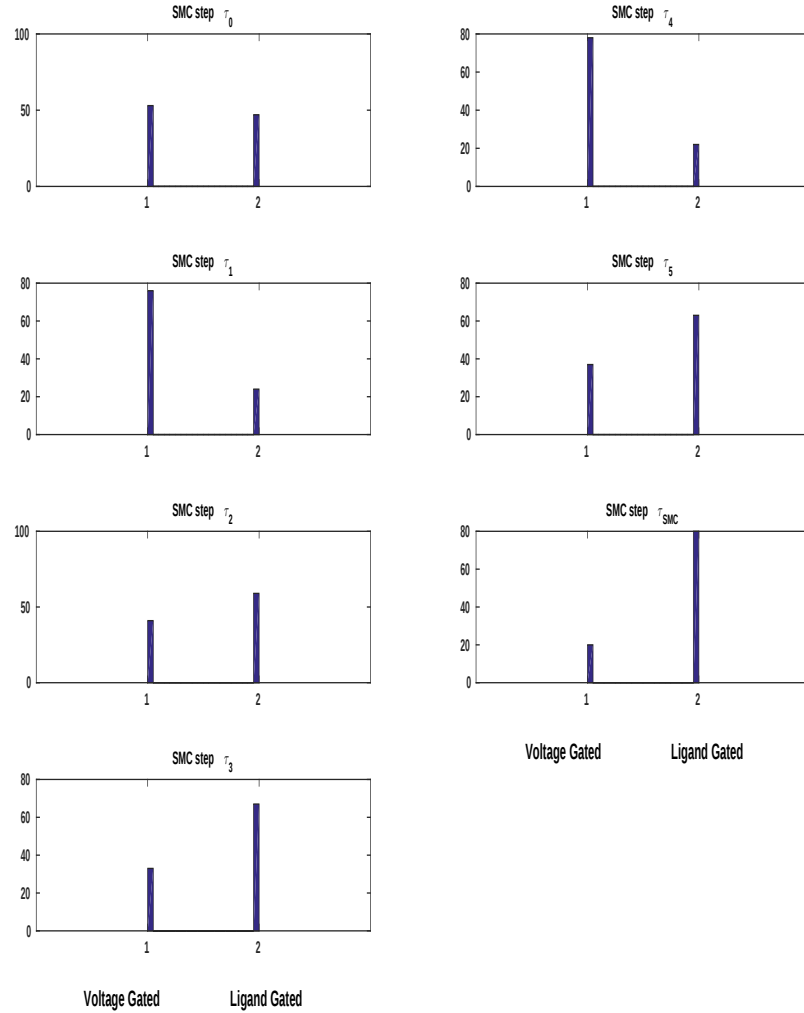


Figure 5.19: The number of particles assigned to each model, the voltage and the ligand gated variant, are plotted as bar graphs. We have plotted these graphs for all the SMC time steps for episode 1 from batch 1 (tomato plant, see Figure 5.15(a)).

5.6 Posterior sensitivity analysis

Apart from elucidating uncertainties associated with model predictions the posterior distributions can be used to find out additional regularities in the parameter space. Presence or absence of such regularities can indicate the extent to which the model (through its parameters) convey information about the stimulus-response episodes. To this end we have used the collection of particles representing the posteriors to carry

out sensitivity analysis. Sensitivity analysis gives us the opportunity to reflect upon the effect that individual as well as a collection of parameters, in our model, have on the net response. Moreover using the output of the inference mechanism, the posterior distribution, for this purpose highlights the crucial connection between the output and the parameters of the model in lieu of observed data. Before presenting the results of this analysis let us introduce the methodology behind using the ABC-SMC output in terms of particles from the final population for carrying out sensitivity analysis. Following [Toni et al. \(2009\)](#) if we carry out PCA on the particles from the final SMC population, collected in a $\mathbb{R}^{N \times D}$ matrix where N is the number of particles and D is the number of parameters, then those principal components which explain most of the variance in this particle collection represent the linear combination of “sloppiest” ([Gutenkunst et al., 2007](#); [Toni et al., 2009](#)) or the least sensitive parameters. Variance of the joint posterior here is accentuated by the principal components predicated upon the notion that lesser sensitivity leads to higher variance in the posterior. Furthermore, the contribution of each parameter in the linear combinations that are the principal components indicates the sensitivity of the parameters in combination. With this idea we ran PCA on the particles obtained after running GP-ABC-SMC, on the voltage gated model, on two consecutive episodes from each batch that have similar time courses. The inference results for chosen episodes from batch 1 (tomato plant) are already shown in Figure [5.15\(f\)](#) & [5.15\(g\)](#) and from batch 2 (cucumber plant) in Figure [5.16\(g\)](#) & [5.16\(h\)](#). The first (component 1) explaining the most and the last (component 15) explaining the least amount of variance of the joint posterior distribution of the parameters are shown in Figure [5.20](#) & [5.21](#). These are obtained from some chosen episodes from batch 1 and 2 respectively. We have also plotted the amount (%) of variance explained by each of the 15 principal components. For the episodes from batch 1, that is tomato plant, we see the stiffest component (component 15) is dominated by parameter k_{22} , the rate of $\delta \mathbf{I}_{ca}$ in the differential equation for the differential membrane voltage $\delta \mathbf{V}_m$, whereas the sloppiest component (component 1) is dominated by separate parameters, k_{ca} (plot on left) and k_B (plot on right), both found in the differential equation for $\delta \mathbf{I}_{ca}$. For the second batch we find almost the identical parameter combination forming the stiffest component, with the dominant ones being μ , k_{11} , k_d . Since μ quantifies the rate of the stimulus current the presence of this parameter asserts our modelling assumption about the crucial triggering role that the stimulus current plays. The appearance of k_{22} and k_d among the most sensitive parameters reinforces our hypothesis about the actuation of the calcium current being related to change in membrane voltage since these parameters control the voltage and calcium current interaction. The sloppiest component for both the episodes is dominated by the parameters k_3 and $\delta \mathbf{I}_{ion}(t = 0)$.

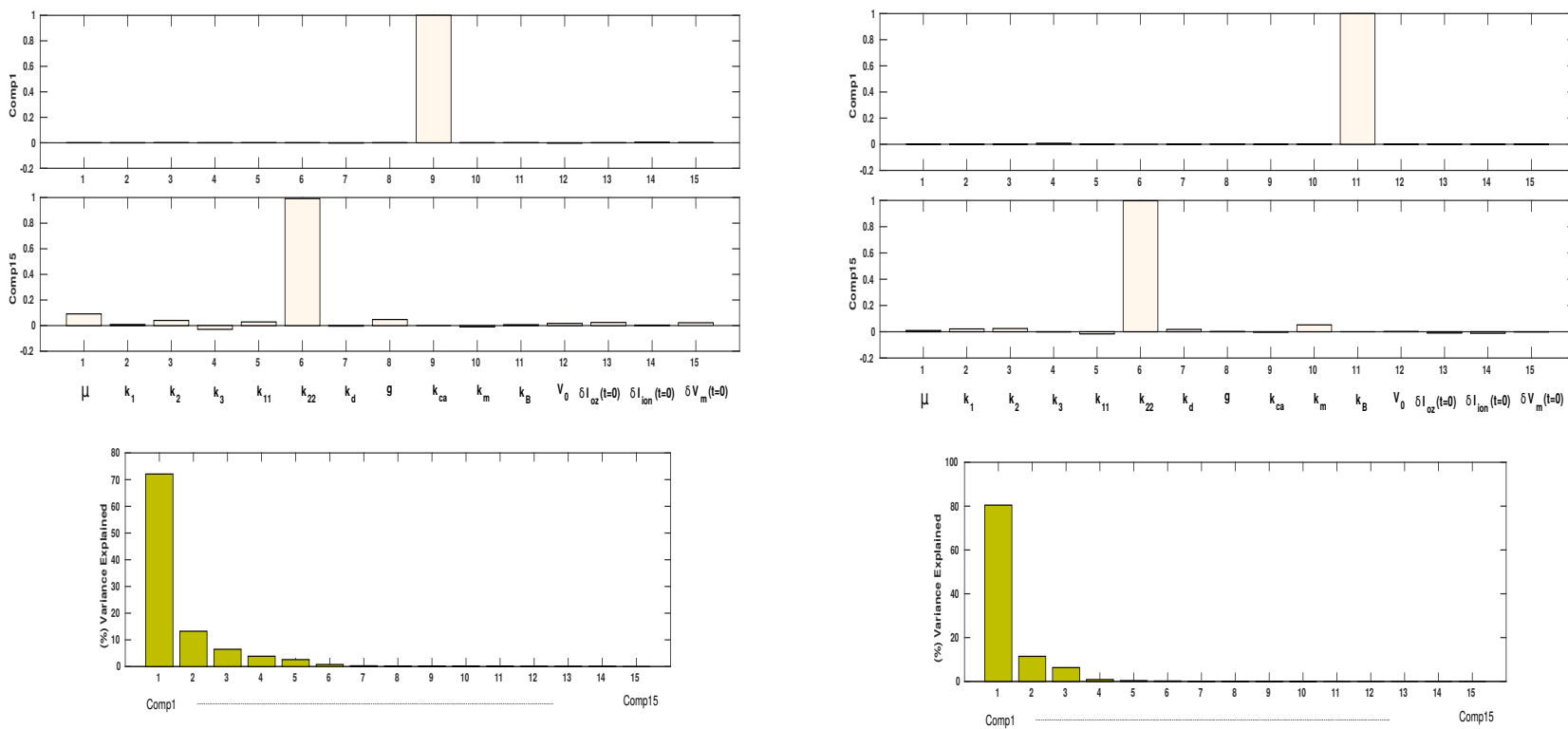


Figure 5.20: Results of the sensitivity analysis are plotted through the principal components and their associated explained variances. The loadings, as the linear parameter combinations, of the 1st (comp 1) and the last (comp 2) principal components are plotted as bar charts in (a) and (b). The height in both direction of the bars indicate the contributions of each of the parameters for both (1st and the last) the components. We have also plotted the (%) of variance explained by all the components in decreasing order.

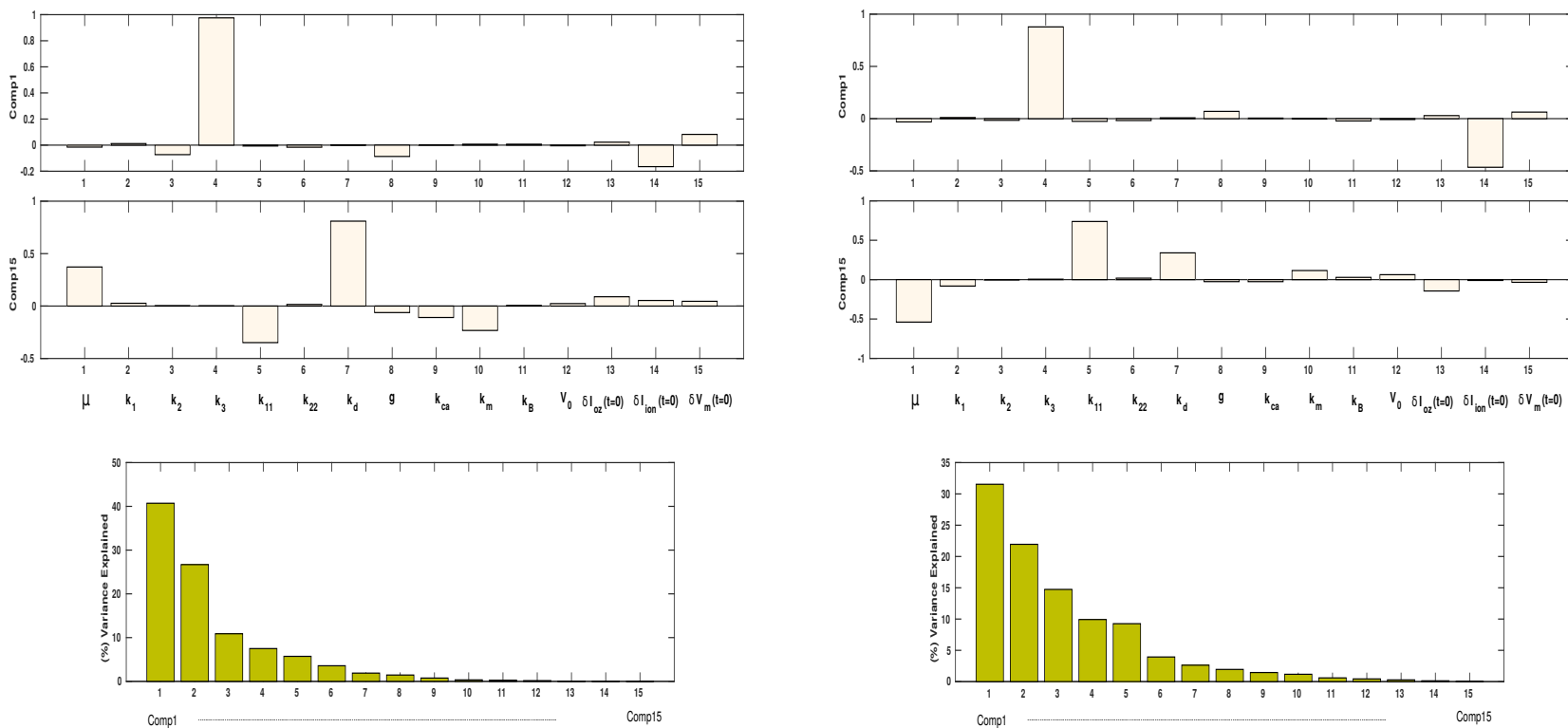


Figure 5.21: Results of the sensitivity analysis on episodes from batch 2.

At this point we like to point out that the results of the sensitivity analysis cannot be trivially generalized across a host of different episodes within the same batch and species. This is due to the massive variability in the time series associated with the episodes. However, from responses with similar time courses the output of the GP-ABC-SMC can be used to unearth the commonalities between the parameters that have the largest effect on the response itself, as furnished above.

5.7 Population approach

The ability of the posterior distributions in summarizing the uncertainty with individual episodes have been established in the previous sections. However, the nature of the variability among the episodes makes it hard to draw a generalized understanding about uncertainty, especially concerning model predictions, while shifting our focus from individual episodes to the batch as a whole. One possible step towards gaining such a holistic perspective leads us towards a population approach. By the term population we mean a population of models such as ours. These models differ from each other through a difference in parameter values. Thus, while one model with its specific parameters can be taken as a representation of the calcium current $\delta \mathbf{I}_{\text{ca}}(t)$ for a single episode, a collection of models then becomes a representation of a batch of episodes. Moreover, drawing these parameter values from the posterior distributions learnt using GP-ABC-SMC will provide the desired holistic representation of uncertainty.

In order to summarize the posterior distributions, considering all the episodes from a certain batch, firstly we collect the particles $\{\phi_{SMC}^{(i)}\}_{i=1,\dots,N}$ from the final population of GP-ABC-SMC for each episode of that batch in a matrix $\Phi \in \mathbb{R}^{MN \times D}$ where M is the number of episodes, $N = 100$ is the number of particles and $D = 15$ is the number of parameters. In this case we are using the voltage gated model. We build the matrix Φ from episodes of batch 1 and then fit a mixture of Gaussian densities to this collection using the expectation maximization algorithm (Bishop, 2006). The fitted mixture of Gaussian densities now represent a distribution on all the particles where each component of the mixture approximates episode specific posterior densities of parameters. From this distribution we can now draw parameter values and simulate a variety of calcium current trajectories. Furthermore, we accept those trajectories of $\delta \mathbf{I}_{\text{ca}}(t)$ which satisfy the following criteria:

$$\left\| \left(\frac{1}{M} \sum_{j=1}^M \widehat{\delta \mathbf{I}_{\text{ca}}}(t) \right) - \delta \mathbf{I}_{\text{ca}}(t) \right\|^2 \leq Tol, \quad (5.15)$$

where $\widehat{\delta \mathbf{I}_{\text{ca}}}(t)$ is the smoothed (by GP regression) calcium current, $M = 8$ is the number of episodes (from batch 1 in this case) and Tol is the threshold of the Euclidean norm between the average of the smoothed currents and the simulated calcium current $\delta \mathbf{I}_{\text{ca}}(t)$.

With the value of Tol we can control how much the simulated calcium currents can deviate from the average behaviour in light of the data. In Figure 5.22 we have plotted 500 such simulated episodes along with the true episodes from batch 1 (tomato) plant.

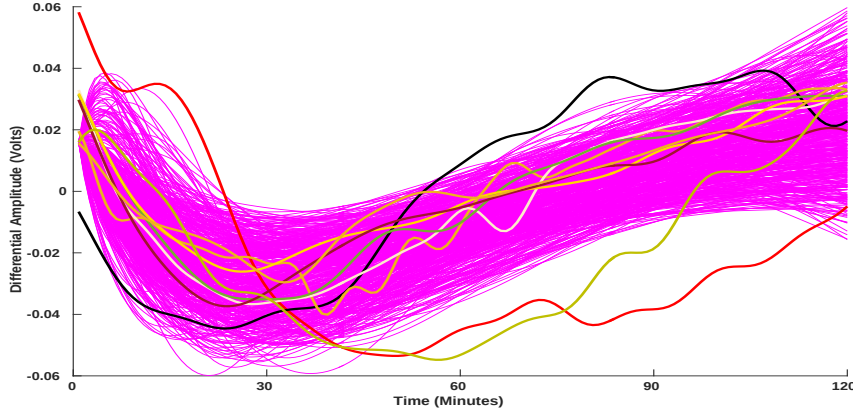


Figure 5.22: The black thick curve representing the average of all the episodes and most of the coloured thick curves representing the smoothed (through GP regression) time courses of $\widehat{\delta I_{Ca}}(t)$ falls within the ranges of the population of simulated calcium currents plotted with thin magenta curves.

It is clearly evident from Figure 5.22 that the population of the simulated current trajectories are able to capture a fair amount of variability and can convey the desired uncertainty for a batch of episodes.

5.8 Conclusion

In this chapter we proposed a phenomenological model to quantify the electrical responses in higher plants subjected to ozone exposure. Furthermore, we have argued that the electrical signals recorded from the surface of the plants are manifestation of plant-wide calcium signals triggered due to the intrusion of ozone in plants. We built mathematical models to describe these calcium signals as calcium ionic currents within the guard cells and then extrapolated our assumptions to tissue level models. We used the GP-ABC-SMC algorithm to fit and compare such models to experimental recordings of ozone driven electrical signals— which we term episodes. Although our models are based on simplistic assumptions of the ionic interactions within the cell and their dynamics at the tissue (the phloem) level, the inference algorithm is able to shed light on some of the interesting aspects of this stimulus-response phenomenon.

Apart from model-fitting and propagating posterior parameter uncertainties to model predictions, the inference algorithm is applied in order to rank two competing model assumptions —(1) a voltage gated model where the calcium current is generated through

a change in cellular membrane voltage and (2) a ligand gated model where the efflux of Ca^{2+} ions leads to the generation of the same. GP-ABC-SMC, the inference algorithm developed in previous chapters, found both the models equally good in explaining test episodes. We also carried out sensitivity analysis using the output of the algorithm for the voltage gated model. From the sensitivity analysis we found out that the most sensitive parameters include k_{11} , k_{22} , k_d and μ . The parameter μ which is the rate of ozone current (the input in our model) is expected to have high sensitivity. However, k_{22} and k_d control the rates of interaction of the calcium current and the membrane voltage, both of which appear in the differential equation for the membrane voltage. Such a result can be interpreted in the support of an underlying voltage gated calcium channel dynamics.

We have also noticed that GP-ABC-SMC captures the regime change, the shrinkage of the electrical response, in the latter episodes from the batch of experiments involving cucumber plants. We attribute this regime change to a gradual adaptation to the repetitive stimulation of the cucumber plant. However, our conclusions here are based on the fact that we have chosen only two batches of experiments which include consecutive episodes with discernible responses. Other batches have only few episodes that show clear electrical responses. For this lack of repeatability strong conclusion can only be drawn in the light of a more concrete evidence within an elaborate dataset. Amidst such repeatability issues we have gone on to fit the voltage gated model to 55 episodes in total including the ones shown in this chapter. In most cases we have noticed similar level of fit to the experimental data.

The dynamical variables of the model other than the calcium current do not show traces of any underlying regularities through their dynamics. Thus, to truly validate our central hypothesis and the model involving a long distance calcium wave (current), we need to carry out further experiments of the same ilk while using some form of calcium channel blocking. Absence of the electrical responses thereafter could be taken forward as a strong evidence in favour of our hypothesis.

Chapter 6

Summary and main contributions

6.1 Summary

In this thesis, we have presented the various stages in the development of a novel Bayesian inference algorithm for deterministic dynamical systems that utilizes some of the recent developments in this field, such as the ABC-SMC algorithm and Gaussian processes for smoothing trajectories of an ODE system. The proposed algorithm was applied to solve an inverse problem in plant electrophysiology, incorporating a novel mathematical model and experimental results. The thesis primarily reports two major achievements— i) Modifying the ABC-SMC algorithm to achieve a significant speedup, ii) Postulating a novel mathematical model that describes the generation of plant-wide electrical signal in response to ozone as a stimulus. In the following paragraphs we will briefly summarize individual chapters of this thesis and highlight the main achievements.

6.1.1 Chapter 1

We started by presenting the task of uncertainty quantification in dynamical systems models described as non-linear ordinary differential equations as a statistical problem. A task that needs to be addressed within the Bayesian statistical framework. We then stated the main motivation behind this work: designing a fast Bayesian inference algorithm that can handle the above mentioned task of uncertainty quantification.

We also introduced our secondary aim of the thesis—Applying the GP based ABC-SMC algorithm to solve an inverse problem in plant electrophysiology. We introduced this problem by providing a brief description of various electrical signals found to exist in plants, and the biophysical motivation behind modelling such signals.

We finish this chapter by furnishing a brief summary of the content presented in the remaining chapters of this thesis.

6.1.2 Chapter 2

In this chapter we pursue a literature review of ABC methods in the context of parameter estimation and model selection in ordinary differential equations. We explain how the ABC-SMC algorithm can be applied for these two tasks, pointing out the major advantages and limitations of this inference method. We also mentioned, where necessary, the connections between ABC-SMC and other major achievements (as well as limitations) of the burgeoning field of ABC.

6.1.3 Chapter 3

Having established our main thesis goal of resolving the computational burden of ABC-SMC, resulting from explicit integration in the previous chapters, we start by introducing Gaussian process regression and show how GPs can be used to model empirical velocity field of an ODE. We then apply this machinery to transform the ABC distance function from state to derivative space, bypassing the requirement of explicit integration and thus the computational bottleneck. We also highlighted the major differences between our approach and previous gradient matching techniques that employ Gaussian processes. Especially, we pointed out that our GP based ABC-SMC does not require the sampling of states and other nuisance parameters as in methods that use GP with exact MCMC. Furthermore, it was also shown how operating in the derivative space frees us from learning the initial values of the states as additional parameters. However, the proposed method achieves these through a compromise. No direct feedback exist between GP regression and sampling of posterior ODE parameters in the proposed method, resulting in a two state approach vulnerable to extremely noisy measurements. Finally, we concluded this chapter by stating other potential advantages that can be gained through speeding up ABC-SMC such as carrying out many trials runs, with variety of algorithmic settings, of the GP-ABC-SMC algorithm to obtain the most accurate posterior estimates. This was previously unattainable, due to a huge computational burden, while using the ABC-SMC algorithm based on explicit integration.

6.1.4 Chapter 4

We applied the GP-ABC-SMC algorithm, proposed in the previous chapter, to estimate parameters of several toy ODE and DDE model systems. For the chosen problems, GP-ABC-SMC obtained similar estimates to that of ABC-SMC based on explicit integration. Although we attempted to compare GP-ABC-SMC to other MCMC algorithms

based on GP, we found that some of these methods were tested with informative priors. Thus, it is difficult to draw conclusion in favour of GP-ABC-SMC or any of these other GP based algorithms. Interestingly, we found that ABC-SMC suffers from population degeneracy when populations are sampled from such narrow priors. However, GP-ABC-SMC performed well on two models, Hes 1 and the signal transduction cascade, that have been used previously for benchmarking other gradient matching algorithms proposed in [Calderhead et al. \(2008\)](#) and [Dondelinger et al. \(2013\)](#); [Wang and Barber \(2014\)](#) respectively. For all the models on which we tested GP-ABC-SMC the process of parameter estimation finished within 1 minute consistently. For the Hes 1 model such performance is better than the state-of-the-art, since GP-ABC-SMC does not require the knowledge and estimation of initial history function of the states.

Utilizing the speed benefits of GP-ABC-SMC we ran multiple trials of the algorithm to test the performance of various perturbation kernels as well as tolerance schedules. Furthermore, we presented a variability analysis of the estimates. To the best of our knowledge such variability analysis of estimation results is seldom done in the context of Bayesian inference in dynamical systems, including ABC-SMC.

Retaining the most important aspect of ABC-SMC, that is providing a holistic platform for uncertainty quantification, we applied the GP based ABC-SMC model selection algorithm for ranking four candidate epidemic models of the SIR family. These models have been previously used for benchmarking the ABC-SMC model selection algorithm using a common-cold dataset. In this experiment real data was used and the GP-ABC-SMC algorithm had to tackle the issue of hidden state. GP-ABC-SMC arrived at similar results as found previously at a significantly reduced execution time.

Through the various experiments presented in this chapter, we showcased all the benefits, as was indicated in the previous chapter, that can be achieved using a gradient matching approach within ABC-SMC.

6.1.5 Chapter 5

We tested various aspects of the GP-ABC-SMC algorithm on benchmarking model systems. Although these models present formidable challenges, the success and also the applicability of the proposed algorithm can be tested through solving new problems. In such novel applications we will be forced to starve the algorithm of any a priori information about the model, making the inference algorithm work harder to achieve desirable estimates. In this chapter we applied the GP-ABC-SMC to fit phenomenological models of electrical responses in plants to data, which is a novel inverse problem of its kind. In this case we developed novel mathematical models that describe plant-wide electrical responses to exposure of ozone as a stimulus. We modelled these electrical responses to be the manifestation of Ca^{2+} ion efflux actuated by the ozone exposure. Furthermore,

we separately modelled the mechanism behind this efflux as being voltage as well as ligand gated. We found, upon applying GP-ABC-SMC, that the voltage gated variant fit well to the experimental data and is chosen as the more suitable variant in terms of Bayesian model selection criteria (relying on the Bayes factor). This fact was supported by the results of sensitivity analysis done using the posterior (as particle population) estimates of the model parameters.

Our main contributions of this chapter are the proposed mathematical models which described an important biophysical phenomenon in plants: electrical responses in plants subjected to a harmful environmental pollutant such as ozone. Apart from modelling, we also showcased how the GP-ABC-SMC can be successfully used to quantify uncertainty related to parameters, structures and predictions of a novel model, using a novel experimental dataset. We believe the work reported in this chapter will pave the path for further explorations of the stimulus-response phenomenon in plants, using a Bayesian statistical framework and hopefully the GP-ABC-SMC algorithm.

6.2 Future work

In this section we will outline some of the future research directions for extending the GP-ABC-SMC and also the stimulus-response mathematical models that we have proposed in the previous chapter.

6.2.1 Feedback between GP and ODE parameters

One way of guiding the GP model to reflect the characteristics of the ODE model is by sampling the GP parameters Φ within the SMC steps and constructing the empirical velocity using these sampled parameters and the observation Y^d . So, in this procedure we start by sampling the GP parameters $\Phi \sim \pi(\Phi)$ as well as the ODE parameters $\theta \sim \pi(\theta)$. Now, we can evaluate the smoothed state $\widehat{\mathbf{X}}(t) = \mathbb{E}[\mathbf{X}(t^*)|Y^d]$ using equation 3.31, where we construct the covariance matrices $K(t^L, t^*; \Phi)$ using the sampled GP parameters Φ . We can then evaluate, using equation 3.32, the empirical velocity $\mathbf{V}(t) = \mathbb{E}[\frac{d}{dt}\mathbf{X}]$ using the smoothed state $\widehat{\mathbf{X}}(t)$. Now we construct the same distance function $\Delta(\mathbf{V}^d(t), \mathbf{f}(\widehat{\mathbf{X}}(t), \theta)) \leq \epsilon_\tau$ to implement ABC. However, in this case a sample of Φ will only be accepted if the tolerance is satisfied depending on the sample of θ . Note that this is still a two step approach as we are evaluating the smoothed state and its derivative in two steps, but now the acceptance of Φ and θ are inter-related. Thus now the acceptance (not the sampling) of Φ is dependent on the ODE dynamics. The implementation of this modified sampling scheme present some challenges. The first challenge is related to the design of suitable prior distribution $\pi(\Phi)$. Following Dondelinger et al. (2013), we can initially use GP regression and then use the point estimates to construct the

support of $\pi(\Phi)$. However, we need to carefully balance between a narrow prior that can cause population degeneracy and a wider one that can lead to lower acceptance. Lower acceptance on the other hand will result in a number of computationally expensive inversion of the GP and derivative GP covariance. Thorough exploration using many model systems needs to be carried out to find the best tradeoff. Carrying out such exploration is a task that we like to pursue in the future.

6.2.2 Geometric tolerance schedule

The choice of a suitable tolerance schedule in ABC-SMC is perhaps the most critical algorithmic setting in the inference process. This schedule essentially determines how the algorithm navigates the parameter space. A carefully designed tolerance schedule ensures that the filtering steps explore the parameter space sufficiently without getting stuck in areas of low likelihood. The α -th quantile method does not require prior empirical knowledge about the model itself but the very choice of α is complex and has an impact on the tradeoff between exploration and computational efficiency. Many different values of α shall be tested systematically to find out those which ensure consistent posterior estimates. This entire task at the least is computationally prohibitive considering ABC-SMC run on the state space (using numerical integration).

Using the speed benefits of the GP-ABC-SMC platform multiple tolerance schedules can be tested for a specific problem for best inferential results within a practical time frame. However, we believe a fair amount of engineering is still required in this regard, for example we need to choose multiple start $\epsilon_{\tau=0}$ and end $\epsilon_{\tau=S_{MC}}$ tolerance values, the quantile adaptation rate α as well as the number S_{MC} of algorithmic steps.

In order to protect the user from having to decide these settings (as well as multiple testing of multiple combinations) without affecting the inference we consider using the gradient and curvature of the distance function surface to automatically construct a suitable tolerance schedule. For non-linear least squares problems the Levenberg-Marquardt algorithm use the knowledge of the curvature of error surface effectively to navigate towards a local minima. We can morph such an optimisation algorithm with the ABC-SMC to automatically construct the best possible tolerance schedule.

The non-linear least squares problem is concerned with finding the optimal parameter vector $\theta \in \mathbb{R}^d$ that minimizes the following cost function:

$$r(\theta) = \frac{1}{2} \sum_{i=1}^L \left(y_i^d - f(x_i, \theta) \right)^2, \quad (6.1)$$

between the observed data $y_i^d \in \mathbb{R}$ and the non-linear function $f(x_i, \theta)$ evaluated at the input $x_i \in \mathbb{R}, i = 1, \dots, L$. The Hessian matrix $\mathbf{H}$ with elements $H(j, l) = \frac{\partial^2 r}{\partial \theta_j \partial \theta_l}$

being the second order partial derivatives of the cost function is used to obtain the curvature information. Using this curvature and the gradient ∇r of the cost function, the Levenberg-Marquardt algorithm updates the parameter vector at each iteration τ as:

$$\theta_{\tau+1} = \theta_{\tau} - (\mathbf{H} + \lambda \mathcal{D})^{-1} \nabla r, \quad (6.2)$$

where $\mathcal{D}$ represents the diagonal elements of $\mathbf{H}$. As λ , a damping factor, gets small the update rule uses curvature information and as λ is bigger then the gradient descent is followed.

In ABC-SMC we choose, before hand, a path on an Euclidean space from a starting point that is the initial tolerance $\epsilon_{\tau=0}$ to the final point which is the final tolerance $\epsilon_{\tau_{SMC}}$. The points on this path, representing each tolerance value ϵ_{τ} , can then be collectively interpreted as the tolerance schedule. For an adaptive schedule the successive points on this path are chosen as the α -th quantile of the previous distances. Alternatively, we can choose the next tolerance level by applying the Hessian to run an implicit Levenberg-Marquardt update over the posterior mode, of the intermediate distributions, calculated from the accepted particles in each SMC step. Thus while the algorithm is at SMC step $\tau - 1$ we approximate the sample mean of the particle population

$$\bar{\theta}_{\tau-1} = 1/N \sum_{i=1}^N \theta_{\tau-1}^{(i)} \quad (6.3)$$

at that step that have satisfied the corresponding tolerance $\epsilon_{\tau-1}$, $\{\theta_{\tau-1}^{(i)}\}_{i=1,\dots,N} = \{\theta_{\tau-1}^{(i)} | \Delta(Y^d, Y^s(\theta_{\tau-1}^{(i)})) \leq \epsilon_{\tau-1}, 1 \leq i \leq N\}$ where N is the population size. We then use the Levenberg algorithm to update this mean (of parameter samples) vector to obtain

$$\bar{\theta}_{\tau} = \bar{\theta}_{\tau-1} - (\mathbf{H} + \lambda \mathcal{D})^{-1} \nabla C(\bar{\theta}_{\tau-1}). \quad (6.4)$$

where $\mathbf{H}$ and $C(\bar{\theta}_{\tau-1})$ are the Hessian and Jacobian of the ODE cost function (see equation C.10 and C.8). The next tolerance in schedule is then adaptively chosen as:

$$\epsilon_{\tau} = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^L \left(Y_k^d(t_i) - f(X_k(t_i, \bar{\theta}_{\tau}), \bar{\theta}_{\tau}) \right)^2, \quad (6.5)$$

where X_k is the k -th state of a K state coupled ODE system. The update given in equation 6.4 is repeated until we find a tolerance lower in value than the previous one. Until that occurs the λ is successively increased by a factor, say 10. Once a new tolerance is found the optimisation is stopped and the rest of the ABC steps are carried out. Note that the choice of α is now replaced by the choice of λ . It is thus interesting to see whether a generic setting of λ can be used for a variety of problems. Furthermore, note that the optimisation is started at different points on the distance surface during different SMC steps. This can be interpreted as starting the optimisation with many initial parameter values, thus being more likely to recover a final ϵ close

to the global minima. The quantile adaptation method on the other hand does not ensure the same. Interestingly, applying the update method as shown in equation 6.4 on the Fisher information matrix (see Appendix C) rather than the Hessian transforms the tolerance adaptation to track the curvature of the log-likelihood surface. Since the Euclidean distance function and the log-likelihood for a normally distributed noise model are directly related, this transformation essentially introduces data dependency (through the covariance matrix associated with the FIM) on the curvature estimation.

To test the fruitfulness of this tolerance adaptation scheme we need to run many experiments comparing the quantile and the optimisation based tolerance schedules. Also, we need to find a way to evaluate the Hessian and the gradient of the distances in the derivative space, when using this tolerance method within GP-ABC-SMC. We would like to take up these tasks in the future.

6.2.3 Partially observed systems

We have so far handled partially observed systems, with one or many hidden states using explicit integration. However, introducing integration for hidden states defeats the purpose of this and any other gradient matching algorithm. Considering the state vector $\mathbf{X} = (X_o, X_h)$, ideally we want to use the GP covariance $K(t, t')$ and the ODE right hand side $\mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta})$ to directly estimate $X_h(t)$ given the observations for $X_o(t)$. The multi-task Gaussian process (MTGP) model (Bonilla et al., 2007; Durichen et al., 2015) can be employed for this task. Modelling a coupled ODE system using the MTGP, we can construct covariance functions on outputs/tasks (states) along with inputs (time points). For a K state ODE system, in GP-ABC-SMC we model each of the K states using different GPs with different covariances. However, using MTGP we can model the correlations across state dimension using a covariance $K(k, k)$ and the time input using a covariance $K(t, t')$. The full joint covariance of the MTGP is then given by a Kronecker product of these two covariances. The resulting scenario is depicted in Figure 6.1. Thus, using this joint covariance we can model inter-state dependencies that could potentially lead us towards estimating the hidden states directly. So far we have not found a systematic way to achieve this, but we believe we can exploit the MTGP framework to solve this issue.

Furthermore, for some systems the output is observed as a mixture of the individual states, such as a linear combination. None of the gradient matching algorithms (including GP-ABC-SMC) can be applied on this sort of systems. We need to get rid of these limitations, using the Gaussian process framework, to make GP-ABC-SMC more widely applicable.

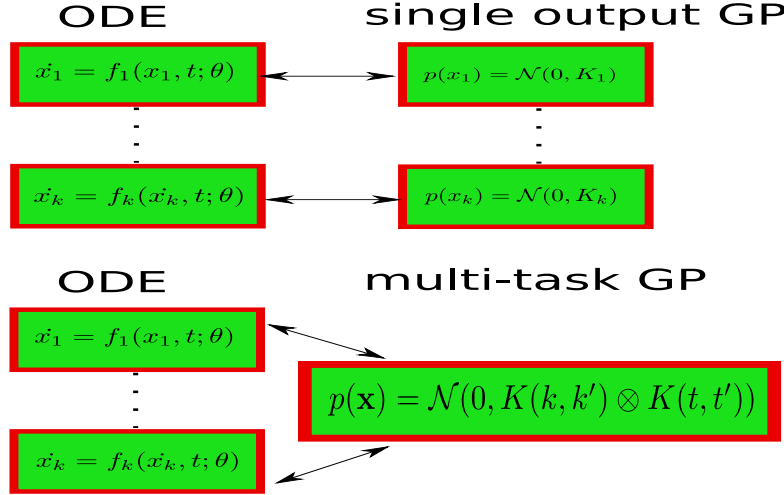


Figure 6.1: The block diagram on top represents the single-output GP modelling framework, wherein separate GP priors are placed on each state of a coupled ODE. In the MTGP framework, a single prior is placed on all the states through the Kronecker product covariance.

6.2.4 Further work regarding the modelling of plant electrical responses

Before suggesting any improvements of the stimulus-response model we need to consider more elaborate validation of our model. We have so far used some of the salient features of GP-ABC-SMC to support our hypothesis around the association of a calcium current to the electrical response. However, for a more rigorous validation we need to design experiments to practically test this hypothesis. Experiments involving calcium channel blockers can be applied to this task. But since we are dealing with plant-wide electric responses thus we have to consider the difficulties associated with blocking channels across the length of phloem. However, we believe, our results can potentially motivate plant scientists to design such experiments in the future.

Apart from validating our hypothesis, experimentally, we can extend the mathematical model to describe spatio-temporal properties of electrical responses. The experimental setup that we are using is not adequate for recording spatial behaviour, such as propagation dynamics, of the electrical responses. Thus, we need to conduct experiments employing more number of electrodes to carefully record the propagation of the electrical responses. We can then attempt to solve the associated inverse problem, using a spatio-temporal model, to elucidate several interesting attributes of the dynamics (spatial and temporal) of long distance electrical responses.

Finally, considering this work as a platform, in the future we like to study and obtain a more generalized representation of stimulus-response episodes combining different stimuli. For example extending our basic idea of investigating the effects of harmful environmental pollutants on electrical responses can be extended to include pollutants such as acid rain. In the current experimental framework, experiments have been done

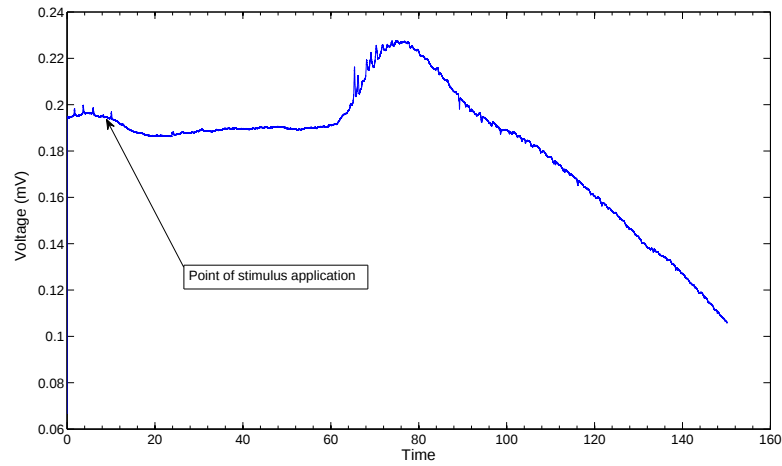


Figure 6.2: Time course of electrical potential obtained after adding 5ml of $0.05M$ H_2SO_4 to the soil.

on plants using H_2SO_4 as a stimulus. In this case H_2SO_4 is applied to the soil and the resulting electrical response is shown in Figure 6.2. Modelling this sort of responses presents new challenges such as: a) modelling the absorption of H_2SO_4 by the plants, b) understanding the mechanism behind the delayed onset of response and c) incorporating repetitive oscillations (bursting) in the model. Since we already have some amount of data for the said stimuli, and the GP-ABC-SMC algorithm for inference, we can start modelling this very interesting phenomenon.

Appendix A

Multivariate Gaussian Identities

A.1 Schur complement

In this appendix we provide details of Schur complement in order to derive multivariate Gaussian distribution following Gallier's work ([Gallier, 2010](#)). Let a $k \times k$ matrix be partitioned as 2×2 block matrix:

$$\begin{bmatrix} \underbrace{\mathbf{A}}_{m \times m} & \underbrace{\mathbf{B}}_{m \times n} \\ \underbrace{\mathbf{C}}_{n \times m} & \underbrace{\mathbf{D}}_{n \times n} \end{bmatrix} \quad (\text{A.1})$$

where A,B,C and D are: $m \times m$, $m \times n$, $n \times m$ and $n \times n$ matrices, respectively. And $k = m + n$. Now if we consider a linear system, where

$$\begin{aligned} Ax + By &= c, \\ Cx + Dy &= d. \end{aligned} \quad (\text{A.2})$$

We can solve the system for y to obtain

$$y = D^{-1}(d - Cx), \quad (\text{A.3})$$

where D is invertible. Substituting y in equation [A.8](#), we get

$$Ax + B(D^{-1}(d - Cx)) = c, \quad (\text{A.4})$$

and

$$(A - BD^{-1}C)x = c - BD^{-1}d. \quad (\text{A.5})$$

By assuming that the matrix $A - BD^{-1}C$ is invertible, the solution becomes:

$$x = (A - BD^{-1}C)^{-1}(c - BD^{-1}d), \quad (\text{A.6})$$

and

$$y = D^{-1}(d - C(A - BD^{-1}C)^{-1}(c - BD^{-1}d)). \quad (\text{A.7})$$

We can rewrite the above equations as

$$\begin{aligned} x &= (A - BD^{-1}C)^{-1}c - (A - BD^{-1}C)^{-1}BD^{-1}d, \\ y &= -D^{-1}C(A - BD^{-1}C)^{-1}c + (D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1})d. \end{aligned} \quad (\text{A.8})$$

And finally the solution for the matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}$ is given as,

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}) \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}. \end{aligned} \quad (\text{A.9})$$

A.2 Multivariate Gaussian

Let x ($x \in \Re$) be an n -dimensional random vector whose each variate has an univariate Gaussian distribution. The probability density function of such a vector is a multivariate Gaussian distribution given as

$$P(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{1/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (\text{A.10})$$

where $\boldsymbol{\mu}$ ($\boldsymbol{\mu} \in \Re$) is the mean vector and $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma} \in \Re^{n \times n}$) is a covariance matrix, which must be positive definite in order to be valid. In order to explain the concept of a multivariate Gaussian vividly, we can consider a bivariate example here by generating a two-dimensional vector $\mathbf{a}$ ($\mathbf{a} \in \Re^2$, $\mathbf{a} = [a_1, a_2]^T$) where

$$\mathbf{a} = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (\text{A.11})$$

The mean and the covariance matrix are given as:

$$\begin{aligned} \boldsymbol{\mu} &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \text{ and} \\ \boldsymbol{\Sigma} &= \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}. \end{aligned} \quad (\text{A.12})$$

$\rho\sigma_1\sigma_2$ is the cross-covariance of the two variates a_1 and a_2 . The correlation between a_1 and a_2 is high with a large correlation parameter ρ , and is zero when the two variates are independent.

We now consider two vectors $\mathbf{a}$ ($\mathbf{a} \in \mathbb{R}$) and $\mathbf{b}$ ($\mathbf{b} \in \mathbb{R}$), which have joint Gaussian distribution:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{B} \end{bmatrix} \right) \quad (\text{A.13})$$

The covariance matrix $\boldsymbol{\Sigma} = [\mathbf{A}\mathbf{C}^T; \mathbf{C}\mathbf{B}]$ is separated into block matrices in equation A.13, in which, $\mathbf{A}$ and $\mathbf{B}$ are corresponding covariance matrices for $\mathbf{a}$ and $\mathbf{b}$, respectively. $\mathbf{C}$ is the cross-covariance matrix between $\mathbf{a}$ and $\mathbf{b}$. With a zero-mean assumption ($\boldsymbol{\mu}_a = 0, \boldsymbol{\mu}_b = 0$), the joint distribution can be written as

$$P(\mathbf{a}, \mathbf{b}) \propto \mathcal{N} \left(-\frac{1}{2} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}^T, \begin{bmatrix} \mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right) \quad (\text{A.14})$$

By the Schur complement, the block matrix is given by

$$\begin{bmatrix} \mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{B}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{C}) & \mathbf{O} \\ \mathbf{O} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{C}^T\mathbf{B}^{-1} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \quad (\text{A.15})$$

The joint distribution in equation A.14 can be written using this as

$$\begin{aligned} P(\mathbf{a}, \mathbf{b}) &\propto \exp \left(\begin{bmatrix} \mathbf{a} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{b} \\ \mathbf{b} \end{bmatrix}^T \begin{bmatrix} (\mathbf{A} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{C}) & \mathbf{O} \\ \mathbf{O} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{a} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{b} \\ \mathbf{b} \end{bmatrix} \right) \\ P(\mathbf{a}, \mathbf{b}) &\propto \exp \left(-\frac{1}{2} (\mathbf{a} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{b})^T (\mathbf{A} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{C})^{-1} (\mathbf{a} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{b}) \right) \exp \left(-\frac{1}{2} \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} \right) \end{aligned} \quad (\text{A.16})$$

The conditional distribution of $\mathbf{a}$ given $\mathbf{b}$ can be obtained as

$$P(\mathbf{a}|\mathbf{b}) = \frac{P(\mathbf{a}, \mathbf{b})}{P(\mathbf{b})} = \mathcal{N}(\mathbf{C}^T\mathbf{B}^{-1}\mathbf{b}, \mathbf{A} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{C}) \quad (\text{A.17})$$

Relaxing the zero-mean assumption, equation above becomes

$$P(\mathbf{a}|\mathbf{b}) = \mathcal{N}(\boldsymbol{\mu}_a + \mathbf{C}^T\mathbf{B}^{-1}(\mathbf{b} - \boldsymbol{\mu}_b), \mathbf{A} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{C}) \quad (\text{A.18})$$

Appendix B

ABC-SMC

The ABC SMC algorithm is derived from the sequential importance sampling. Let π be the target distribution we want to sample from. If it is impossible to sample from π directly then one can sample from a suitable proposal distribution, η , and use importance sampling weights to approximate π . For example let φ be a distribution having the same support as π and we wish to obtain the expectation of φ w.r.t π . This is given as

$$E_{\pi}[\varphi(x)] = \int \varphi(x)\pi(x)dx. \quad (\text{B.1})$$

The above equation can be written as

$$\begin{aligned} E_{\pi}[\varphi(x)] &= \int \varphi(x) \frac{\pi(x)}{\eta(x)} \eta(x) dx \\ &= \int \varphi(x) w(x) \eta(x) dx, \end{aligned} \quad (\text{B.2})$$

where $w(x) = \pi(x)/\eta(x)$ is the unnormalized *importance weight function*. Equation [B.2](#) is known as the importance sampling identity. By sampling N particles (samples) $\{X^{(i)}\}$ from the proposal distribution η and substituting the Monte Carlo approximation

$$\eta(dx) = \frac{1}{N} \sum_{i=1}^N \delta_{X^{(i)}}(dx), \quad (\text{B.3})$$

where $\delta_{x_0}(x)$ is the Dirac delta function which is defined as

$$\int_x f(x) \delta_{x_0}(x) dx = f(x_0), \quad (\text{B.4})$$

of this distribution into equation B.2 we can obtain the approximation of $E_\pi[\varphi(x)]$. Therefore using the importance weights the target density can be approximated point-wise as

$$\hat{\pi}(dx) = \frac{1}{N} \sum_{i=1}^N w(X^{(i)}) \delta_{X^{(i)}}(dx). \quad (\text{B.5})$$

Hence one can, in order to sample from a target distribution π , sample from the proposal distribution, η , and then weight the samples by importance weights, w . In SIS, one reaches the target distribution π_T through a series of intermediate distributions π_t , where $t = 1, \dots, T - 1$ are the indexes of these intermediate distributions. By using the importance sampling idea described above one can sample from a series of proposal distributions η_t and weight the samples by importance weights

$$w_t(x_t) = \frac{\pi_t(x_t)}{\eta_t(x_t)}. \quad (\text{B.6})$$

Within the SIS, the proposal distributions are defined as

$$\eta_t(x_t) = \int \eta_{t-1}(x_{t-1}) k_t(x_{t-1}, x_t) dx_{t-1}, \quad (\text{B.7})$$

where η_{t-1} is the previous proposal distribution and k_t is a Markov kernel (also called a perturbation kernel in SMC literature) which defines a random walk around the sample x_{t-1} .

To apply SIS in the ABC context, we need to define the intermediate and proposal distributions. Considering the SIS framework as the base we can define the ABC-SMC algorithm to be a special case of the SIS algorithm. We now choose the intermediate and proposal distributions in an ABC fashion (incorporating the distance metric $\Delta(Y_d, Y_s) \leq \epsilon$ while defining these distributions). The intermediate distributions are defined as

$$\pi_t(x_t) = \pi(x) \mathbb{1}(\Delta(Y_d, Y_s) \leq \epsilon_t) \quad (\text{B.8})$$

where $\pi(x)$ is the prior distribution; $\mathbb{1}(x)$ is the indicator function and ϵ_t is the tolerance satisfied by particles contributing to the intermediate distribution $\pi_t(x_t)$. We define the first proposal distribution equal to the prior $\eta_1 = \pi$ and the proposal distributions at time t ($t = 2, \dots, T$), η_t , is defined as the perturbed intermediate distribution at time $t - 1$, π_{t-1} , such that for every particle we have $\pi(x) \geq 0$, satisfying the condition $\pi_t(x_t) \geq 0 \Rightarrow \eta_t(x_t) \geq 0$. Hence we have

$$\eta_t(x_t) = \mathbb{1}(\pi_t(x_t) \geq 0) \times \int \pi_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t) dx_{t-1}, \quad (\text{B.9})$$

where K_t denotes the perturbation kernel.

To calculate the weights defined as

$$w_t(x) = \frac{\pi_t(x)}{\eta_t(x)}, \quad (\text{B.10})$$

we need to evaluate $\eta_t(x_t)$ in an appropriate way. Standard Monte Carlo approximation can be used

$$\int \eta_{t-1}(x) K_t(x_{t-1}, x_t) dx = \frac{1}{N} \sum_{i=1}^N K_t(x_{t-1}, x_t), \quad (\text{B.11})$$

as shown in [Del Moral et al. \(2006\)](#) to obtain

$$\begin{aligned} \eta_t(x_t) &= \mathbf{1}(\pi_t(x_t) \geq 0) \times \int \pi_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t) dx_{t-1} \\ &= \mathbf{1}(\pi_t(x_t) \geq 0) \times \int w_{t-1}(x_{t-1}) \eta_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t) dx_{t-1} \\ &= \mathbf{1}(\pi_t(x_t) \geq 0) \times \frac{1}{N} \sum_{x_{t-1}^{(i)} \sim \pi_{t-1}} w_{t-1}(x_{t-1}^{(i)}) K_t(x_{t-1}^{(i)}, x_t), \end{aligned} \quad (\text{B.12})$$

where N denotes the number of particles and $x_{t-1}^{(i)}$, $i = 1, \dots, N$, are all the particles from the intermediate distribution π_{t-1} . The unnormalized weights can then be calculated as

$$w_t(x_t) = \frac{\pi(x_t)}{\sum_{x_{t-1}^{(i)} \sim \pi_{t-1}} w_{t-1}(x_{t-1}^{(i)}) K_t(x_{t-1}^{(i)}, x_t)}, \quad (\text{B.13})$$

for all accepted particles x_t .

Appendix C

Fisher information matrix for ODEs

C.1 Evaluating the Fisher information matrix for ODEs

Let us consider a coupled ODE system having K states where Y^d are the experimental measurements of the states $\mathbf{X}(t^L, \boldsymbol{\theta})$ at discrete time points $t^L \triangleq \{t_i\}_{i=1, \dots, L}$. We further assume that Y_k^d is the state trajectory X_k (a vector of state values evaluated at t^L) corrupted by additive Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 \mathbb{I}_L)$, where σ_k^2 is the standard deviation for the k -th state. For such a Gaussian noise model we have the log likelihood given by

$$\begin{aligned} \mathcal{L}(Y^d | \mathbf{X}, \boldsymbol{\sigma}) &= \prod_{k=1}^K \mathcal{N}(Y_k^d | X_k, \sigma_k^2 \mathbb{I}_L) \\ &\propto \sum_{k=1}^K (Y_k^d - X_k)^T \Sigma_k^{-1} (Y_k^d - X_k), \end{aligned} \tag{C.1}$$

where $\Sigma_k^{-1} = \sigma_k^2 \mathbb{I}_L$. It is worth noting that with a constant noise for each of the state dimensions we can approximate the Mahalanobis distance in the above equation with an Euclidean distance. This results in the likelihood to be equivalent to the ABC distance function that we have used previously:

$$\begin{aligned} \mathcal{L} &\propto \sum_{k=1}^K (Y_k^d - X_k)^T (Y_k^d - X_k) \\ &= \sum_{k=1}^K \sum_{i=1}^L \left(Y_k^d(t_i) - X_k(t_i) \right)^2 \\ &= \Delta(Y^d, Y^s). \end{aligned} \tag{C.2}$$

The derivative of the log likelihood with respect to each of the elements of the parameter vector $\boldsymbol{\theta}$ is given by

$$\frac{\partial \mathcal{L}}{\partial \theta_j} \propto \sum_{k=1}^K (Y_k^d - X_k)^T \Sigma_K^{-1} \mathbf{Z}_{kj}, \quad (\text{C.3})$$

where the L -dimensional sensitivity vector $\mathbf{Z}_{kj}$ has its elements as the gradient of the k -th state with respect to the j -th parameter, denoted as $z_{kj}(t) = \frac{\partial X_k(t)}{\partial \theta_j}$. One way to obtain these first order sensitivities is to use numerical derivatives using finite differences, while compromising accuracy. The best way to calculate these sensitivities is by augmenting the original set of differential equations $\frac{d}{dt} \mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta})$ with the first order sensitivity equations given as:

$$z_{kj}(t) = \frac{\partial}{\partial \theta_j} \frac{\partial X_k(t)}{\partial t} = \sum_{i=1}^K \frac{\partial f_k}{\partial X_i(t)} \frac{\partial X_i(t)}{\partial \theta_j} + \frac{\partial f_k}{\partial \theta_j}. \quad (\text{C.4})$$

The entire set of equations describing the state and the sensitivities can now be solved together using any standard numerical ODE solver. Although this later method of obtaining the sensitivities is the most accurate one, it introduces additional computational overload by increasing the number of differential equations needed to be solved. Since the log likelihood given in equation C.1 has the form of a product of multivariate Gaussian densities, we can use the expression of FIM $\mathbf{I}$ for a single multivariate normal $\mathcal{N}(\mu(\boldsymbol{\theta}), \Sigma)$ given by (Porat and Friedlander, 1986)

$$I(\boldsymbol{\theta})_{j,l} = \frac{\partial \mu^T}{\partial \theta_j} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_l}, \quad (\text{C.5})$$

to calculate the FIM $\mathbf{I}_{ODE}$ for the ODE model trivially as (Girolami and Calderhead, 2011):

$$\mathbf{I}_{ODE}(\boldsymbol{\theta})_{j,l} = E \left[\left\{ \frac{\partial}{\partial \theta_j} \mathcal{L}(Y^d | \mathbf{X}, \boldsymbol{\sigma}) \right\} \left\{ \frac{\partial}{\partial \theta_l} \mathcal{L}(Y^d | \mathbf{X}, \boldsymbol{\sigma}) \right\} \right] = \sum_{k=1}^K \mathbf{Z}_{kj}^T \Sigma_K^{-1} \mathbf{Z}_{kl}. \quad (\text{C.6})$$

This expression of the FIM is used in Filippi et al. (2013) to design the covariance of a multivariate perturbation kernel and in Girolami and Calderhead (2011) to design a geometrically motivated Hamiltonian Monte Carlo algorithm.

C.2 Approximation of the Hessian matrix for ODEs

Apart from the FIM we can use the sensitivity equations to approximately evaluate the Hessian of the least squares cost function (which is same as the Euclidean ABC distance function) for an ODE system. Considering the same ODE model with K states we can

define the least squares cost function $C(\boldsymbol{\theta})$ as:

$$C(\boldsymbol{\theta}) = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^L \left(Y_k^d(t_i) - X_k(t_i) \right)^2. \quad (\text{C.7})$$

We can define the gradient of this cost function as:

$$\frac{\partial C}{\partial \theta_j} = - \sum_{k=1}^K \sum_{i=1}^L \frac{\partial X_k(t_i)}{\partial \theta_j} = - \sum_{k=1}^K \sum_{i=1}^L r_k(t_i) z_{kj}(t_i) \quad (\text{C.8})$$

where $r_k(t_i) = Y_k^d(t_i) - X_k(t_i)$ is defined as the residual at time t_i . The curvature of $C(\boldsymbol{\theta})$, the Hessian matrix, can be evaluated by taking the second derivative of the cost as:

$$\frac{\partial^2 C}{\partial \theta_j \partial \theta_l} = \sum_{k=1}^K \sum_{i=1}^L z_{kj}(t_i) z_{kl}(t_i) - \sum_{k=1}^K \sum_{i=1}^L r_k(t_i) \frac{\partial z_{kj}(t_i)}{\partial \theta_l}. \quad (\text{C.9})$$

The second term in equation C.9 can be neglected if the residuals are small. The element of the Hessian matrix $\mathbf{H}_{ODE}$ at (j, l) is then approximated by

$$\mathbf{H}_{ODE}(j, l) \approx \sum_{k=1}^K \sum_{i=1}^L z_{kj}(t_i) z_{kl}(t_i) \quad (\text{C.10})$$

References

- E. A. Ainsworth, C. R. Yendrek, S. Sitch, W. J. Collins, and L. D. Emberson. The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annual review of plant biology*, 63:637–661, 2012.
- R. M. Anderson, R. M. May, and B. Anderson. *Infectious diseases of humans: dynamics and control*. Oxford University press, 1991.
- S. Barber, J. Voss, M. Webster, et al. The rate of convergence for approximate bayesian computation. *Electronic Journal of Statistics*, 9:80–105, 2015.
- M. A. Beaumont, J. M. Cornuet, J. M. Marin, and C. P. Robert. Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990, 2009.
- M. A. Beaumont, W. Zhang, and D. J. Balding. *Genetics*, 162(4):2025–2035, 2002.
- M. J. Beilby. Action potential in charophytes. *International review of cytology*, 257: 43–82, 2007.
- J. M. Bernardo and A. FM. Smith. *Bayesian theory*. IOP Publishing, 2001.
- C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- E. V. Bonilla, K. M. Chai, and C. Williams. Multi-task gaussian process prediction. In *Advances in neural information processing systems*, pages 153–160, 2007.
- P. Buschmann and D. Gradmann. Minimal model for oscillations of membrane voltage in plant cells. *Journal of theoretical biology*, 188(3):323–332, 1997.
- B. Calderhead and M. Girolami. Estimating Bayes factors via thermodynamic integration and population MCMC. *Comput. Stat. Data Anal.*, 53:4028–4045, 2009.
- B. Calderhead, M. Girolami, and N. D. Lawrence. Accelerating Bayesian Inference over Nonlinear Differential Equations with Gaussian Processes. In *Proc. Advances in Neural Information Processing Systems 21*, pages 217–224, 2008.
- R. Capone, B. S. Tiwari, and A. Levine. Rapid transmission of oxidative and nitrosative stress signals from roots to shoots in arabidopsis. *Plant Physiology and Biochemistry*, 42(5):425–428, 2004.

- S. K. Chatterjee, S. Das, K. Maharatna, E. Masi, L. Santopolo, S. Mancuso, and A. Vitaletti. Exploring strategies for classification of external stimuli using statistical features of the plant electrical response. *Journal of The Royal Society Interface*, 12(104), 2015.
- S. K. Chatterjee, S. Ghosh, S. Das, V. Manzella, A. Vitaletti, E. Masi, L. Santopolo, S. Mancuso, and K. Maharatna. Forward and inverse modelling approaches for prediction of light stimulus from electrophysiological response in plants. *Measurement*, 53:101–116, July 2014.
- WG. Choi, M. Toyota, SH. Kim, R. Hilleary, and S. Gilroy. Salt stress-induced ca^{2+} waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proceedings of the National Academy of Sciences*, 111(17):6497–6502, 2014.
- A. Christmann and E. Grill. Plant biology: Electric defence. *Nature*, 500(7463):404–405, 2013.
- A. Christmann, E. W. Weiler, E. Steudle, and E. Grill. A hydraulic signal in root-to-shoot signalling of water shortage. *The Plant Journal*, 52(1):167–174, 2007.
- H. Clayton, M. R. Knight, H. Knight, M. R. McAinsh, and A. M. Hetherington. Dissection of the ozone-induced calcium signature. *The Plant Journal*, 17(5):575–579, 1999.
- H. Cramér. *Mathematical methods of statistics*, volume 9. Princeton university press, 1999.
- P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436, June 2006.
- P. Del Moral, A. Doucet, and A. Jasra. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, 22(5):1009–1020, 2012.
- X. Didelot, R. G. Everitt, A. M. Johansen, and D. J. Lawson. Likelihood-free estimation of model evidence. *Bayesian analysis*, 6(1):49–76, March 2011.
- F. Dondelinger, M. Filippone, S. Rogers, and D. Husmeier. ODE parameter inference using adaptive gradient matching with Gaussian processes. In *Proc. 16th Int’l Conf. on Artificial Intelligence and Statistics*, pages 216–228, 2013.
- C. C. Drovandi and A. N. Pettitt. Estimation of parameters for macroparasite population evolution using approximate bayesian computation. *Biometrics*, 67(1):225–33, March 2011.
- R. Durichen, M. AF. Pimentel, L. Clifton, A. Schweikard, and D. A. Clifton. Multitask gaussian processes for multivariate physiological time-series analysis. *IEEE Transactions on Biomedical Engineering*, 62(1):314–322, 2015.

- N. H. Evans, M. R. McAinsh, A. M. Hetherington, and M. R. Knight. Ros perception in arabidopsis thaliana: the ozone-induced calcium response. *The Plant Journal*, 41(4):615–626, 2005.
- H. H. Felle and M. R. Zimmermann. Systemic signalling in barley through action potentials. *Planta*, 226(1):203–214, 2007.
- S. Filippi, C. P. Barnes, J. Cornebise, and M. PH. Stumpf. On optimality of kernels for approximate bayesian computation using sequential monte carlo. *Statistical applications in genetics and molecular biology*, 12(1):87–107, 2013.
- J. Fromm and S. Lautner. Electrical signals and their physiological significance in plants. *Plant. Cell Environ.*, 30:249–257, 2007.
- J. Gallier. The schur complement and symmetric positive semidefinite (and definite) matrices, <http://www.cis.upenn.edu/~jean/schur-comp.pdf/>, 2010.
- A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*. 2003.
- M. Girolami. Bayesian inference for differential equations. *Theoretical Computer Science*, 408(1):4–16, 2008.
- M. Girolami and B. Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.
- L. Glass and M. C. Mackey. Pathological conditions resulting from instabilities in physiological control systems. *Annals of the New York Academy of Sciences*, 316(1):214–235, 1979.
- D. Gradmann and P. Buschmann. Oscillatory interactions between voltage gated electroenzymes. *Journal of experimental botany*, 48(Special Issue):399–404, 1997.
- D. Gradmann and H. Mummert. Action potentials in acetabularia: Measurement and simulation of voltage-gated fluxes. *Membrane Biology*, 124:265–273, 1991.
- A. Grelaud, C. P. Robert, JM. Marin, F. Rodolphe, JF. Taly, et al. Abc likelihood-free methods for model choice in gibbs random fields. *Bayesian Analysis*, 4(2):317–335, 2009.
- 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 computational biology*, 3(10):e189, 2007.
- B. Hille. *Ion channels of excitable membranes*, volume 507. Sinauer Sunderland, MA, 2001.
- R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the american statistical association*, 90:773–795, 1995.

- K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.
- B. Macdonald, M. Niu, S. Rogers, Ma. Filippone, and D. Husmeier. Approximate parameter inference in systems biology using gradient matching: a comparative evaluation. *BioMedical Engineering OnLine*, 2016.
- D. JC. MacKay. Introduction to gaussian processes. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166, 1998.
- D. JC. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
- S. Mancuso. Hydraulic and electrical transmission of wound-induced signals in vitis vinifera. *Functional Plant Biology*, 26(1):55–61, 1999.
- JM. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.
- D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- G. Mayraz and G. E. Hinton. Recognizing handwritten digits using hierarchical products of experts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):189–197, 2002.
- G. Miller, K. Schlauch, R. Tam, D. Cortes, M. A. Torres, V. Shulaev, J. L. Dangl, and R. Mittler. The plant nadph oxidase rbohD mediates rapid systemic signaling in response to diverse stimuli. *Science Signaling*, 2(84):ra45–ra45, 2009.
- NAM. Monk. Oscillatory expression of Hes1, p53, and NF- κ B driven by transcriptional time delays. *Current Biology*, 13:1409–1413, 2003.
- S. AR. Mousavi, A. Chauvin, F. Pascaud, S. Kellenberger, and E. E. Farmer. Glutamate receptor-like genes mediate leaf-to-leaf wound signalling. *Nature*, 500(7463):422–426, 2013.
- Y. Murakami. Bayesian parameter inference and model selection by population annealing in systems biology. *PloS one*, 9(8):e104057, 2014.
- J.D. Murray. *Mathematical Biology: I. An Introduction*. Springer, 2002.
- R. M. Neal. Regression and Classification Using Gaussian Process Priors. *Bayesian Statistics*, 6:475–501, 1998.
- A. O’Hagan and JFC. Kingman. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B*, 40:1–42, 1978.

- B. G. Pickard. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *The Botanical Review*, 39(12):172–201, December 1973.
- B. Porat and B. Friedlander. Computation of the exact information matrix of gaussian time series with stationary random components. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(1):118–130, Feb 1986. ISSN 0096-3518.
- J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798, 1999.
- C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91, 1945.
- J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B*, 69(5):741–796, November 2007.
- C. E. Rasmussen and H. Nickisch. Gaussian processes for machine learning (gpml) toolbox. *The Journal of Machine Learning Research*, 9999:3011–3015, 2010.
- C. E. Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006.
- C. P. Robert, JM. Cornuet, JM. Marin, and N. S. Pillai. Lack of confidence in approximate bayesian computation model choice. *Proceedings of the National Academy of Sciences*, 108(37):15112–15117, 2011.
- J. I. Schroeder and S. Hagiwara. Cytosolic calcium regulates ion channels in the plasma membrane of vicia faba guard cells. *Nature*, 338:427–430, 1989.
- M. Secrier, T. Toni, and M. PH. Stumpf. The abc of reverse engineering biological signalling systems. *Molecular BioSystems*, 5(12):1925–1935, 2009.
- J. R. Shewchuk. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. *Science (80-.)*, 49:64, 1994.
- S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.
- S. A Sisson, Y. Fan, and M. M. Tanaka. Correction for sisson sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 106(39):16889, 2009.
- E. Solak, R. Murray-Smith, W.E. Leithead, C.E. Rasmussen, and D.J. Leith. Derivative observations in Gaussian process models of dynamic systems. In *Proc. Advances in Neural Information Processing Systems 15*, pages 1033–1040, 2002.

- R. Stahlberg and D. J. Cosgrove. Rapid alterations in growth rate and electrical potentials upon stem excision in pea seedlings. *Planta*, 187:523–531, 1992.
- B. Stankovic, T. Zawadzki, and E. Davies. Characterization of the Variation Potential in Sunflower. *Plant Physiol.*, 115(3):1083–1088, November 1997.
- V. Sukhov, E. Akinchits, L. Katicheva, and V. Vodeneev. Simulation of variation potential in higher plant cells. *The Journal of membrane biology*, 246(4):287–296, 2013.
- V. Sukhov and V. Vodeneev. A mathematical model of action potential in cells of vascular plants. *Journal of Membrane Biology*, 232(1-3):59–67, 2009.
- T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P.H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, February 2009.
- M. K. Transtrum, B. B. Machta, and J. P. Sethna. Geometry of nonlinear least squares with applications to sloppy models and optimization. *Physical Review E*, 83(3):036701, 2011.
- M. K. Transtrum and P. Qiu. Optimal experiment selection for parameter estimation in biological differential equation models. *BMC bioinformatics*, 13(1):181, 2012.
- T. Vahisalu, I. Puzõrjova, M. Brosché, E. Valk, M. Lepiku, H. Moldau, P. Pechter, YS. Wang, O. Lindgren, J. Salojärvi, et al. Ozone-triggered rapid stomatal response involves the production of reactive oxygen species, and is controlled by slac1 and ost1. *The Plant Journal*, 62(3):442–453, 2010.
- JM. Varah. A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing*, 3(1):28–46, 1982.
- V. Vyshemirsky and M. A. Girolami. Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6):833–839, 2008.
- Y. Wang and D. Barber. Gaussian Processes for Bayesian Estimation in Ordinary Differential Equations. In *Proc. 31st Int’l. Conf. Machine Learning*, pages 1485–1493, 2014.
- A. G. Wilson and R. P. Adams. Gaussian process kernels for pattern discovery and extrapolation. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1067–1075, 2013.
- M. R. Zimmermann, H. Maischak, A. Mithöfer, W. Boland, and H. H. Felle. System potentials, a novel electrical long-distance apoplastic signal in plants, induced by wounding. *Plant Physiology*, 149(3):1593–1600, 2009.