University of Auckland

Thesis Proposal

Parametrisation and Identification Methods for Epidemic Systems

David Wu

supervised by

Dr. Oliver Maclaren

and

Dr. Vinod Suresh

Department of Engineering Science

November 2019

Abstract

Although mathematical modelling in epidemiology is well-established, there are still many open questions about how best to relate these models to real-world data. While conclusions drawn from purely mathematical analysis are useful for describing overarching behaviours of the system, these results are less useful for quantitative inferences and prediction. For these purposes, estimation of the parameters and the state of the model is required.

Trajectory-matching methods are the most common methodology for fitting epidemiological models such as ordinary differential equations (ODEs) to data. These are usually based on maximum likelihood methods that, under Gaussian error assumptions, reduce to repeatedly computing the least-squares error between the solution of the forward model and the data. However, this process does not allow for errors in the model structure itself, and can be computationally expensive to compute. Furthermore, naive trajectory matching methods can fail badly in the presence of identifiability problems, and detecting these issues adds further difficulties and computational effort. Other methods such as gradient matching or Markov Chain Monte Carlo (MCMC) suffer from similar problems due to identifiability issues, structural model error and/or computation time.

The aim of this project is to adapt a method from the field of functional data analysis for fitting ODEs to data, and apply it in the field of epidemic modelling. In this context, the mathematical model is viewed as a regularisation term in an optimisation problem, and need only be approximately satisfied. Furthermore, in this formulation, the need to solve the model directly can be bypassed. Automatic differentiation tools enable the resulting fitting problem to be efficiently solved using derivative-based optimisation methods.

Initial analyses of toy models and real systems have already been performed. We also consider the method's usefulness for solving identifiability problems and develop ways of performing uncertainty quantification. In future, we are interested in further incorporating uncertainty quantification into the methodology, and extending this method to multi-dimensional models, in both space and time, to answer questions about epidemic propagation. We are working in collaboration with researchers from the School of Population Health at the University of Auckland on applying these methods to the recent measles and mumps outbreaks in New Zealand.

Contents

List of Figures

Nomenclature

u_x	Derivative of u with respect to x
$\dot{u}$	Derivative of u with respect to time, t
θ	Vector of parameters
y	Vector of observations
x	Vector of state variables
$\|\cdot\|$	The L-2 norm
inf	Infinum
min	Minimum
$\mathcal{N}(\mu, \Gamma)$	Gaussian distribution with mean μ and covariance Γ
$\mathcal{GP}(\mu, k(\cdot, \cdot))$	Gaussian process with mean μ and covariance function k
S	Number of Susceptibles
E	Number of Exposed
I	Numberof Infectious
R	Number of Removed
N	Size of the at-risk population
$\mathcal{R}_0$	Basic reproduction number

1 Background

1.1 Problem Overview

In much of mathematical biology, real-world biological systems are often idealised in the modelling process. For understanding the overarching dynamics and structure of such biological systems, this idealisation is both inevitable and invaluable. However, this can lead to problems when trying apply the results of theoretical analysis to real-world scenarios.

Modelling is a powerful tool for answering problems about the behaviour of systems. For epidemiological applications, these problems fall roughly into two categories: prevention and control. Though analytical results, such as the ones elaborated by Rass and Radcliffe [1], Capasso [2], or Diekmann [3], are useful in providing general guidelines for asymptotic behaviours, the transient dynamics may be more useful to practitioners. Therefore a more specific prediction must be able to be made.

Such predictions can be made by estimating the parameters of the mathematical model, and using these estimates to generate an estimate of the state of the system in the past and into the future. From this, control strategies can be implemented and simulated, creating meaningful predictions, and allowing for informed decisions to be made. This type of modelling and prediction is particularly important for infectious diseases, where the consequences can be globally significant, and intervention requires significant investment of resources.

This thesis aims to explore the use of optimisation methods and functional data analysis techniques to perform the parametrisation of mathematical models, as a complementary tool to typical least-squares and Bayesian MCMC methods. In particular, we are interested in fitting some (epidemiological) model of the form

$$\mathcal{D}u = f(u;\theta) \tag{1.1}$$

to some data y, where u is the state variable, $\mathcal{D}$ is some differential operator, f some arbitrary function, and θ a finite-dimensional vector of parameters. This model can be solved to produce a solution for u, if given θ, initial, and boundary conditions. We term this the *forward problem*.

An example of a model of this form is the ordinary differential equation:

$$\frac{du}{dt} = f(u;\theta), \tag{1.2}$$

which is the typical form of compartmental models in mathematical epidemiology. These compartmental models of epidemiology have a relatively long history, dating back to the seminal Kermack and McKendrick [4] paper. The general idea is to partition the population into several compartments that describe their state of infection. The simplest type, the SIR model as presented in (1.3), is still in wide use as a baseline model that some practitioners use to fit their data to. The model can be written as

$$\begin{aligned} \dot{S} &= -\beta SI/N, \\ \dot{I} &= \beta SI/N - \alpha I, \\ (\dot{R} &= \alpha I), \\ N &= S + I + R. \end{aligned} \tag{1.3}$$

Here, S represents the number of susceptibles, I the number of infected, R the number of removed/recovered, N the total size of the at-risk population, α the recovery rate, and β the effective contact/transmission rate.

To make predictions using this model, we would wish to determine the values of the parameters α and β, as well as initial conditions of the state variables S, I and R (we could assume that $I(0) = 1$, $S(0) = N - 1$, $R(0) = 0$ to reduce this to estimating just one more value). This would then allow us to solve the problem forward in time for prediction, or perturb the system to model control. We term this estimation of the parameters the *inverse problem*.

1.2 State of the Art

The estimation of the parameters (also known as model calibration) for dynamic (time-varying) models, such as the SIR model, is typically done by relating the parameter estimation problem to a state estimation problem. A common approach is to assume a normally distributed, additive error in the observation of the system [5]–[8], which can be represented

$$y = u + \epsilon, \ \epsilon \sim \mathcal{N}(0, \sigma^2), \tag{1.4}$$

where y is the data, and u is the state of the model system. Taking a log-likelihood of this error model gives what can be termed *trajectory matching* [9], where the sum of squares error between a model-generated trajectory and the data is minimised. This is also known as the nonlinear least-squares approach, and can be written as

$$\theta_{opt} = \inf_{\theta} \ \|y - u(\theta, u_0)\|^2 \tag{1.5}$$

where θ is the vector of parameters, and u_0 the initial state of the system (often also estimated). In this formulation, the state of the model system u is generated through the exact solution of the differential equation system in (1.1), i.e. solving the forward problem, and imposing the model exactly. The solution of the forward problem is also required to be performed at every iteration of the optimisation procedure, meaning that the solution of the inverse problem in this way requires multiple solutions of the forward problem.

Unfortunately, this method displays some problems typical of nonlinear, nonconvex objective functions. Namely, the parameter estimate recovered is sensitive to the initial iterate, or *unstable*. Some methods have been proposed for the selection of this initial iterate [10]; typically though, this problem is side-stepped by simply performing the optimisation many times with a large number of initial iterates [6], [11], [12].

This problem is exacerbated in biological systems, where the model is generally inadequate for capturing the dynamics of the observations [13], or we have low confidence in our model. There are explicit ways of formulating a model which can account for this sort of error [14], but the forward problem is generally computationally expensive to solve. Even without this additional complexity, the forward problem can be expensive to solve, for example, if there are many states,

i.e u is high-dimensional, such as in PDE models. Indirect methods can be used to solve this problem, by performing least-squares on the vector field, instead of the trajectory, and avoiding solving the forward problem [9]. However, these methods require full observability of all state variables, or make very strong assumptions about the trajectories of the unobserved states. This makes them difficult to use in practice, since typically data is only measured for a subset of the states of the system. Methods have been developed to solve this partial observation problem, such as the all-at-once method in the geophysics literature [15], or the generalised profiling (also parameter cascading) method in the statistical literature, the latter of which we explore in detail in Section 3.1. These attempt to estimate the underlying state and the parameters at the same time, which interpolates data where it exists, and uses the model to infer the unobserved states.

One other problem with the above optimisation methods is that they require strong assumptions about the form of error in the model in order to quantify the uncertainty, such as having only independent, additive, Gaussian observation noise. To address this, the standard tool is Bayesian analysis. This consists of computing the conditional probability of the parameters given the data, from a likelihood model (probability of the data conditioned on the parameters) and a prior on the parameters. Typically, this is not an analytically tractable problem, due to the probability distributions used, or the form of the likelihood model. Thus, numerical methods are used to generate a set of samples that converge onto the posterior distribution, allowing it to be characterised. Sampling methods such as Monte Carlo Markov Chain (MCMC) are now used extensively for this purpose. However, these methods are computationally expensive as they require a high number of forward model solves in order to generate the sample distribution. Additionally, MCMC methods do not alleviate the model inadequacy and computational expense problems of optimisation-based trajectory matching. This is because the likelihood model is typically constructed using the trajectory matching functional [16]. Computational expense problems are further exacerbated from the sampling procedure. This compounds if the model inadequacy problem is addressed, by modelling process error, for example by simulating a stochastic process.

A more thorough review of the literature and associated theory is provided in the attached literature review [17], along with some description of epidemiological models.

1.3 Motivation

The act of parametrisation is not merely a mathematical or statistical exercise. Infectious diseases have sweeping negative effects on countries if not properly controlled. Providing quantitative predictions of these effects is important in the decisions for executing different control strategies. Further, with social factors leading to decreases in vaccination coverage, it becomes more important that the risks of future outbreaks be assessed, and suitable control strategies evaluated. The recent mumps and measles outbreaks in Auckland have emerged as an interesting case study for answering questions about the policy surrounding responses to epidemic outbreaks. Questions have arisen regarding the efficacy of the control strategy implemented for mumps, and its effect on the later measles outbreak, and policies for future response, which forms part of the future work in Section 5.1.

2 Scope

The objectives of this research project are to:

1. Develop a reusable package for the parameterisation and identifiability analysis of epidemic models, based on generalised profiling, a tool from functional data analysis
2. Extend the methods to multidimensional problems
3. Analyse a range of different epidemiological models including
    - 2019 measles outbreak in New Zealand
    - Relations with and imbued immunity from the 2017 mumps outbreak and response
4. Explore the Bayesian equivalents for variants of this method, and determine efficiency gains from the optimisation simplifications

The following tasks will be undertaken as a part of the proposed research:

1. Construct a classification of relevant epidemiological models and their applicability to specific diseases
2. Develop a package for fitting non-spatial and implicitly spatial dynamic models
3. Extend the package to explicitly spatial dynamic models
4. Collate data for relevant disease outbreaks and controls
5. Perform parameter estimation studies and hypothetical modelling of relevant disease outbreaks and controls
6. Implement the “Bayesianisation" of the method
7. Compare the uncertainty of estimates under the two statistical frameworks
8. Explore hybrid/multi-paradigm methods that blend the optimisation and Bayesian frameworks

3 Methodology

3.1 Generalised Profiling

The generalised profiling method (also parameter cascading) was introduced by Ramsay, Hooker, Campbell, *et al.* [18] as a way of parametrising differential equation models. The method is motivated by the nonparametric statistical analysis field of functional data analysis. The underlying concept is that the data represents some underlying (smooth) function [19]. By decomposing this underlying function by projecting onto a basis, such as a basis of splines, the underlying function can be recovered through optimisation procedures. Extending this idea to dynamical systems, where the underlying state is modelled as a function on a basis of splines, and enforcing some model-induced penalty gives rise to the generalised profiling method. The method can also be thought of as a relaxation of the trajectory matching problem with the objective

$$\mathcal{L}(\theta, c|y) = \|y - \Phi c\|^2 + \rho\|\mathcal{D}(\Phi c) - f(\Phi c, \theta)\|^2 \tag{3.1}$$

where Φ represents a basis on which the state estimate is expanded, c the coefficient vector for that basis, and ρ a tuning parameter for the objective.

In the literature[18], [20]–[22], the above objective is used in an inner loop to estimate c conditioned on a proposed θ and given y:

$$\hat{c}(\theta) = \inf_c \mathcal{L}(c|\theta, y) \tag{3.2}$$

This is then used to perform least-squares optimisation in the "outer" optimisation to determine the optimal parameters $\hat{\theta}$:

$$\hat{\theta} = \inf_\theta H(\theta) := \inf_\theta \|y - \Phi\hat{c}(\theta)\|^2 \tag{3.3}$$

This allows for the standard properties of ordinary least squares to be used on the outer objective to recover confidence intervals.

One of the oddities of the generalised profiling method as introduced by Ramsay, Hooker, Campbell, *et al.* [18] is the discrepancy between the outer and inner objective functions. The reasoning for the form of the outer objective, as given in [18] is:

> Because $\hat{c}_i(\theta, \sigma; \lambda)$ is already regularized, criterion H does not require further regularization and is a straightforward measure of fit such as error sum of squares, log-likelihood or some other measure that is appropriate given the distribution of the errors e_{ij}.

with the context that there exist observation errors $e_{ij} \sim \mathcal{N}(0, \sigma^2)$. We take a slightly different conceptual standpoint. Because the model penalty accounts for the process error that is not captured in trajectory matching, it is an integral part of the objective function, as opposed to being a simple regularisation tool. Thus the outer objective is computed as the inner objective, with c being profiled out at its minima, i.e.

$$\mathcal{L}(\theta|y) = \min_c \mathcal{L}(\theta, c|y). \tag{3.4}$$

This also allows for computational savings, since there is little need to perform a nested optimisation, and c and θ can be optimised simultaneously to generate point estimates of interest.

Indeed, initial experimentation into performing nested optimisation did increase the solve time of a single iteration significantly.

One of the drawbacks of the generalised profiling method is that the solution is dependent on the regularisation parameter ρ. If the model f is linear, then those properties can be exploited in order to derive an explicit form of the generalised cross validation criterion, which can then be used to select ρ [9]. However, most biological models do not possess this property. This means that other methods must be used.

Interestingly, the objective function bears striking resemblance to the regularised inverse problem:

$$\mathcal{L}(\theta|y) = \|y - A(\theta)\|^2 + \lambda R(\theta) \tag{3.5}$$

for a forward model A, data y, parameters θ and regularisation function R. This leads to the interpretation of the generalised profiling method as a data interpolation problem regularised by a model fitting penalty. Using this framework we can repurpose methods of determining regularisation parameters for the choice of the tuning parameter in the generalised profiling method. A priori methods, such as the Morozov discrepancy principle [23], can be used if the magnitude of the least-squares error is known, but this is rarely the case in practice. *A posteriori* methods, such as the L-curve criterion [24], cross validation [25] or the Lepskii balancing principle [26] analyse the behaviour of the objective as λ is varied. All these methods have some ad hoc element - the L-curve criterion assumes some trend of regularisation trade-off to determine a regularisation parameter; cross validation is typically expensive, so its approximations are performed with ad hoc termination; the balancing principle chooses arbitrary threshold functions to determine the regularisation parameter. We choose to use the L-curve criterion for its simplicity, and for its ability to extend to the determination of multiple regularisation parameters [27], [28]. Furthermore, this method also corresponds in spirit to the technique used in [21] We also make the choice to propagate forward the solutions of c and θ as ρ is varied, as we expect the objective minima to vary smoothly as ρ is changed.

3.2 Identifiability

Regardless of the approach for parameter estimation, the problems of identifiability and estimability will present themselves. These refer to the questions:

1. Given an infinite amount of perfect data, can the parameters be uniquely recovered?
2. Given a finite sample of imperfect data, can the parameters be recovered to good precision?

These questions relate to the problems of structural identifiability and practical identifiability, respectively.

Structural identifiability questions can generally be answered a priori, and an array of differential algebra techniques have been developed and utilised. Two prominent techniques are Differential Algebra Identifiability of Systems (DAISY)[29] and Exact Arithmetic Rank (EAR)[30]. These both utilise computational differential algebra techniques to determine global and local structural identifiability respectively. Consider the process (3.6a) and observation (3.6b) models:

$$\dot{x}(t) = f(x(t), \theta), \tag{3.6a}$$
$$y(t) = g(x(t), \theta), \tag{3.6b}$$

for some state x and observations y.

DAISY uses Ritt's algorithm to eliminate the state x to generate algebraic functions of the parameters θ, which can then be analysed for identifiability globally as well as locally. EAR methods instead construct a Jacobian matrix of the output y with respect to the initial state and parameters (about arbitrary states and parameters), and evaluate its rank to determine local identifiability.

Practical identifiability problems generally are data-dependent. Raue, Kreutz, Maiwald, *et al.* [31] introduces the profile likelihood technique to determine practical identifiability. Though introduced for the trajectory matching objective (eq. (1.5)), consider the general objective $\mathcal{L}(\theta)$ that defines the amount of misfit for parameters θ. Then the profile likelihood $\mathcal{L}_p$ with respect to some function of the parameters $\Omega(\theta)$ constrained to some value ω is

$$\mathcal{L}_p(\omega) = \min_{\theta|\Omega(\theta)=\omega} \{\mathcal{L}(\theta|y)\}. \tag{3.7}$$

This represents the subset of the parameter space that minimises the objective function, for every value of ω. The special case where $\Omega(\theta) = \theta_i$ generates the profile of the objective with respect to a specific parameter θ_i, and can be used to determine the identifiabiltiy of a particular parmeter. As θ_i is varied, if the model is non-identifiable, then the profile will be flat, as the underlying combination of non-identifiable parameters can remain unchanged. The idea of finite confidence intervals can also be applied to the profile to give a quantitative threshold of "flatness". If the log-likelihood (i.e. objective function) remains below the confidence interval threshold for a given quantile, then the parameter is said to be practically identifiable. Another powerful application of the profile likelihood method is to confirm whether specific combinations of parameters are identifiable, such as $\mathcal{R}_0 = \frac{\beta}{\alpha}$ in the epidemiological literature.

Other methods, such as Monte Carlo simulation [32], [33] and sensitivity analysis techniques[34] are also used to analyse practical identifiability. The first of these is similar in intent to the profile likelihood method — it intends to determine if there is correlation between parameters of the model [32]. It does this by drawing a number of samples from a (true) synthetic trajectory with additive error of known variance. Then, parameters are fitted to the sample trajectories, and correlations between parameters are computed from the error in the fitted parameters. As a side-effect of this computation, unidentifiable parameters can be detected, which will have a large amount of error compared to the noise in the trajectory. This method suffers from the need to be performed on known synthetic data, which makes assumptions, such as model adequacy, that cannot be guaranteed for real datasets. Variants of this method are known as bootstrap methods [34]. Sensitivity analysis techniques use the inverse Fisher Information Matrix as a surrogate for the sample covariance to estimate correlation [34]. The Fisher Information Matrix can be computed as the Hessian of the system, and can be extracted from automatic differentiation techniques. By making a quadratic approximation of the objective function, confidence intervals can be estimated.

4 Applications

The regularised generalised profiling technique has been implemented for ordinary differential equations in Python using the `CasADi` framework [35]. The code is available at https://github.com/dwu402/pypei. The functionality of the procedure is driven by `CasADi MX` symbolics, `bspline` functions, and plugin integration with `IPOPT` as the nonlinear program solver.

The implementation has then been applied to synthetic data generated from toy ODE models and standard epidemiological models. The fitting is done with partially observed, noisy data. This method was also applied to an SEIR model in dimensional form, fitting against the reported measles case data in Auckland 2019.

4.1 A Toy Model

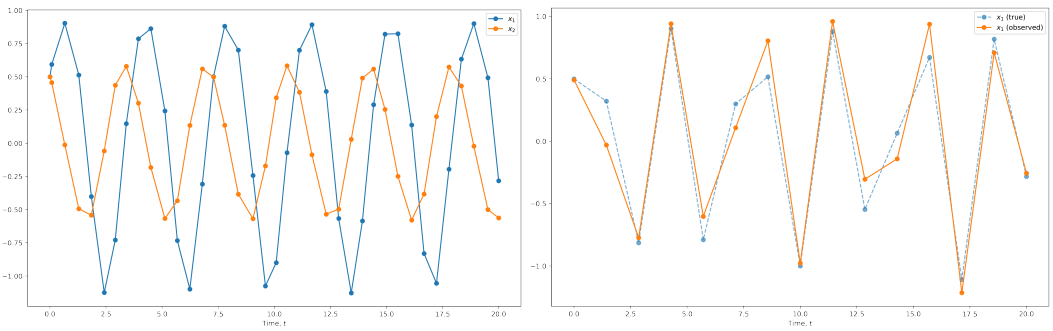
To demonstrate the properties of the methods, we examine the system:

$$\begin{aligned} \dot{x}_1 &= cx_2(K - x_1), \\ \dot{x}_2 &= -cx_1. \end{aligned} \tag{4.1}$$

We consider the observation model

$$y = x_1 + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma^2). \tag{4.2}$$

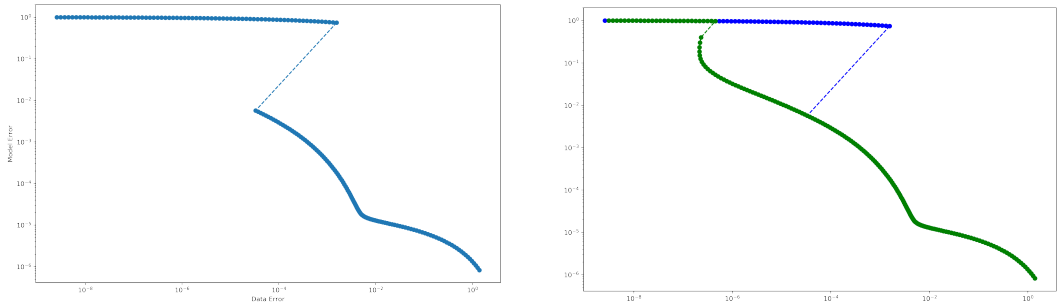
Trajectories of the system and a single realisation of the observation model are provided in Figure 4.1.



(a) Trajectories of x_1, x_2 **(b)** Realisation of y

Figure 4.1: Toy model trajectories and observation model realisation for $c = 1, K = 3, \sigma = 0.15$, where x_1 is sample uniformly at 15 points in $[0, 20]$

Fitting the generated data using the generalised profiling objective function with IPOPT for the tuning parameters $\rho \in [1 \times 10^{-6}, 1 \times 10^{6}]$ and $\alpha = (\log_{10}(p) + 6) \times 10^{-5}$, we can produce an L curve by plotting the two major error components the data fit error and model fit error against each other on log-log scales as in Figure 4.2a.



(a) L Curve **(b)** Back-Propagated L Curve

Figure 4.2: L curve generated by fitting the toy model for $\rho \in [1 \times 10^{-6}, 1 \times 10^{6}]$ on the left, and the back-propagated curve overlay on the right.

We observe an unexpected discontinuity in the L curve. This is likely to do with a nonconvexity leading to multiple near-optima, which switch on some bifurcation of the tuning parameters. We can propagate the solution backwards (in the direction of decreasing ρ) to examine if this is the case (Figure 4.2b).

We also profile the likelihoods of the objectives, to examine the identifiability of the parameters.

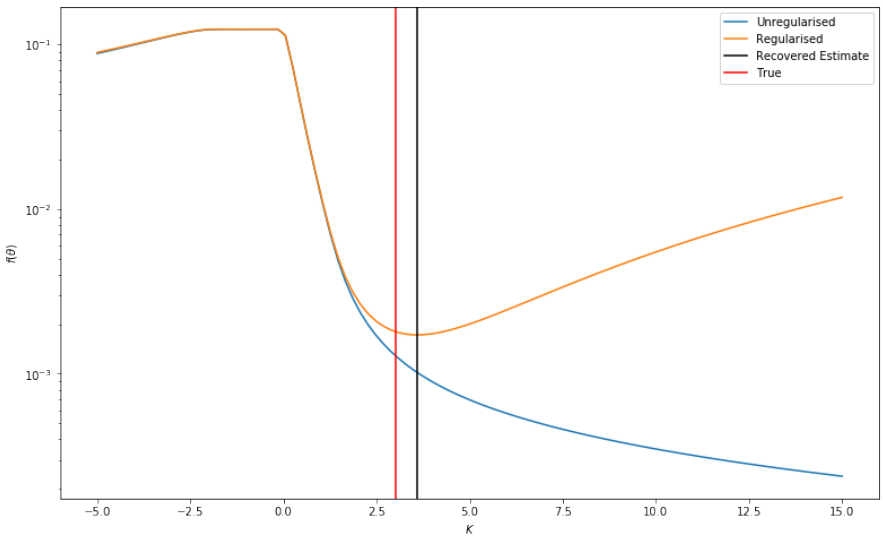


Figure 4.3: Profile of the parameter K with $\rho \approx 0.1335$, and $\alpha = 5 \times 10^{-5}$ for regularisation.

We see that the parameter K is one-sided identifiable — all we know is that the value of K must be larger than some threshold value. The regularisation therefore specifies the parameter value by choosing the tolerable amount of error as we increase the parameter from the left. We note that the true parameter value is still smaller than the recovered parameter — but a sensible result is recoverable from a likelihood maximisation method.

4.2 Synthetic Epidemics

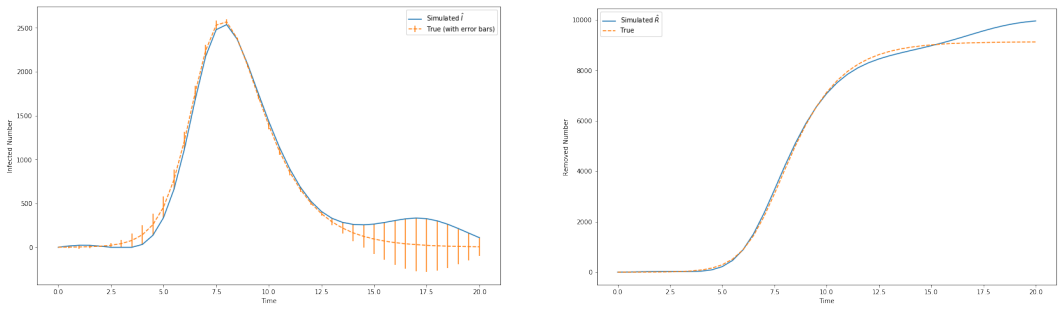
To examine the properties of the generalised profiling and profile likelihood methods in the field of interest, we generate synthetic data with the simple SIR model introduced in Equation (1.3). We impose the noise model:

$$\begin{aligned} \hat{I}(t) &\sim \mathcal{GP}\left(I(t), \sigma^2 \exp\left(-\frac{1}{2\delta^2}(t_i - t_j)^2\right)\right), \\ \hat{R}(t) &= \alpha \int_0^t \hat{I}(\tau) d\tau \end{aligned} \tag{4.3}$$

That is, the noisy infected values are modelled as a process with correlated errors, which can also be interpreted as adding process error.

$$\frac{d\hat{I}}{dt} = \frac{dI}{dt} + \epsilon.$$

The noisy removed values are the corresponding removed individuals as prescribed by the SIR model. A realisation is provided in Figure 4.4



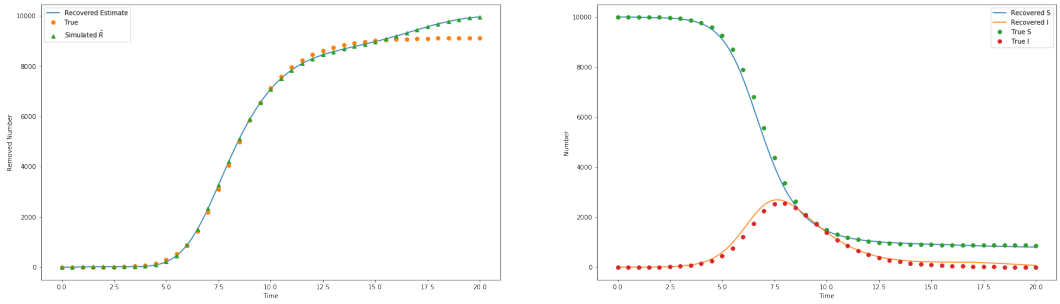
(a) Realisation of $\hat{I}$ **(b)** Realisation of $\hat{R}$

Figure 4.4: Realisation of Synthetic SIR model with Gaussian process I, $\delta = 2$, $a = 10^4$, $\alpha = 0.75$, $\Delta t = 0.5$

Trajectories are recovered with $\rho = 0.77$, which is chosen to minimise the validation error computed as

$$\frac{1}{n}\sum_{i=0}^{n}\left((S_{recovered}(t_i) - S_{true}(t_i))^2\right). \tag{4.4}$$

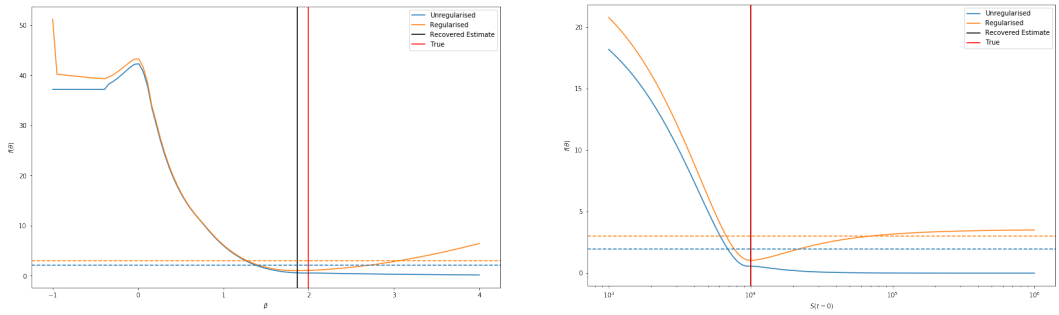
Plots of the recovered trajectory for the observed (R) and unobserved (S, I) states are provided in Figure 4.5



(a) Recovery of R **(b)** Recovery of S, I

Figure 4.5: Recovered trajectories of the model with $\rho = 0.77$, separated by state observability

We also generate profiles for three parameters: β, $S(0) = N + 1$ and α. We expect these to be nonidentifiable, due to the appearance of $\frac{\beta}{N}$ in the model, and the form of the observation model (for α). Indeed, examining the profile (log-)likelihoods of β and $S(0)$ in Figure 4.6 we can see that they possess a one-sided identifiability. An instinctive approach to bounding this interval is to apply regularisation, for example in Figure 4.6. The reconstructed intervals, which we will term penalised likelihood intervals, are now finite. However, the reduction in this intervals has a



(a) Profile of β **(b)** Profile of $S(0)$

Figure 4.6: Profiles $\mathcal{L}_p(\theta_i)$, χ^2 confidence intervals and associated penalised likelihood intervals, computed as $1.96 + \min(\mathcal{L}_p(\theta_i))$, are shown for non-regularised and regularised (Tikhonov) forms of the objective function as β and $S(0)$ are varied.

bias tradeoff - we are no longer guaranteed the asymptotic result that our interval will contain the true parameter value the corresponding percentage of samples. This effect on the likelihood intervals is characterised in more detail for the α profile, where the regularisation parameter is tuned as $\{0, 0.01, 0.5, 100\}$, respectively unregularised, weakly regularised, regularised, and strongly regularised in Figure 4.7.

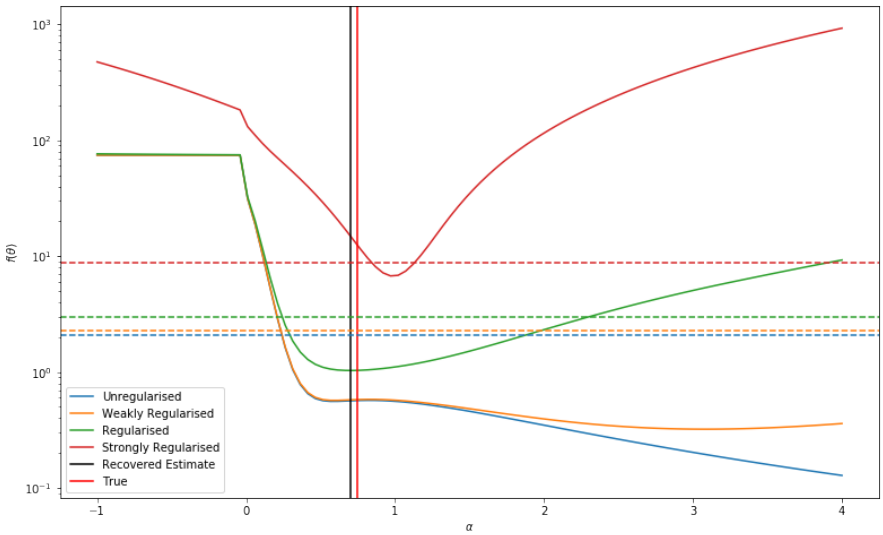


Figure 4.7: Profile $\mathcal{L}_p(\alpha)$, χ^2 confidence intervals and associate penalised likelihood intervals ($1.96+\min(\mathcal{L}_p(\alpha))$) shown for various levels of regularisation ($\{0, 0.01, 0.5, 100\}$) as the parameter α is varied

As the regularisation parameter is increased, the interval shrinks, but as displayed by the profile when $\rho = 100$, may no longer bound the true value with the same probability. This suggests that regularisation is not the (complete) solution to identifiability, and may even give misleading results, if it is not properly treated.

4.3 Measles

The 2019 outbreak of measles is an sobering reminder of the effects of New Zealand's less than optimal vaccination policies and attitudes. Concentrated in the Auckland and Upper North Island regions, particularly in the Counties Manukau DHB area, the number of reported cases reached over 1500 as of the end of September 2019 [36].

Data on the number of reported cases collated by the ESR, and weekly reports were made available to the public. This was used as the data for fitting to an SEIR model, consistent with the latency (and initial non-infectivity) patterns of measles infections:

$$\begin{aligned} \dot{S} &= -\beta SI/N, \\ \dot{E} &= \beta SI/N - \gamma E, \\ \dot{I} &= \gamma E - \alpha I, \\ \dot{R} &= \alpha R, \\ N &= S + E + I + R, \\ \theta &= \{\alpha, \beta, \gamma\}. \end{aligned} \tag{4.5}$$

We impose the observation model:

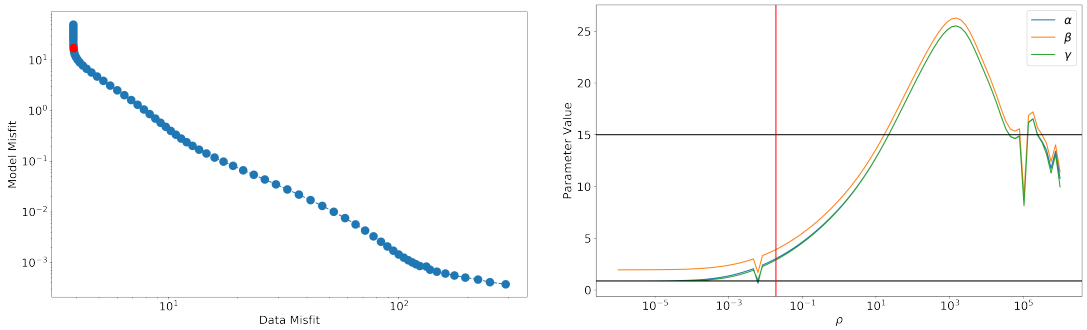
$$y = R + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{4.6}$$

where y is the cumulative weekly incidence count. The interpretation is that the reported cases correspond to the number of newly reported symptomatic cases. Assuming a 100% reporting rate, this will correspond to the number of recovered cases. Due to the fact that the system is autonomous and thus invariant under time shifts, the fact that the reported cases should lag behind the infected cases can be neglected.

We make a note here that this is not standard. Typically, the data is interpreted as the number of infected present at any given time point ($y \sim \mathcal{N}(\dot{I}, \sigma^2)$) [20] [16]. We find this incongruous with our understanding of the reporting system.

Regularisation was applied to two roughly known parameters $\alpha \approx 7/8, \gamma \approx 7/10$, representing an latent period of approximately 10 days ($7/\gamma$), and an infectious period of 8 days ($7/\alpha$). We choose to impose this prior knowledge in a less strict way to allow for deviations in the qualities of the disease. This also allows for a supplementary ad hoc criterion for determining correct regularisation parameters.

Performing the fitting, we see that the deviation from the regularisation happens when $\rho \approx 1e-3$ in Figure 4.8b, which corresponds closely with being in the region of a turn in the L-curve in Figure 4.8a.A trajectory for this value of ρ is produced in Figure 4.9.



(a) L-curve as ρ is varied.

(b) Plot of the parameters as ρ is varied.

Figure 4.8: L-curve and plot of recovered parameter estimates as ρ is varied, with $\alpha = 0.1$. The red dot (left) and line (right) correspond to $\rho = 0.005$, which was used to recover the trajectory in Figure 4.9. Black horizontal lines (left) correspond to regularisation values of $\beta = 15$ (upper) and $\alpha, \gamma = 0.875$ (lower).

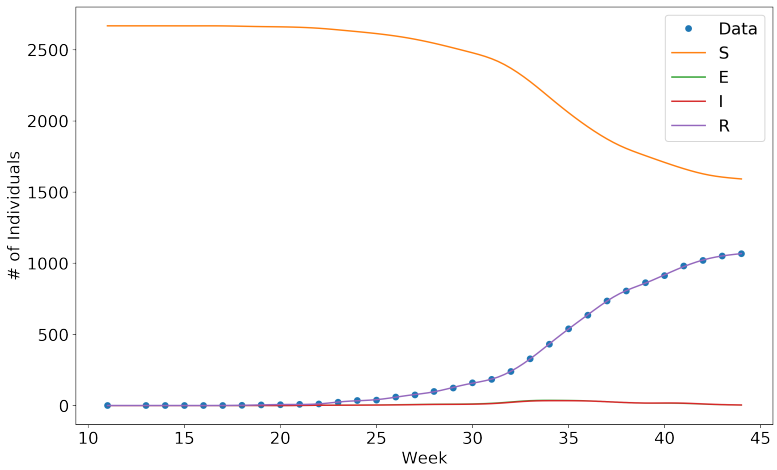


Figure 4.9: Recovered trajectory for the measles case when $\rho = 0.005$, $\alpha = 0.1$.

4.3.1 Identifiability and Other Notes

When solving for this system, we have elected to also estimate the total at-risk population N. Due to this decision, we see that the profiles of the parameters display signs of practical identifiability problems (Figure 4.10). This is consistent with the findings in [33], where the SEIR model is found to be practically unidentifiable, despite its supposed structural identifiability (when $S(0)$ and N are known).

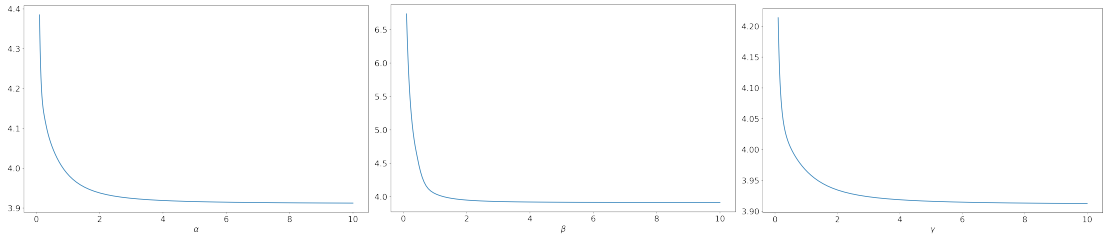


Figure 4.10: Profiles of parameters $\{\alpha, \beta, \gamma\}$ with $\rho = 0.005$ for the measles case.

In fact, if the at-risk population is not known, it can be shown that the model is actually also structurally non-identifiable. Using DAISY (Ritt's algorithm), the coefficients of the normalised input-output equation for the SEIR model are given in [33] as:

$$\frac{\beta}{N}, \qquad \alpha + \gamma, \qquad \frac{\alpha\beta + \beta\gamma}{N}, \qquad \frac{\alpha\beta\gamma}{N}$$

which can be reduced to show that β is free, with $N = \frac{N_q}{\beta_q}\beta$, alongside the previously known pair of solutions $(\alpha, \gamma) \in \{(\alpha_q, \gamma_q), (\gamma_q, \alpha_q)\}$. To resolve this non-identifiability, either the at-risk population must be known, or a value of $\mathcal{R}_0$ must be specified; alongside knowledge of α or γ. This means that our prior knowledge of the presentation of measles is not sufficient, even if it is strictly imposed. Unfortunately, the value of $\mathcal{R}_0$ is also not specified narrowly enough by previous outbreaks and studies [37] that we get a usable estimate for it.

However, there is a glimmer of hope. If we consider the final size relation from classical epidemiology, we find a surprising result. The final size relation gives the (theoretical) number of individuals that become infected over the course of the infection:

$$\log\left(\frac{S_\infty}{S_0}\right) = \mathcal{R}_0 \frac{S_\infty - N}{N} \tag{4.7}$$

$$R_\infty = N - S_\infty \tag{4.8}$$

where R_∞ is the final size, S_0 is the initial susceptible population, and $\mathcal{R}_0 \coloneqq \frac{\beta}{\alpha}$ is the basic reproduction number. We can see that the final epidemic size is relatively stable even as ρ is changed in Figure 4.11. This is despite the fact that the estimated size of the initial population is relatively volatile as ρ is varied. This may show that the final size of the epidemic is an identifiable parameter combination of the system.

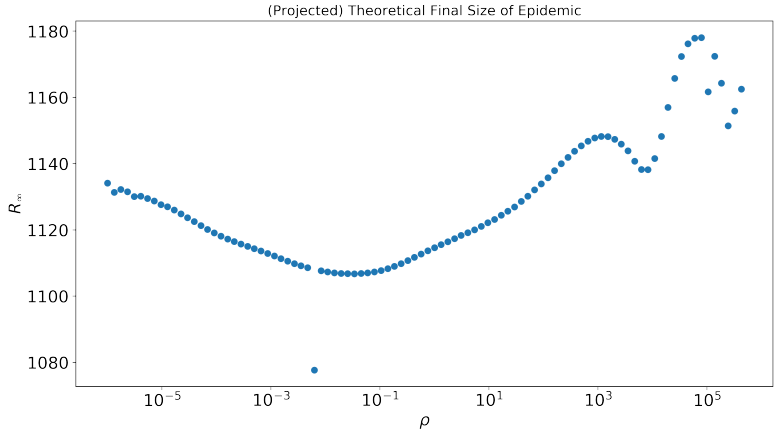


Figure 4.11: Final size of the measles epidemic as ρ is varied, as calculated from the final size relation.

5 Further Work

5.1 Control Strategy Efficacy

One of the curiosities of the 2019 Measles outbreak was the prior implementation of a vaccination campaign in the Waitemata and Auckland District Health Board (WDHB, ADHB) regions for a previous outbreak of mumps in 2017. A question that the two DHBs have is what effect the campaign had on the measles in the two regions [38], considering they did not experience the same intensity of outbreak as in the Counties Manukau DHB region. The ultimate goal is to use this information to inform cost-benefit analyses of similar vaccination strategies, augmenting analyses performed at a holistic level by [39]

Considering the difficulty in estimating the at-risk population, there is a high likelihood that there are still a significant number of susceptible individuals, particularly in the adolescent and young adult age group [40]. The concern is that outbreaks are not of sufficient size to affect the entire at-risk population. This makes the prediction of epidemic recurrence difficult, as outbreaks will be of potentially unknown size. This could also mark the importance of contact networks, as there may be sufficiently "disjoint" sub-networks such that the outbreak will exhaust one of them in one outbreak, but leave the rest susceptible to further outbreaks.

Other complicating mechanisms that could affect the modelling of this type of include the (external) sources of infection [41]. Particularly in Pasifika communities, the amount of non-endemic contact induced by familial visitation, for example, can introduce new sources of infection, or even create a pool of infected. An example of the latter is the recent Samoan measles epidemic that has resulted from contact with infected populations in New Zealand, which can later retransmit to the same population.

Previous modelling work has been done for the optimal timing of the MMR vaccine [42], [43], using a deterministic SIR model, and generating predictions using the R_0 and R_v (secondary cases with vaccination) constants, and a trajectory matching method. Due to the data availability and computational limits at the time, analysis based on age or ethnicity structure was not performed, which would be the aim of this project.

5.2 Extension to Multiple Dimensions

One clear extension we can make to the model is to introduce more complex differential operators than $\frac{d}{dt}$. This would mean an extension into spatial dimensions.

It is rather simple to construct a reaction-diffusion equation for fitting:

$$\frac{\partial u}{\partial t} = D\nabla^2 u + f(u,\theta) \iff \left(\frac{\partial}{\partial t} - D\nabla^2\right) u = f(u,\theta) \tag{5.1}$$

where D is an element of θ.

This can then be written in the standard generalised profiling form with

$$\mathcal{D}(\cdot) = \left(\frac{\partial}{\partial t} - D\nabla^2\right)(\cdot) \tag{5.2}$$

and utilising the standard tensor product definition of multi-dimensional B-splines [44]:

$$B(\mathbf{x}) = \prod_i B_i(x_i),\ \mathbf{x} = \{t, x_1, x_2, \dots\} \tag{5.3}$$

where B_i represents a univariate B-spline in the i-th dimension of $\mathbf{x}$, and B is the multidimensional spline.

In fact, this method (as well as a Bayesian near-equivalent) has already been investigated by Xun, Cao, Mallick, *et al.* [22] for a 1D, single-state, linear problem. However, the models used in epidemiology typically have nonlinear reaction terms, as well as being naturally 2D and has multiple states.

One avenue of exploration is the use of non-spline bases, for example, employing classical finite element techniques. Indeed, this has been done for a homogenous anisotropic diffusion model in [45] to model rainfall. The generalisation to reaction-diffusion equations for epidemiological modelling would mean that the linear properties of the problem disappear, and also inherit the identifiability problems seen in Section 3.2.

Another technique for spatial modelling in the epidemiological literature is the use of metapopulation models, where individuals are also categorised into spatial compartments in addition to infection compartments [46], [47]. This allows for the spatio-temporal system to be written as a purely temporal differential equation. This form of model is a likely candidate for the model to be used in the control efficacy study above (Section 5.1).

5.3 Bayesian Framework Interpretation

We can motivate a comparison between the frequentist approach of generalised profiling and Bayesian inference by making some typical assumptions about the error structures in the model.

Consider the observation model to be:

$$y = g(x) + e_o,\ e_o \sim \mathcal{N}(0, \sigma_o^2 I) \tag{5.4}$$

i.e. there is Gaussian, iid additive noise. Then we can say that

$$\pi(y|x) = \pi_{e_o}(y - g(x)) = \mathcal{N}(g(x), \sigma_o^2 I) \tag{5.5}$$

by integrating the observation model over e_o.

Next, we motivate a heuristic representation of process error, ignoring the nuances of stochastic differential equation theory.

$$\mathcal{D}x = f(x;\theta) + e_p,\ e_p \sim \mathcal{N}(0, \sigma_p^2 I), \tag{5.6}$$

which, after considering a discrete-time formulation, gives rise to

$$\pi(x|\theta) \propto \prod_{\tau=t_0}^{t} \left(\pi_{e_p}(\mathcal{D}x(\tau) - f(x(\tau);\theta))\right)^{d\tau}, \tag{5.7}$$

using product integral notation. A more detailed derivation is presented in Appendix A.

Recall that the multivariate normal distribution has the density function

$$x \sim \mathcal{N}(\mu, \Gamma) = \frac{1}{\sqrt{(2\pi)^k |\Gamma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Gamma^{-1}(x-\mu)\right) \tag{5.8}$$

for $x, \mu \in \mathbb{R}^k$.

Then, taking the log of the posterior gives

$$\begin{aligned} \log \pi(\theta|y) &= \log\left(\pi(y|x)\pi(x|\theta)\pi(\theta)\right) && (5.9)\\ &= \log \pi(y|x) + \log \pi(x|\theta) + \log \pi(\theta) \\ &= -\frac{1}{2}\left[\sigma_o^{-2}\|y-x\|^2 + \sigma_p^{-2}\|\mathcal{D}x - f(x;\theta)\|^2 + \sigma_\theta^{-2}\|\theta-\theta_0\|^2\right] + const. && (5.10) \end{aligned}$$

where we have introduced the prior $\pi(\theta) \sim \mathcal{N}(\theta_0, \sigma_\theta^2 I)$.

An estimator of the posterior distribution is the *maximum a posteriori* (MAP) estimate, which maximises the posterior probability:

$$\begin{aligned} \theta_{MAP} &= \arg\max_\theta \{\pi(\theta|y)\} && (5.11)\\ &= \arg\max_\theta \{\log \pi(\theta|y)\} \\ &= \arg\min_\theta \{-2\log \pi(\theta|y)\} \\ &= \arg\min_\theta \left\{\sigma_o^{-2}\|y-x\|^2 + \sigma_p^{-2}\|\mathcal{D}x - f(x;\theta)\|^2 + \sigma_\theta^{-2}\|\theta-\theta_0\|^2\right\} && (5.12) \end{aligned}$$

recovering the generalised profiling objective function, if we consider the projection of x onto a basis such that $x = \Phi c$. Thus, we see that the generalised profiling method can be thought of as an estimator of a specific case of a much more general error model.

This raises questions about the properties of the estimator if different error structures are used. For example, if we use the Poisson observation model (like in [16]), we can derive a different likelihood, and thus log-likelihood function. Does this allow for different objective function structures to be equally valid estimators? And how does the frequentist approximations of identifiability and confidence breakdown, as compared to the distributions generated by Bayesian methods? Can optimisation methods be used to make Bayesian methods more efficient?

This final question is of particular interest, when considering the inherent computational tradeoff of using a Bayesian method. With optimisation methods, we traverse the posterior distribution in a particular direction to converge onto the MAP estimate. However, to quantify the uncertainty in a Bayesian sampling framework, we are interested in the density of samples in a large area around the MAP estimate. This requires a larger number of samples, and due to needing them to be representative of the posterior, this also means a large number of samples generated that are rejected by the algorithm of choice. Some of these problems can be mitigated by using optimisation methods to get an idea of the behaviour of the posterior around its maxima — its location and local behaviour are represented by the MAP estimate and the profile likelihood respectively. An alternate use of optimisation could be to generate informative priors in order to better regularise the problem. Of course, there is also the obvious benefit of using optimisation methods to detect non-identifiability before attempting the expensive sampling procedures.

6 Resources

6.1 Facilities

It is expected that the bulk of the work for this thesis will be computational in nature. For this, the sufficient resources would be a computer or network of computers able to run code in a reasonable time horizon. The bulk of implementation is expected to be in Python, with dabblings in other languages as required by other tools and plugins. There is potential for the use of a compiled language to accelerate computational speed. Regardless, standard computational resources should be sufficient.

It is expected that for some portions of the project, sensitive healthcare data will need to handled and analysed. For this, it is expected ethics approval to be required. Up to the current point in time, this has not been a requirement, since data has been sourced from publicly accessible databases.

In terms of practical experimentation, it would be impractical, if not wholly unethical, to run realisations of the models. Research relationships with labs and projects more focused on (micro-)biological aspects of infection will be sought. We are currently in contact with epidemiologists in the School of Public Health and representatives from the Auckland and Waitemata DHBs with regard to data identification and acquisition as well as providing modelling guidance.

6.2 Budget

It is expected that the bulk of non-conference-related expenditure for this project would come from data acquisition, and any outsourced computational power. We do not have firm beliefs on expected costs for these, as we are not pursuing these avenues as of yet.

The budget for this project should be covered by the PRESS allocation.

7 Deliverables

By End of November 2019

Completed by the time of provisional candidate assessment.

1. Implementation of the generalised profiling optimisation method for ODEs
2. Parameter estimation for a simple epidemic model (measles) with synthetic and real data
3. Literature review on epidemiological modelling and parameter estimation.

7.1 In the Next Year

By May 2020

1. Validation study of the method on a synthetic model
2. Implementation and release of a general package for generalised profiling and identifiability analysis of ODEs

By November 2020

1. Analysis of effect of mumps control programme on measles outbreak
2. Basic implementation of generalised profiling methods for PDEs
3. Preliminary analysis of the spatial spread of measles

Expected Outputs and Events

1. Conference: ANZIAM 2020 (February)
2. Conference: SMB 2020 (September)
3. Output: Submission of article on generalised profiling/identifiability for measles

7.2 Within 2 Years

By End of May 2021

1. Validation experiments for the PDE methods
2. Implementation of generalised profiling in a Bayesian sampling method
3. Implementation of generalised profiling as a pre-analysis for Bayesian methods

By End of November 2021

1. Packaged generalised profiling method for PDEs
2. Comparison of uncertainty estimates between Bayesian and frequentist implementations of generalised profiling

References

[1] L. Rass and J. Radcliffe, *Spatial deterministic epidemics*. American Mathematical Society, 2003, p. 261, ISBN: 9780821804995.

[2] V. Capasso, *Mathematical Structures of Epidemic Systems*, ser. Lecture Notes in Biomathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, vol. 97, ISBN: 978-3-540-56526-0. DOI: 10.1007/978-3-540-70514-7.

[3] O. Diekmann, *Mathematical epidemiology of infectious diseases : model building, analysis, and interpretation*, eng, ser. Wiley series in mathematical and computational biology. Chichester ; New York: John Wiley, 2000, ISBN: 0471492418.

[4] W. O. Kermack and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 115, no. 772, pp. 700–721, Aug. 1927, ISSN: 1364-5021. DOI: 10.1098/rspa.1927.0118.

[5] G. Chowell, "Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts," *Infectious Disease Modelling*, vol. 2, no. 3, pp. 379–398, Aug. 2017, ISSN: 2468-0427. DOI: 10.1016/J.IDM.2017.08.001.

[6] J. P. D'Silva and M. C. Eisenberg, "Modeling spatial invasion of Ebola in West Africa," *Journal of Theoretical Biology*, vol. 428, pp. 65–75, Sep. 2017. DOI: 10.1016/J.JTBI.2017.05.034.

[7] A. Smirnova, L. DeCamp, and H. Liu, "Inverse Problems and Ebola Virus Disease Using an Age of Infection Model," in *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases*, Cham: Springer International Publishing, 2016, pp. 103–121. DOI: 10.1007/978-3-319-40413-4_8.

[8] A. Smirnova and G. Chowell, "A primer on stable parameter estimation and forecasting in epidemiology by a problem-oriented regularized least squares algorithm," *Infectious Disease Modelling*, vol. 2, no. 2, pp. 268–275, May 2017, ISSN: 2468-0427. DOI: 10.1016/J.IDM.2017.05.004.

[9] J. Ramsay and G. Hooker, *Dynamic Data Analysis*, ser. Springer Series in Statistics. New York, NY: Springer New York, 2017, ISBN: 978-1-4939-7188-6. DOI: 10.1007/978-1-4939-7190-9.

[10] I. Dattner, "A model-based initial guess for estimating parameters in systems of ordinary differential equations," *Biometrics*, vol. 71, no. 4, pp. 1176–1184, Dec. 2015, ISSN: 0006341X. DOI: 10.1111/biom.12348.

[11] A. Gábor and J. R. Banga, "Robust and efficient parameter estimation in dynamic models of biological systems.," *BMC systems biology*, vol. 9, p. 74, Oct. 2015, ISSN: 1752-0509. DOI: 10.1186/s12918-015-0219-2.

[12] N. Goeyvaerts, L. Willem, K. Van Kerckhove, *et al.*, "Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season-specific influenza-like illness incidence," *Epidemics*, vol. 13, pp. 1–9, Dec. 2015, ISSN: 1755-4365. DOI: 10.1016/J.EPIDEM.2015.04.002.

[13] J. Brynjarsdóttir and A. O'Hagan, "Learning about physical parameters: the importance of model discrepancy," *Inverse Problems*, vol. 30, no. 11, p. 114 007, Nov. 2014, ISSN: 0266-5611. DOI: 10.1088/0266-5611/30/11/114007.

[14] R. E. Morrison, T. A. Oliver, and R. D. Moser, "Representing Model Inadequacy: A Stochastic Operator Approach," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 2, pp. 457–496, Jan. 2018, ISSN: 2166-2525. DOI: 10.1137/16M1106419.

[15] E. Haber, U. M. Ascher, and D. W. Oldenburg, "Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach," *GEOPHYSICS*, vol. 69, no. 5, pp. 1216–1228, Sep. 2004, ISSN: 0016-8033. DOI: 10.1190/1.1801938.

[16] A. Chatzilena, E. van Leeuwen, O. Ratmann, *et al.*, "Contemporary statistical inference for infectious disease models using Stan," Aug. 2019. arXiv: 1903.00423.

[17] D. Wu, "Parametrisation and Identification of Epidemic Models: Literature Review," unpublished, 2019.

[18] J. O. Ramsay, G. Hooker, D. Campbell, *et al.*, "Parameter estimation for differential equations: a generalized smoothing approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 5, pp. 741–796, Nov. 2007, ISSN: 13697412. DOI: 10.1111/j.1467-9868.2007.00610.x.

[19] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, ser. Springer Series in Statistics. New York, NY: Springer New York, 2005, ISBN: 978-0-387-40080-8. DOI: 10.1007/b98888.

[20] G. Hooker, S. P. Ellner, L. D. V. Roditi, *et al.*, "Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario.," *Journal of the Royal Society, Interface*, vol. 8, no. 60, pp. 961–74, Jul. 2011, ISSN: 1742-5662. DOI: 10.1098/rsif.2010.0412.

[21] D. Campbell and O. Chkrebtii, "Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates," *Mathematical Biosciences*, vol. 246, no. 2, pp. 283–292, Dec. 2013, ISSN: 0025-5564. DOI: 10.1016/J.MBS.2013.03.011.

[22] X. Xun, J. Cao, B. Mallick, *et al.*, "Parameter Estimation of Partial Differential Equation Models," *Journal of the American Statistical Association*, vol. 108, no. 503, pp. 1009–1020, Sep. 2013, ISSN: 0162-1459. DOI: 10.1080/01621459.2013.794730.

[23] T. Raus and U. Hämarik, "About the balancing principle for choice of the regularization parameter," *Journal of Physics: Conference Series*, vol. 135, no. 1, p. 012 087, Nov. 2008, ISSN: 1742-6596. DOI: 10.1088/1742-6596/135/1/012087.

[24] P. C. Hansen, "The L-curve and its Use in the Numerical Treatment of Inverse Problems," in *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*, WIT Press, 2000, pp. 119–142.

[25] R. R. Picard and R. D. Cook, "Cross-Validation of Regression Models," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 575–583, Sep. 1984, ISSN: 0162-1459. DOI: 10.1080/01621459.1984.10478083.

[26] P. Mathé, "The Lepskii principle revisited," *Inverse Problems*, vol. 22, no. 3, pp. L11–L15, Jun. 2006, ISSN: 0266-5611. DOI: 10.1088/0266-5611/22/3/L02.

[27] M. Belge, M. E. Kilmer, and E. L. Miller, "Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet transform domain," A. Mohammad-Djafari, Ed., vol. 3459, International Society for Optics and Photonics, Sep. 1998, pp. 328–336. DOI: 10.1117/12.323812.

[28] ——, "Efficient determination of multiple regularization parameters in a generalized L-curve framework," *Inverse Problems*, vol. 18, no. 4, p. 314, Aug. 2002, ISSN: 02665611. DOI: 10.1088/0266-5611/18/4/314.

[29] G. Bellu, M. P. Saccomani, S. Audoly, *et al.*, "DAISY: A new software tool to test global identifiability of biological and physiological systems," *Computer Methods and Programs in Biomedicine*, vol. 88, no. 1, pp. 52–61, Oct. 2007, ISSN: 01692607. DOI: 10.1016/j.cmpb.2007.07.002.

[30] J. Karlsson, M. Anguelova, and M. Jirstrand, "An Efficient Method for Structural Identifiability Analysis of Large Dynamic Systems," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 941–946, Jul. 2012, ISSN: 1474-6670. DOI: 10.3182/20120711-3-BE-2027.00381.

[31] A. Raue, C. Kreutz, T. Maiwald, *et al.*, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, Aug. 2009, ISSN: 1460-2059. DOI: 10.1093/bioinformatics/btp358.

[32] N. Tuncer, H. Gulbudak, V. L. Cannataro, *et al.*, "Structural and Practical Identifiability Issues of Immuno-Epidemiological Vector-Host Models with Application to Rift Valley Fever," *Bulletin of Mathematical Biology*, vol. 78, no. 9, pp. 1796–1827, Sep. 2016, ISSN: 0092-8240. DOI: 10.1007/s11538-016-0200-2.

[33] N. Tuncer and T. T. Le, "Structural and practical identifiability analysis of outbreak models," *Mathematical Biosciences*, vol. 299, pp. 1–18, May 2018, ISSN: 0025-5564. DOI: 10.1016/J.MBS.2018.02.004.

[34] F. Fröhlich, F. J. Theis, and J. Hasenauer, "Uncertainty Analysis for Non-identifiable Dynamical Systems: Profile Likelihoods, Bootstrapping and More," in Springer, Cham, Nov. 2014, pp. 61–72. DOI: 10.1007/978-3-319-12982-2_5.

[35] J. A. E. Andersson, J. Gillis, G. Horn, *et al.*, "CasADi – A software framework for nonlinear optimization and optimal control," *Mathematical Programming Computation*, vol. 11, no. 1, pp. 1–36, 2019. DOI: 10.1007/s12532-018-0139-4.

[36] ESR. "Public Health Surveillance - Measles Report." (2019), [Online]. Available: https://surv.esr.cri.nz/surveillance/WeeklyMeaslesRpt.php (visited on 10/16/2019).

[37] F. M. Guerra, S. Bolotin, G. Lim, *et al.*, "The basic reproduction number (R0) of measles: a systematic review.," *The Lancet. Infectious diseases*, vol. 17, no. 12, e420–e428, Dec. 2017, ISSN: 1474-4457. DOI: 10.1016/S1473-3099(17)30307-9.

[38] J. McQueen, C. Gillmer, G. Tucker, *et al.*, *Mumps outbreak immunisation programme, "Community Immunity" in Auckland and Waitemata DHBs*, 2019.

[39] D. Hayman, J. Marshall, N. French, *et al.*, "Cost-benefit analyses of supplementary measles immunisation in the highly immunized population of New Zealand," *Vaccine*, vol. 35, no. 37, pp. 4913–4922, Sep. 2017, ISSN: 0264-410X. DOI: 10.1016/J.VACCINE.2017.07.077.

[40] G. Reynolds, C. Dias, S. Thornley, *et al.*, "Analysis of the Auckland 2014 measles outbreak indicates that adolescents and young adults could benefit from catch-up vaccination," *New Zealand Medical Journal*, vol. 128, no. 1422, pp. 53–62, 2015, ISSN: 11758716.

[41] D. T. S. Hayman, J. C. Marshall, N. P. French, *et al.*, "Global importation and population risk factors for measles in New Zealand: a case study for highly immunized populations," *Epidemiology and Infection*, vol. 145, no. 9, pp. 1875–1885, Jul. 2017, ISSN: 0950-2688. DOI: 10.1017/S0950268817000723.

[42] M. G. Roberts and M. I. Tobias, "Predicting and preventing measles epidemics in New Zealand: application of a mathematical model," *Epidemiology and Infection*, vol. 124, no. 2, pp. 279–287, Apr. 2000, ISSN: 0950-2688. DOI: 10.1017/S0950268899003556.

[43] M. Tobias and M. Roberts, "Modelling Measles. Predicting and Preventing Measles Epidemics in New Zealand: Application of a mathematical model," Ministry of Health, Wellington, Tech. Rep., 1998.

[44] K. Höllig and J. Hörner, *Approximation and modeling with B-splines*. Philadelphia: SIAM, 2013, p. 214, ISBN: 9781611972948.

[45] M. S. Bernardi, M. Carey, J. O. Ramsay, *et al.*, "Modeling spatial anisotropy via regression with partial differential regularization," *Journal of Multivariate Analysis*, vol. 167, pp. 15–30, Sep. 2018, ISSN: 0047-259X. DOI: 10.1016/J.JMVA.2018.03.014.

[46] J. Arino and P. Van Den Driessche, "A multi-city epidemic model," *Mathematical Population Studies*, vol. 10, no. 3, pp. 175–193, Jan. 2003, ISSN: 08898480. DOI: 10.1080/08898480306720.

[47] J. M. Hyman and T. Laforce, "Modelling the Spread of Influenza among Cities," in *Bioterrorism : mathematical modelling applications to homeland security*, H. T. Banks and C. Castillo-Chavez, Eds., Philadelphia: Society for Industrial and Applied Mathematics, 2003, ch. 10, pp. 211–236, ISBN: 0-89871-549-0.

A Derivation of Process Model Likelihood

We aim to derive an expression for $\pi(x|\theta)$. We have the heuristic process model[1] from eq. (5.6)

$$\mathcal{D}x(t) = f(x(t);\theta) + e_p(t),\ e_p(t) \sim \mathcal{N}(0, \sigma_p^2) \tag{A.1}$$

for the differential operator $\mathcal{D}$, state vector $x \in \mathbb{R}^k$, parameter vector $\theta \in \mathbb{R}^p$, and some function $f : \mathbb{R}^k \times \mathbb{R}^p \mapsto \mathbb{R}^k$.

We approximate the differential operator, for $\mathcal{D} = \frac{d}{dt}$, as a discrete-time state space model.

$$\mathcal{D}x(t) = \frac{dx}{dt}(t) \approx \frac{x(t) - x(t - \Delta t)}{\Delta t}.$$

Thus we can write

$$x(t) = x(t - \Delta t) + f(x(t - \Delta t);\theta)\Delta t + e_p(t)\Delta t, \tag{A.2}$$

giving

$$\pi(x(t)|x(t - \Delta t), \theta, e_p(t)) = \delta\left(x(t) - x(t - \Delta t) - f(x(t - \Delta t);\theta)\Delta t - e_p(t)\Delta t\right). \tag{A.3}$$

where δ is the Dirac delta function.

Using the standard consideration of $\pi(x(t), x(t - \Delta t), \theta, e_p)$ we get

$$\pi(x(t), e_p|x(t - \Delta t)\theta) = \pi(x(t)|x(t - \Delta t), \theta, e_p)\pi(e_p), \tag{A.4}$$

which we marginalise over e_p to get the transition probability

$$\begin{aligned}
\pi(x(t)|x(t - \Delta t), \theta) &= \int \pi(x(t), e_p|x(t - \Delta t)\theta)de_p \\
&= \int \pi(x(t)|x(t - \Delta t), \theta, e_p)\pi(e_p)de_p \\
&= \int \delta(x(t) - x(t - \Delta t) - f(x(t - \Delta t);\theta)\Delta t - \Delta t e_p(t))\pi(e_p)de_p \\
&= \pi_{e_p}\left(\frac{x(t) - x(t - \Delta t)}{\Delta t} - f(x(t - \Delta t);\theta)\right).
\end{aligned} \tag{A.5}$$

Then to get the likelihood of the entire process, we have

$$\pi(x(t)|\theta) = \pi(x(t_0)|\theta) \prod_{\tau=t_0+\Delta\tau}^{t} \pi(x(\tau)|x(\tau - \Delta\tau), \theta). \tag{A.6}$$

[1] A more rigorous construction would be to consider a properly stochastic differential equation, and use the related theory to construct the likelihood

Substituting gives

$$\pi(x(t)|\theta) = \pi(x(t_0)|\theta) \prod_{\tau=t_0+\Delta\tau}^{t} \pi_{e_p}\left(\frac{x(\tau) - x(\tau-\Delta\tau)}{\Delta\tau} - f(x(\tau-\Delta\tau);\theta)\right). \tag{A.7}$$

which at the limit as $\Delta\tau \to 0$ is

$$\begin{aligned}\pi(x(t)|\theta) &= \pi(x(t_0)|\theta) \prod_{\tau=t_0}^{t} \left(\pi_{e_p}(\mathcal{D}x(\tau) - f(x(\tau);\theta))\right)^{d\tau} \\ &= \pi(x(t_0)|\theta) \exp\left(\int_{\tau=t_0}^{t} \log\left(\pi_{e_p}(\mathcal{D}x(\tau) - f(x(\tau);\theta))\right) d\tau\right).\end{aligned} \tag{A.8}$$

utilising the product integral notation.

Absorbing $\pi(x_0|\theta)$ into the prior, and taking the logarithm of the remaining product gives

$$\begin{aligned}\log \pi(x(t)|\theta) &= \int_{\tau=t_0}^{t} \log\left(\pi_{e_p}(\mathcal{D}x - f(x;\theta))\right) d\tau \\ &= \int_{\tau=t_0}^{t} \sigma_p^{-2}(\mathcal{D}(x(\tau)) - f(x(\tau);\theta))^2 d\tau + (t-t_0)const.\end{aligned} \tag{A.9}$$

Computing the integral at finite collocation points over a finite time domain gives

$$\log \pi(x(t)|\theta) = \sigma_p^{-2} \left\|\mathcal{D}(x(t) - f(x(t);\theta))\right\|^2 + const. \tag{A.10}$$