

# Likelihood-based estimation and prediction for misspecified epidemic models: an application to measles in Samoa

David Wu <sup>\*1</sup>, Helen Petousis-Harris<sup>2</sup>, Janine Paynter<sup>2</sup>, Vinod Suresh<sup>1,3</sup>, and Oliver J. Maclaren<sup>1</sup>

<sup>1</sup>*Department of Engineering Science, University of Auckland, New Zealand*

<sup>2</sup>*Department of General Practice and Primary Health Care, University of Auckland, New Zealand*

<sup>3</sup>*Auckland Bioengineering Institute, University of Auckland, New Zealand*

---

Prediction of the progression of an infectious disease outbreak in a population is an important task. Differential equations are often used to model an epidemic outbreak's behaviour but are challenging to parametrise. Furthermore, these models can suffer from misspecification, which biases the estimates. Stochastic models can help with misspecification but are even more expensive to simulate and perform inference with. Here, we develop an explicitly likelihood-based variation of the generalised profiling method as a tool for prediction and inference under model misspecification. Our approach allows us to carry out identifiability analysis and uncertainty quantification using profile likelihood-based methods without the need for marginalisation. We provide additional justification for this approach by introducing a new interpretation of the model approximation component as a stochastic constraint. This interpretation preserves the rationale for using profiling rather than integration to remove nuisance parameters while still providing a link back to stochastic models. We applied an initial version of this method during an outbreak of measles in Samoa in 2019-2020 and found that we achieved relatively fast, accurate predictions. Here present the most recent version of our method and its application to this measles outbreak, along with additional validation.

---

## 1 Introduction

The prediction of the progression of epidemic outbreaks of infectious diseases is notoriously difficult, with multiple confounding sources of uncertainty. Despite this, even approximate predictions are important in informing the response to outbreaks. This importance is emphasised in areas where resources restrict response infrastructure or rapid response is required. Early predictions are typically model-based, as the available data alone are usually insufficient for understanding the key dynamics of an outbreak.

---

\*Corresponding Author. dwu402@aucklanduni.ac.nz

The models used to characterise the spread of infectious diseases are usually mechanistic models, and these are well-established in the mathematical epidemiological community [1]. From the early deterministic models of Ross [2] and Kermack and McKendrick [3], to recent developments in both deterministic [4–6] and stochastic models [7–9], there are a range of models that can be used to describe various aspects of the dynamics of infectious disease outbreaks. However, there are still difficult open problems on how best to relate these models to data and extract meaningful information about model behaviour specific to particular outbreaks [10, 11].

While there are many methods for performing parameter inference for mechanistic models, there remain well-known difficulties in performing inference on complex, nonlinear systems. From a user perspective, inference methods for such systems are typically computationally expensive or technically difficult to use [12]. Additionally, limitations in data collection, combined with complex model structures, may mean parameter estimates are not uniquely determinable, or *nonidentifiable* [13, 14]. There is also the subtle problem of *model misspecification* [15–17]. Misspecification is the phenomenon of discrepancies between the model and the true data generating process. It can cause bias in inference and inaccuracies in forecasting if not considered carefully.

Many recent advances in inference methods incorporate uncertainty from misspecification into the modelling process as an additional stochastic process, under both likelihood/frequentist frameworks [17] and Bayesian frameworks [18]. However, this compounds the computational expense of the inference problem, due to the increased complexity. The core of the expense is the additional marginalisation step (over the stochastic process model) that this introduces to the calculation of the likelihood function. Even with state-of-the-art methods, these stochastic methods can often take around five times longer than deterministic counterparts [18].

An alternative is to only approximately solve or enforce a deterministic model. This approach allows for violations of the deterministic model without entirely sacrificing the model-based information required for forecasting. This can help prevent large prediction errors from strict enforcement of a misspecified model. Only approximately solving the equations also reduces the computational expense, which is useful in forecasting where results are time-sensitive and resources limited.

We follow this approach here and utilise an existing statistical method based on this idea, called *generalised profiling* [19]. We introduce a slight variant of this approach under a new statistical interpretation and describe how this assists in interpreting and addressing nuisance parameters, model misspecification and uncertainty quantification. Our statistical interpretation differs from that of introducing an explicit stochastic process model; instead, we enforce the model as a stochastic constraint following ideas introduced in the statistics and econometrics literature [20, 21]. This can also be thought of as a *prior likelihood* [22] term. This provides a previously lacking explicit motivation for profiling (maximising) out nuisance parameters in the style of profile likelihood instead of marginalising them out.

The rest of the article is laid out as follows. In Section 2, we describe the general model fitting problem in more detail. In Section 3, we describe our methods, including a variant of generalised profiling. In Section 4, we apply the methods to a case study of a measles outbreak in Samoa, where the data is non-ideal and some classical methods break down. Finally in Section 5 we discuss benefits and limitations to the approach.

## 2 Background

Mechanistic models in biology are often deterministic dynamical systems, where the evolution of state of the system, denoted by  $x$ , is defined either in discrete time,  $x_{t+1} = F(x_t; \theta)$ , or more

commonly as a set of differential equations:

$$\frac{dx}{dt} = f(x; \theta), \quad (1)$$

where  $f$  is some vector field, and  $\theta$  its parameters. The state  $x$  itself may also be discrete or continuous-valued.

The traditional method of parameter estimation is based on minimising some measure of the distance  $d(y, x(t; \theta))$  between the data,  $y$ , and the parameter-dependent state of the model,  $x(t; \theta)$ , under some observation model  $g$ , for example:

$$d(y, x(t; \theta)) = \|y - g(x(t, \theta))\|_2^2. \quad (2)$$

The above corresponds to a typical ‘least squares’ formulation and has a natural (though not necessary) interpretation under the assumption of normally-distributed measurement errors. This formulation has a few difficulties, however [23]. Firstly, it may be computationally expensive to evaluate  $x$  since this is usually carried out by numerical integration, especially in the presence of stiffness (time-scale separation) or other related phenomena. Compounding this, estimating parameters by minimising this distance using iterative methods means numerical integration must usually be carried out repeatedly for different parameter values. Furthermore, many models have regions of parameter space where they exhibit unexpected or computationally expensive behaviour.

Secondly, this formulation implicitly assumes that the models  $f$  and  $g$  are “correct”, in the sense that the system that generates the data  $y$  is exactly described by  $f$  and  $g$ . This may not be true: the models may be *misspecified*. This causes problems with biased estimates of the parameters and state. One way to model this is to use an explicit stochastic process, typically a Gaussian process, to capture the misspecification [15]. This adds another stochastic term to the model, which is then used to account for discrepancies between the best-fitting model and the data. Similarly, stochasticity can be incorporated directly into the dynamics of the model. A typical continuous-time, continuous-state model is then a stochastic differential equation (SDE) of the form:

$$dx = f(x; \theta)dt + \sigma(x; \theta)dW_t, \quad (3)$$

where  $W_t$  represents a Wiener process [24, 25]. The error represented by the second term can be interpreted as either inherent stochasticity or as representing a form of misspecification due to missing influences, though these interpretations are subject to a number of subtleties [24].

Regardless of how the stochastic model is interpreted, inference on such models is significantly more expensive than the deterministic ODE system. The state  $x$  is typically unobserved, and the likelihood function  $p(y|\theta, \sigma)$  thus requires marginalising over this. Much of the recent literature on estimating the parameters of complex dynamic models hence aims at developing more efficient ways of marginalising over latent states to evaluate this form of the likelihood or some approximate replacement for it. These approaches usually either exploit the partially-observed Markov process/state-space formalism or use simulation-based inference [18, 26–33]. Discrete-time, and/or discrete-state, stochastic models are often used for modelling complex systems in biology and statistics [34, 35], and these introduce similar computational burdens.

## 3 Methods

### 3.1 Generalised Profiling

An alternative to explicitly representing the model misspecification is to weakly, or approximately, enforce the model. We do this by using the generalised profiling method introduced by Ramsay

et al. [19]. The generalised profiling method, also known as the parameter cascade method, was built on the methods of functional data analysis [36], in order to incorporate functional assumptions about the underlying data in the form of differential equations. We show that, under a stochastic constraint interpretation, this can also be used to approximately enforce a differential equation model and hence allow for misspecification.

### 3.1.1 Standard formulation

In its classical formulation [19, 23], the generalised profiling method is comprised of two nested optimisation problems, extending the two-stage nonlinear least squares algorithm. The inner problem is a smoothing problem, regularised by the model for some given fixed  $\theta$ .

$$\begin{aligned} \hat{c}(\theta) = \arg \min_{c|\theta} \quad & \|y - g(x)\|_2^2 + \lambda \int \left( \frac{dx}{dt} - f(x; \theta) \right)^2 dt, \\ \text{s.t.} \quad & x = \Phi c. \end{aligned} \tag{4}$$

where the state  $x$  is represented in terms of some basis comprising the columns of  $\Phi$ , which is typically a basis of B-splines. This inner problem is tuned with the hyperparameter  $\lambda$ , which acts as a tradeoff between interpolation (when  $\lambda$  is low) and model matching (when  $\lambda$  is high). This produces some optimal value for  $c$  for the given  $\theta$ , which we denote  $\hat{c}(\theta)$ . This value is then passed into an outer optimisation problem

$$\min_{\theta} \quad \|y - g(\Phi \hat{c}(\theta))\|_2^2, \tag{5}$$

which performs the parameter optimisation. Because the inner objective (Equation (4)) uses a spline representation to perform state estimation, we can avoid integration by collocating on a chosen set of points in the time domain, since the derivatives in the integral term can be directly acquired through the spline representation. This also allows for unobserved states to be dealt with, since they are also simultaneously estimated.

The generalised profiling approach is closely related to spline smoothing and other forms of non-/semi-parametric regression. In particular, the inner problem can be seen as a generalisation of standard smoothness penalties to differential equation penalties building on the functional data analysis literature [23, 36]. As discussed by Wahba [37], Ruppert et al. [38], the solution to such smoothing problems can often be interpreted as a form of Bayes estimate or as a predictive estimate for random-effects  $c$ . Similar points were raised in the discussion of [19]. Ramsay et al. [19], Ramsay and Hooker [23], however, stress that they do not view their estimates as arising from an explicit stochastic process model.

Also of note, both conceptually and computationally, is that the ‘nuisance’ parameters  $c$  are maximised out to give  $\hat{c}(\theta)$ , rather than averaged out (which would be expected for random effects). This act of replacing the nuisance parameters with their maximum likelihood estimates, for each value of the interest parameter, is known as *profiling* [39]. Hence, it appears that the outer objective is intended as the profile likelihood of the inner objective, with respect to  $\theta$ , though this is not made explicit in the original paper by Ramsay et al. [19]. We next consider a reformulation of the profiling approach which we feel more naturally unifies the inner and outer optimisation problems and, also leads to an explicit justification for profiling  $c$  out. We first motivate this in the linear case.

### 3.1.2 Reformulation: linear case

Consider the linear regression problem for data  $y$  of the form

$$y = G\Phi c + e, \quad (6)$$

where  $G$  is a linear observation operator,  $\Phi$  consists of a basis with a large number of basis functions ('columns' of  $\Phi$ ) relative to the number of observations ('rows' of  $G$ ), and  $e$  is observational noise with mean 0 and covariance  $\Gamma$ . As stated this problem typically requires regularisation of some form to constrain the  $c$ . Typical statistical interpretations of such regularisation are in terms of Bayesian models [37, 38] or, in non-Bayesian inference, in terms of random-effects models [38]. In both cases the  $c$  coefficients are taken to be random.

An alternative way to introduce a similar regularising effect based on external or 'prior' information with a statistical interpretation, but in which  $c$  is non-random, was introduced by Theil and Goldberger [21], building on the work of Durbin [20]. This amounts to considering, in addition to Equation (6), the additional linear regression equation for another observable random variable  $r$ :

$$r = H\Psi c + \nu. \quad (7)$$

Here  $H$  is a linear observation operator,  $\Psi$  consists of a potentially different set of basis functions, and  $\nu$  is observational noise with mean 0 and covariance  $\Sigma$ . This will be assumed to be independent of  $e$ , representing an independent source of information on  $c$ . In this context  $c$  is assumed fixed (non-random), while  $y$  and  $r$  are random. The Theil and Goldberger [21] procedure amounts to assuming we have observations of both random variables, i.e. observations of  $y$  and  $r$ , and carrying out a standard analysis given these observations. Here the observation of  $r$  plays the role that the prior or overall mean typically plays in Bayesian or random effects models. Theil and Goldberger [21] call this *mixed estimation*, though this is to be distinguished from mixed modelling in which  $c$  would typically be random. This is closely related to the concept of a *prior likelihood* introduced by Edwards [22], which is additional information in the form of a likelihood for a fixed parameter such as  $c$  based on real or hypothetical prior data on a random variable such as  $r$ . A recent treatment of the method of stochastic restrictions in the sense of Theil and Goldberger [21] can be found in Rao et al. [40]. H-likelihood [41, 42] is another related, though distinct, idea.

The above equations can be 'stacked' to give

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} G\Phi \\ H\Psi \end{bmatrix} c + \begin{bmatrix} e \\ \nu \end{bmatrix} \quad (8)$$

This can be solved using the generalised least squares method, using the notation  $\|f\|_A^2 = f^T A f$

$$\min_c \|y - G\Phi c\|_{\Gamma^{-1}}^2 + \|r - H\Psi c\|_{\Sigma^{-1}}^2, \quad (9)$$

which has the usual explicit solution

$$\hat{c} = (\tilde{G}^T \tilde{G})^{-1} \tilde{G}^T \tilde{y}, \quad (10)$$

where  $\tilde{G} = \begin{bmatrix} LG\Phi \\ MH\Psi \end{bmatrix}$ ,  $\tilde{y} = \begin{bmatrix} Ly \\ Mr \end{bmatrix}$ , where  $L$  and  $M$  are whitening matrices of  $\Gamma$  and  $\Sigma$  respectively ( $L^T L = \Gamma^{-1}$ ,  $M^T M = \Sigma^{-1}$ ), and  $\tilde{G}^T \tilde{G}$  is assumed to be invertible.

We now consider the case where the auxiliary information of Equation (7) takes the form of a linear differential equation. We first write this in operator form in the exactly specified case as

$$r = L(\theta)x \quad (11)$$

where  $L(\theta) = \mathcal{D} - A(\theta)$  represents the autonomous system  $\mathcal{D}x = f(x; \theta) = A(\theta)x$ , where  $A(\theta)$  is a linear operator acting on  $x$  that depends on parameters  $\theta$ , and  $r$  represents external forcing terms not depending on  $x$ . We have written  $r$  on the left-hand side to mimic Equation (7). If we instead only enforce this as a *stochastic constraint*, for stochastic  $r$ , analogously to Equation (7), we obtain the least squares problem

$$\min_{c|\theta} \|y - G\Phi c\|_{\Gamma^{-1}}^2 + \int (m(r - L(\theta)\Phi c))^2 dt \quad (12)$$

where, in comparison to the regression formulation Equation (9), we have taken  $\Psi = L\Phi = (\mathcal{D} - A)\Phi$  and have enforced the stochastic constraint in the continuous limit with weighting function  $m(t)$  (this is solved discretely in practice, however). The notation  $\min_{c|\theta}$  indicates that the above is a least-squares problem for  $c$  given each choice of  $\theta$ .

The infinite-dimensional nature of the differential equation constraint requires care. A stochastic interpretation of this infinite-dimensional constraint can be given by interpreting it as enforcing Equation (11) stochastically, i.e. as enforcing

$$r = L(\theta)x + \Sigma DW, \quad (13)$$

where  $DW$  stands for the (formal) time derivative of a multi-dimensional Wiener process [25]. To ensure a consistent interpretation, when this is implemented discretely as a set of regression equations in terms of ‘observations’ of the model, we use an Euler-Maruyama discretisation scheme [43] for the derivatives. This amounts to taking, for the  $i^{\text{th}}$  ‘observation’ (discretisation grid point) of the equation,

$$\begin{aligned} (Dx)_i &\approx \frac{x_{i+1} - x_i}{\Delta t} \\ (DW)_i &\approx \frac{1}{\sqrt{\Delta t}} e_i, \quad e_i \sim \mathcal{N}(0, I). \end{aligned} \quad (14)$$

Multiplying this through by  $\Delta t$  to put it in differential form gives the standard Euler-Maruyama discretisation.

The objective Equation (12) reduces to the standard inner objective for generalised profiling of Equation (4) when the differential equation is linear and, in particular, when our *stochastic prior information consists of the ‘observation’*  $r = 0$ . We will assume this in general, which is analogous to taking the overall mean in the Bayes/random effects formulations as zero, but note that this is not a necessary assumption and could be modified to allow for (random observations of) non-zero forcing terms. Importantly, *maximising* out  $c$  for each  $\theta$  corresponds to *profiling* a standard joint likelihood function  $\mathcal{L}(\theta, c; y, r)$  under appropriate conditions.

### 3.1.3 Reformulation: general case and link to standard formulation

We now reformulate the general nonlinear problem in terms of a single overall objective,  $l(\theta, c)$ :

$$\begin{aligned} \min_{\theta, c} \quad & l(\theta, c) = \|y - g(x)\|_{\Gamma^{-1}}^2 + \int (m(Dx - f(x; \theta) - r))^2 dt \\ \text{s.t.} \quad & x = \Phi c \end{aligned} \quad (15)$$

We will in general take  $r = 0$  and hence drop this from now. As discussed above, we can interpret the second term using an Euler-Maruyama discretisation and write:

$$\begin{aligned}
\min_{\theta, c} \quad & l(\theta, c) = \|y - g(x)\|_{\Gamma^{-1}}^2 + \|Dx - f(x; \theta)\|_{\widehat{\Sigma}^{-1}} \\
\text{s.t.} \quad & x = \Phi c
\end{aligned} \tag{16}$$

where  $\widehat{\Sigma} = \Sigma \Delta t$ . This is not quite the final form of our objective: we consider additional covariance terms in more detail in the following subsection. First we relate the above version to the standard generalised profiling approach, however.

As indicated above, this formulation effectively replaces the objective of the outer optimisation problem in Equation (5) by a *profile* of the overall objective over a vector ‘nuisance’ parameter  $c$ , in the same sense as profiling (negative log-)likelihoods [39]. For example, denoting the overall objective by  $l(\theta, c; \Gamma, \Sigma)$ , for fixed  $\Gamma, \Sigma$  and for the constraint  $x = \Phi c$  imposed, our problem can be decomposed into two analogous stages:

$$\min_{\theta} \min_{c|\theta} l(\theta, c; \Gamma, \Sigma). \tag{17}$$

While the inner optimisation step is the same here as before, our outer optimisation retains the model misfit term. By defining

$$l_p(\theta, \hat{c}(\theta); \Gamma, \Sigma) = \min_{c|\theta} l(\theta, c; \Gamma, \Sigma), \tag{18}$$

our outer optimisation then corresponds to

$$\min_{\theta} l_p(\theta, \hat{c}(\theta); \Gamma, \Sigma). \tag{19}$$

Existing literature using the generalised profiling method [5, 19, 44–46] appears to focus on parameter inference, and neglect the predictive properties of the method. Some work has been done in prediction in [5], but in the context of tuning the smoothing hyperparameter  $\lambda$  of the classical formulation. This was done by integrating the model exactly from a set of points chosen on the approximated state. This integration was never extended past the time horizon of the data, meaning that true prediction was never done. Further, it is noted that depending on the choice of  $\lambda$ , the integral of the model and the estimated state may not necessarily agree. Thus this raises the question of how to utilise and choose information gained from the inference process for prediction. This was also a key question raised by Ramsay et al. [19]. To address this, we introduce a natural method for performing prediction using our objective function. By extending the fitting time-window beyond that supplied by the data, the overall objective naturally performs extrapolation of the data as informed by the model. This is due to the model-misfit term, which imposes the trajectory’s behaviour further forward in time, while respecting the information from the data near the regions where it is available. Practically, this means that the prediction process can be done as part of the inference process, and no further decisions need to be made in order to predict. This is another benefit to viewing the inner and outer optimisation problems as simply different profiles of the same overall objective.

#### 3.1.4 Covariance Estimation and Iterative Solution Process

In the classical formulation, the tuning of the hyperparameter  $\lambda$  can be done efficiently in the linear case, through the application of the generalised cross-validation criterion [37], among other, typically prediction-oriented, data-driven methods. However, when the model  $f$  becomes nonlinear, then the approaches are typically limited to grid search methods [23, 45].

In our reformulation, the estimation of the covariances of the errors are analogous to the tuning of  $\lambda$ . Hence, unlike the standard least squares form of generalised profiling, here covariance-related

terms in the negative log-likelihood are not dropped from our final form of the objective. Following above assumptions about the errors we have in the model, we get for Gaussian errors the following negative log-likelihood objective function.

$$l(\theta, x) = \|y - g(x)\|_{\Gamma^{-1}}^2 + \|Dx - f(x; \theta)\|_{\hat{\Sigma}^{-1}}^2 - 2 \log |\Gamma| - 2 \log |\hat{\Sigma}| \quad (20)$$

However, the concurrent estimation of both the parameters and covariances is difficult due to computational issues with convergence, as well as robustness concerns. Such problems are typically resolved by using iterative methods, such as generalised least squares methods [47], which alternate between estimation of the parameters and covariances. These are early-stopping versions of iteratively-reweighted least squares [47]. The number of iterations  $N$  to take are hard to mathematically justify in most cases, and empirical experiments done by Carroll and Ruppert [47] show that  $N \geq 2$  is advised, but determining an optimal number of iterations remains an open problem and is likely problem-dependent. We also note that this iterative procedure can be used in the form of quasi- or pseudo-likelihood estimation where, for example, the estimates of the mean (state) function parameters are based on quasi-likelihood/generalised least squares procedures, while estimates of the covariances use (typically normal-theory) maximum likelihood or related methods [38, 47]. This allows the use of non-normal likelihood expressions, at least for estimating the parameters of the mean (state) function. The iterative process used is given in Algorithm 1.

---

**Algorithm 1:** Iteratively reweighted least squares algorithm

---

**Data:** initial iterate  $x_0$ , initial weights  $w_0$ , negative log-likelihood function  $l(x, w)$ , maximum iterations  $N$ , error threshold  $\epsilon$

**Result:** optimal values  $\hat{x}$ , optimal weights  $\hat{w}$

```

1 begin
2   initialization
3    $f_0 \leftarrow l(x_0, w_0)$ 
4   for  $i = 1$  to  $N$  do
5      $x_i \leftarrow \arg \min_x l(x, w_{i-1})$ 
6      $w_i \leftarrow \arg \min_w l(x_i, w)$ 
7      $\Delta f_i \leftarrow f_i - f_{i-1}$ 
8     if  $\Delta f_i < -\epsilon$  then
9       // divergence
10       $\hat{x} \leftarrow x_{i-1}$ 
11       $\hat{w} \leftarrow w_{i-1}$ 
12      return
13    $\hat{x} \leftarrow x_i$ 
14    $\hat{w} \leftarrow w_i$ 

```

---

### 3.2 Uncertainty Quantification

Next, we consider statistical uncertainty quantification, in the sense of forming appropriate uncertainty intervals with (approximate) coverage properties.

In the classical approach, the intervals can be extracted by treating the coefficients  $c : x = \Phi c$  as nuisance parameters that depend on the model parameters  $\theta$ . The outer objective function then allows the uncertainty to be propagated from the data space  $y$  to parameter space, via



a linearisation. This leads to Wald-style confidence intervals; however, if the underlying log-likelihood is not well approximated by a quadratic, then this approach can give misleading results [39].

An alternative way to perform analysis on the uncertainty of quantities of interest is using the profile likelihood [32, 39]. We can apply typical frequentist sampling theory methods to form likelihood-based confidence intervals for particular quantities of interest from their profile likelihoods. For a scalar quantity of interest, this is typically done by choosing a cutoff point  $k$  such that

$$k = \exp \left( -\frac{1}{2} \chi_{1,(1-a)}^2 \right)$$

where  $\chi_{1,(1-a)}^2$  is the  $(1-a)$ -th percentile of the chi-squared distribution with 1 degree of freedom. This cutoff then defines an asymptotic  $100(1-a)\%$  confidence interval under appropriate regularity conditions [39]:

$$\left\{ \omega, \frac{\mathcal{L}_p(\omega)}{\mathcal{L}_p(\hat{\omega})} > k \right\} \quad (21)$$

where  $\omega$  is the quantity of interest,  $\mathcal{L}_p(\omega)$  is the profile likelihood of  $\omega$ , and  $\mathcal{L}_p(\hat{\omega})$  is the profile likelihood of  $\omega$  evaluated at the MLE, for which  $\omega = \hat{\omega}$ . This confidence can also inform the identifiability of the quantity of interest [13, 48].

Another approach is to use a bootstrap-style sampling distribution [49, 50], determined by repeatedly solving the (discretised) optimisation problem:

$$\begin{aligned} \min_{\theta, c} \quad & \|y - g(x) + e\|_{\Gamma^{-1}}^2 + \|Dx - f(x; \theta) + \nu\|_{\hat{\Sigma}^{-1}}^2 \\ \text{s.t.} \quad & x = \Phi c \end{aligned} \quad (22)$$

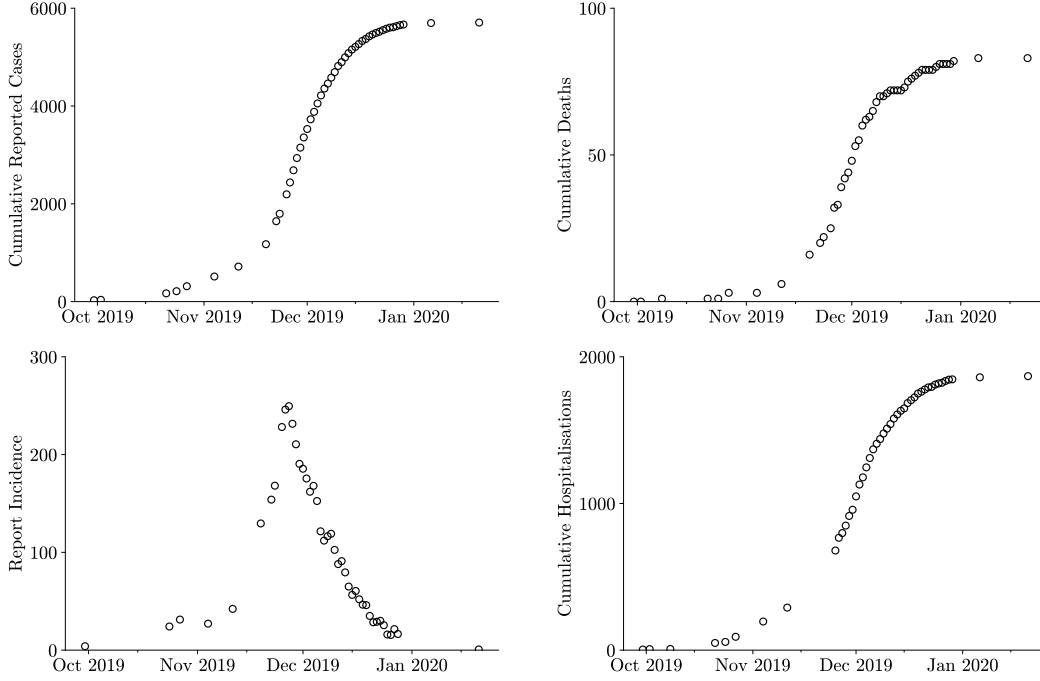
for a series of realisations of  $e$  and  $\nu$ , which have zero mean and covariances  $\Gamma$  and  $\hat{\Sigma}$ . This can be done efficiently using automatic differentiation, and can be carried out in perfect parallel. In the inverse problems literature, the above approach falls under the umbrella of Randomised Maximum Likelihood (RML) [51] methods that include randomised Maximum a Posteriori (rMAP) [52] and Randomize-then-Optimize (RTO) [53]. These methods are usually considered approaches to determining approximate Bayesian posteriors, or proposal distributions for sampling Bayesian posteriors. In contrast, we view our approach as a way of determining a sampling distribution for the maximum likelihood estimator under prior information in the form of stochastic constraints in the style of Theil and Goldberger [21], or a prior likelihood in the style of Edwards [22]. Our interpretation of RML here is thus as a classical bootstrap method, and hence we will also refer to this as ‘RML bootstrap’. Using the bootstrap interpretation, we can construct  $100(1-2a)\%$  percentile interval [50]:

$$\left[ t_a^*, t_{(1-a)}^* \right] \quad (23)$$

where  $t^*$  are the bootstrap samples and  $t_p^*$  is the  $p$ -th percentile of the bootstrap samples.

## 4 Case Study

We now consider an application of these ideas to a case study. Between late September 2019 and early January 2020, the South Pacific nation of Samoa (population  $\sim 200,000$ ) experienced an unprecedented outbreak of measles. Data on the progression of the epidemic was released by various departments of the Government of Samoa in the form of press releases [54]. Initially these were published by the Ministry of Health, but during the emergency period the function



**Figure 1:** Plots of the available data for the 2019 Samoan measles outbreak (current hospitalisations and cumulative discharges available, but not shown). Top-left: cumulative reported cases; top-right: cumulative reported deaths; bottom-left: report incidence (daily rate of new cases since last report); bottom-right: cumulative reported hospitalisations.

was performed by the National Emergency Operations Centre (NEOC). Initial press releases by the Ministry of Health were generally sparse, whereas NEOC press releases were daily. The data is presented in Figure 1. Note that there is missing data for certain dates near the beginning of the outbreak.

In a preliminary effort, we used a less developed version of this method to make predictions for the Samoan measles epidemic, when contacted by the Samoan Observer in late November 2019 [55, 56]. For our analysis here, we use data up to the 29th of November to reproduce similar conditions. This is the point in time where two pertinent questions are raised: how long will the epidemic persist, and how bad will it be?

#### 4.1 Model

To model the dynamics, we build upon the SEIR compartmental model, adding compartments for the additional data available (deaths and hospitalisations).

$$\dot{S} = -\beta S \frac{I}{N}, \quad (24a)$$

$$\dot{E} = \beta S \frac{I}{N} - \gamma E, \quad (24b)$$

$$\dot{I} = \gamma E - (\alpha + \eta + \mu)I, \quad (24c)$$

**Table 1:** Data to state variable correspondences for all observable states of the measles model

Model State Variable	Corresponding Available Data
$I_c$	Cumulative Reports
$D$	Cumulative Deaths
$H$	Current Admissions
$H_c$	Cumulative Hospitalisations
$G$	Cumulative Discharges

$$\dot{R} = \alpha I, \quad (24d)$$

$$\dot{H} = \eta I - \rho H, \quad (24e)$$

$$\dot{G} = \rho H, \quad (24f)$$

$$\dot{D} = \mu I, \quad (24g)$$

$$\dot{I}_c = \gamma E, \quad (24h)$$

$$\dot{H}_c = \eta I. \quad (24i)$$

$S$  represents the susceptible population, who become exposed ( $E$ ) but not infectious at rate  $\beta \frac{I}{N}$  on contact with the infectious ( $I$ ). Exposed individuals become infectious  $I$  at a rate  $\gamma$ , and report their infection (tracked by  $I_c$ ). Infectious individuals can then recover to  $R$  at rate  $\alpha$ , become hospitalised ( $H$ ) at rate  $\eta$ , or die ( $D$ ) at rate  $\mu$ . Hospitalisation is tracked by  $H_c$ , and hospitalised individuals recover and are discharged ( $G$ ) at rate  $\rho$ . We also track the total number of reported cases ( $I_c$ ), the total number of hospitalisations ( $H_c$ ), and the total at-risk population  $N = S + E + I + R + H + G$ . We neglect vital dynamics, due to the relatively short duration of the outbreak, and also vaccination effects, which were implemented too late into the epidemic to have a significant effect. Pre-existing vaccinated individuals are modelled as a non-zero  $R(t = 0)$ . Because we are assuming this existing population of vaccinated individuals, we define a critical threshold parameter  $\mathcal{R}_c$  following [57, 58] as

$$\mathcal{R}_c = \frac{\beta}{\alpha + \eta + \mu} \frac{S(t = 0)}{N(t = 0)} \quad (25)$$

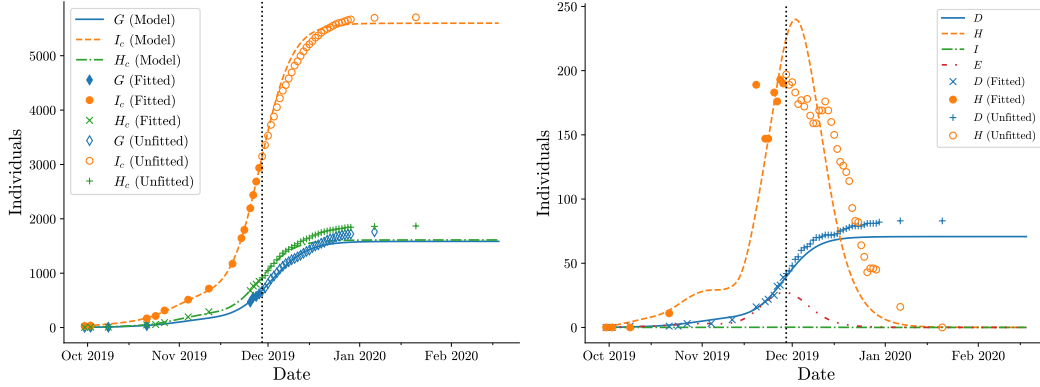
We note that not all states have corresponding data, that is we have a partially observed model. The observed states are described in Table 1. Partially observed models introduce additional problems when fitting a differential equation model. At a basic level, it becomes difficult to quantify model misfit without estimating the state simultaneously. It also can introduce identifiability problems, depending on the state variables that can be observed [59].

## 4.2 Results

The methods were implemented in `Python` as described in Section 3.1, and are initially validated on an SIR model with synthetic data (available in the Supplementary Material). The estimated trajectory of the Samoan measles outbreak, using data up to 29 November 2019, is presented in Figure 2.

We compute confidence intervals for the total reported cases and total deaths, as well as  $\mathcal{R}_c$ , using profile likelihood and RML bootstrap methods described above. We draw 200 samples for the RML bootstrap method. Intervals are computed for a 95% confidence level.

The bootstrap interval for final cases is [5381 – 6041], and the profile likelihood interval is [5244 – 6000], as in Figure 3. Figure 4 plots the RML state estimates for the total reported cases



(a) Cumulative reports ( $I_c$ ), cumulative hospitalisations ( $H_c$ ) and discharges ( $G$ ) (b) Deaths ( $D$ ), incidence of non-infectious individuals ( $E$ ) and infectious individuals ( $I$ ), and hospitalisations ( $H$ )

**Figure 2:** Recovered state estimates at  $N = 3$  iterations of the IRLS algorithm, which is the chosen early stopping point.

over time, which shows that the bootstrap estimate captures the final size, but does not capture the intermediate timepoints correctly.

Due to the implicit assumption that the model is misspecified, the individual parameters are unlikely to correctly reflect the true physical rates. We see this reflected in the spread of most of the parameters in Figure 5, which suggests an element of nonidentifiability in the model.

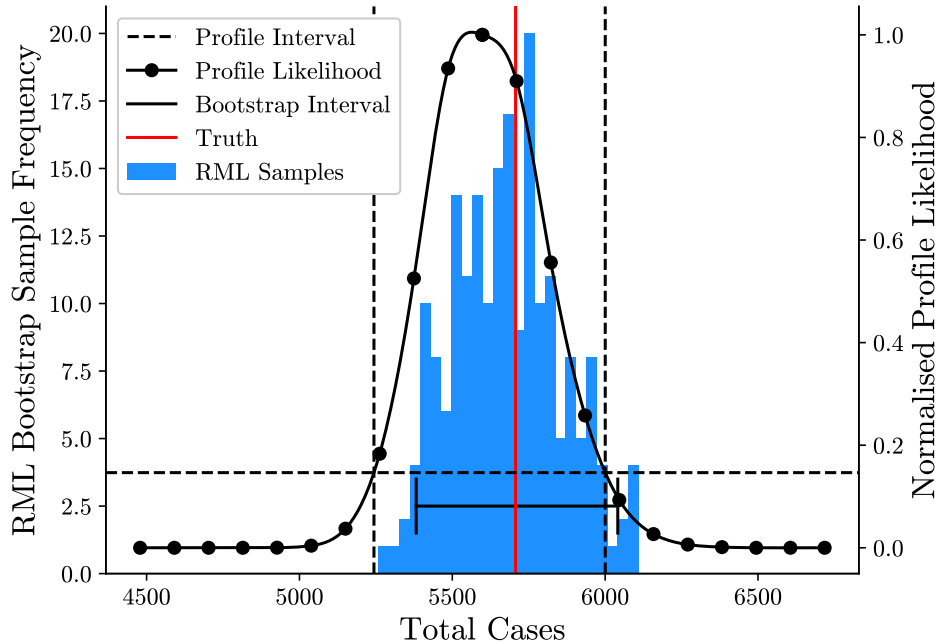
We produce a profile of  $\mathcal{R}_c$  in Figure 6, giving a profile likelihood confidence interval of  $[1.0223 - 1.0432]$ , which is larger than the bootstrap interval  $[1.0242 - 1.0273]$ .

We also extract the state estimates along the  $\mathcal{R}_c$  profile in Figure 7a which can then characterise the behaviour of the expected state, given the data and that  $\mathcal{R}_c$  is a given value. For clarity, Figure 7b provides a profile of the final number of cases along the profile of  $\mathcal{R}_c$ . We see that this does entrap the truth, unlike the bootstrap samples.

We can also analyse the bootstrap samples for estimates of the death counts. We plot a histogram of the estimated deaths from the outbreak in Figure 8, and derive an 95% bootstrap confidence interval of  $[67.2 - 77.5]$ , which does not capture the true value of 83 deaths. The profile interval of  $[65.7 - 76.8]$  also does not. If we look at the estimated case fatality rate (CFR), as estimated by evaluating  $D/cE$  at the final time (the state-derived estimate) or by  $\mu/(\mu + \alpha + \eta)$  (the parameter-derived estimate), we see that they lie around the region of 1.2% to 1.3%. We know that the CFR for this outbreak was 1.4%, so there is some systematic bias in its estimation with our method. Looking at the CFR over time as computed from the empirical data in Figure 9, we see that the CFR increases over time, which is likely due to a delay between reporting the infection and recovering (or in this case death).

## 5 Discussion

The estimation and prediction of epidemics is a difficult task, in both theory and practice. We have presented a likelihood-based reformulation of the standard generalised profiling method introduced and developed by Ramsay et al. [19]. This is based on ideas of mixed estimation in the sense of Theil and Goldberger [21] and prior likelihood in the sense of Edwards [22]. This



**Figure 3:** Profile likelihood and RML bootstrap samples ( $n = 200$ ) of the total reported cases, with corresponding 95% confidence intervals marked.

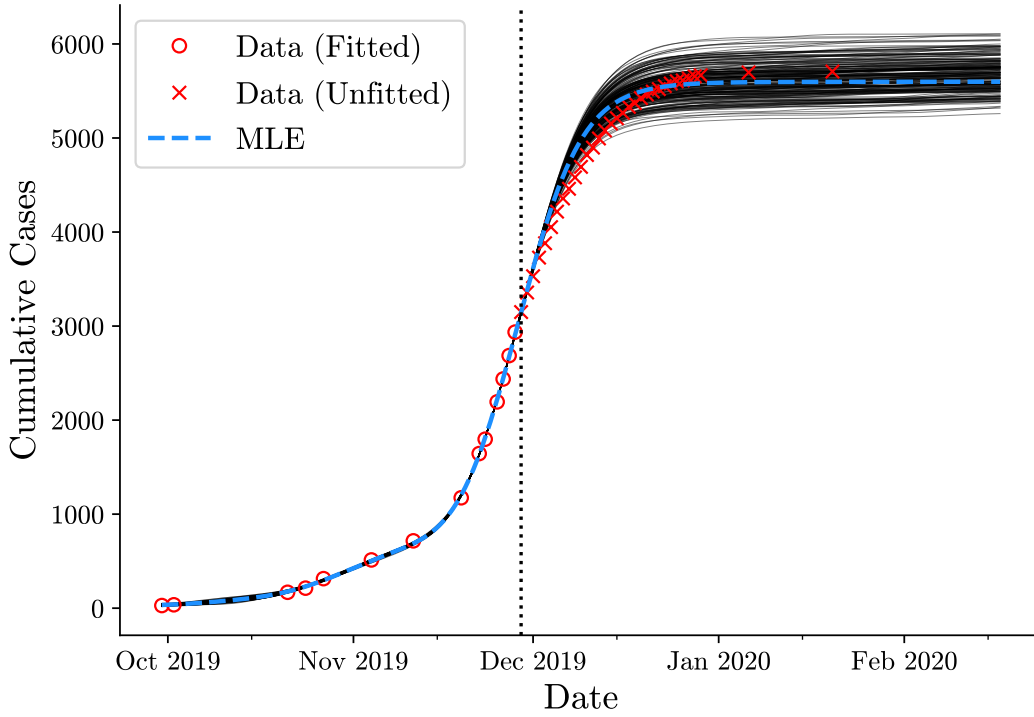
provides a natural rationale for profiling as a method of eliminating nuisance parameters and targeting interest parameters. We then use this to construct a framework for parameter and state inference of a model given data while allowing for model misspecification.

The standard formulation can be seen as a form of profiling, where the outer objective is the  $\theta$ -profile of the inner objective, albeit evaluated at  $\lambda = 0$ , where the model is exactly enforced. Our reformulation can thus be seen as more natural in a profile likelihood sense, as it interprets both stages as profiling a single log-likelihood containing both observational and model error. It also leads to the use of profile likelihood intervals [39] that allow us to perform uncertainty and identifiability analysis on the parameters and states.

We have then applied these methods to a case study of a measles outbreak in Samoa, and shown that the estimates capture the truth, for important in-sample quantities of interest, like  $\mathcal{R}_c$  and the total number of cases.

### 5.1 Profile Interpretation of Generalised Profiling

A key methodological difference in our approach is in the treatment of the smoothing hyperparameter in the standard formulation in terms of covariance matrices of errors. This is similar to the ideas developed in the mixed-effects/random-effects/multi-level regression literature [38, 60], which embed spline models in mixed effect models using random effects. These approaches also relate directly to Bayesian interpretations and the work of Wahba [37] on spline models. As noted above, however, a key motivation for our interpretation in terms of stochastic *constraints*, rather than an explicit stochastic process model, is to provide a more natural rationale for using profiling (maximisation) rather than marginalisation for eliminating nuisance parameters. There

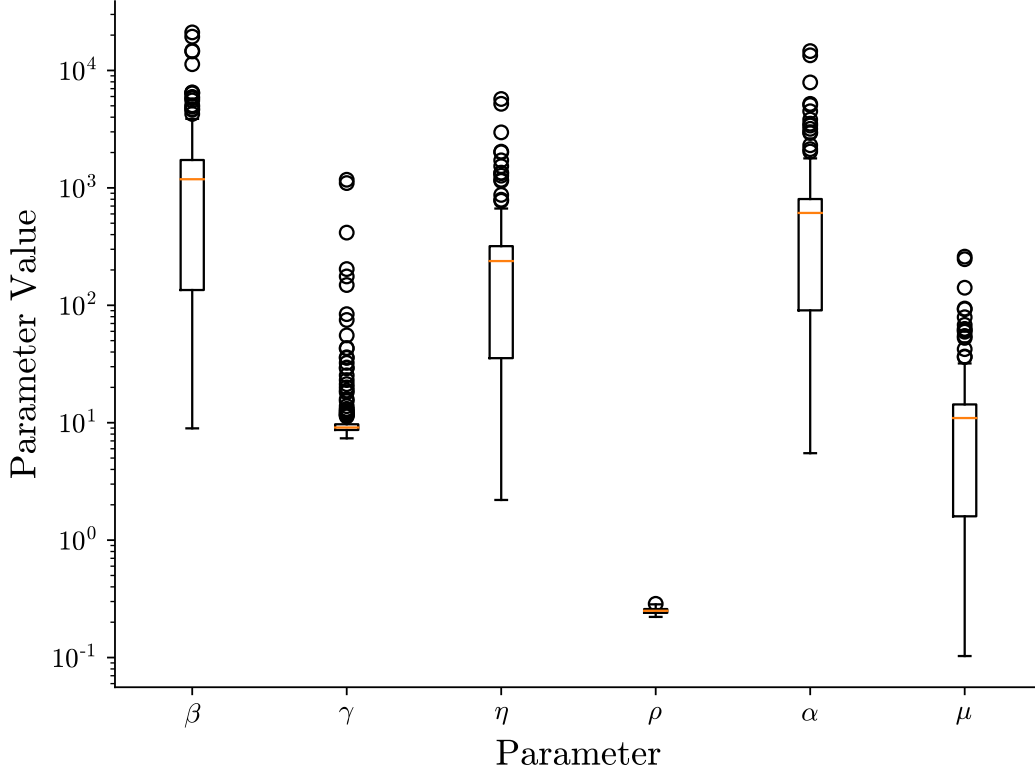


**Figure 4:** Timeseries of RML bootstrap samples of total reported cases ( $n = 200$ ).

is ongoing debate whether profiling or marginalisation better captures properties of likelihood when eliminating nuisance parameters [61–63].

When estimating covariances with our approach, there is still the open question of an optimal stopping criterion, or optimal number of cycles. Furthermore, it is unclear how well covariance matrices can be estimated in this setting, reflecting similar difficulties noted in the mixed effects literature [60]. Validation methods have been used with generalised profiling, such as the forwards prediction error suggested by [64], but there are problems when using this for validating emerging outbreaks such as the Samoan measles data, since prediction is typically most sensitive to the more recent data. More recently, Bayesian approaches to mixed effects models often include additional information on covariances in the form of hyperpriors. Huang et al. [65] have considered this for their Bayesian reinterpretation of the standard formulation of generalised profiling, setting priors on the tuning parameter  $\lambda$ . In principle we could take an analogous approach in terms of our mixed estimation (stochastic constraints) formulation, but choosing appropriate forms for these additional terms faces similar challenges to those raised in specifying hyperpriors in mixed effect models [60]. In general, estimating these hyperparameters remains a key challenge for all approaches.

For the case study at hand, we decided to take the pragmatic option, and compare each iteration using out-of-sample data that we have not used for the fitting process. This includes the estimated total population, compared to the official 2016 Census data [66]; and the estimated vaccinated proportion, compared to the known MCV1 metric (at least one dose of measles-containing vaccine) as of 2018 [67]. We find that the iteration that had the lowest justifiable error



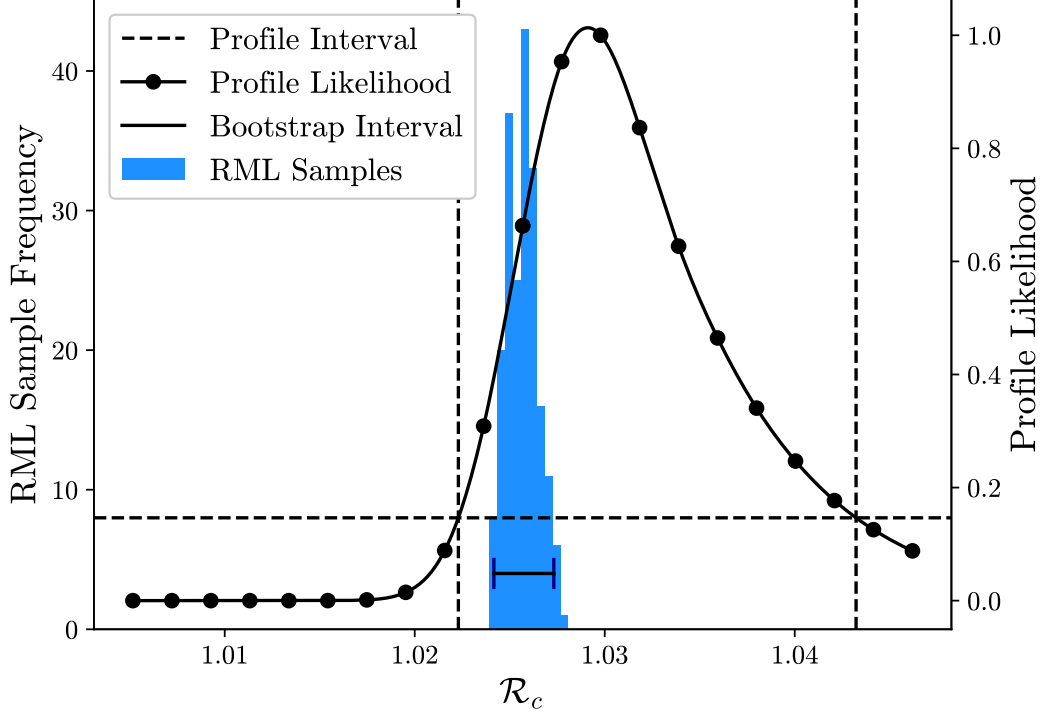
**Figure 5:** Boxplot of bootstrap samples of the model parameters for the Samoan measles case study ( $n=200$ ). These suggest that  $\beta$  (transmission rate),  $\eta$  (hospitalisation rate),  $\alpha$  (recovery rate) and  $\mu$  (death rate) are likely nonidentifiable, due to their spread, whereas  $\rho$  (hospital discharge rate) is identifiable, and  $\gamma$  (incubation rate) is one-sidedly identifiable.

in these two categories also best matched the out-of-sample epidemic data that we withheld in the fitting process.

## 5.2 Identifiability and Uncertainty Quantification

Due to the ability to use the standard likelihood toolbox with the reformulation, we are also able to perform identifiability analysis in a straightforward manner. We see in Figure 5 that there are some parameters that are not identifiable in our model - notably the contact parameter  $\beta$  and the recovery rate  $\alpha$ , which both display a huge range of variability in the bootstrap samples that cross multiple orders of magnitude. This is in agreement with the work by Tuncer and Le [59] that showed an SEIR-type model is practically nonidentifiable to cumulative incidence data (which is what we are fitting to). We note that their analyses of structural identifiability make strong assumptions about having knowledge of the initial conditions and total population size, which are left as free in our analysis.

We see that  $\mathcal{R}_c$  is identifiable in our analysis, which is consistent with the findings of Roosa and Chowell [68] that show that  $\mathcal{R}_0$  is identifiable even if there are identifiability issues with the other parameters. However, we do see some disagreement between the width of the intervals



**Figure 6:** Profile likelihood and RML bootstrap samples ( $n = 200$ ) of  $\mathcal{R}_c$ , with associated 95% confidence intervals marked.

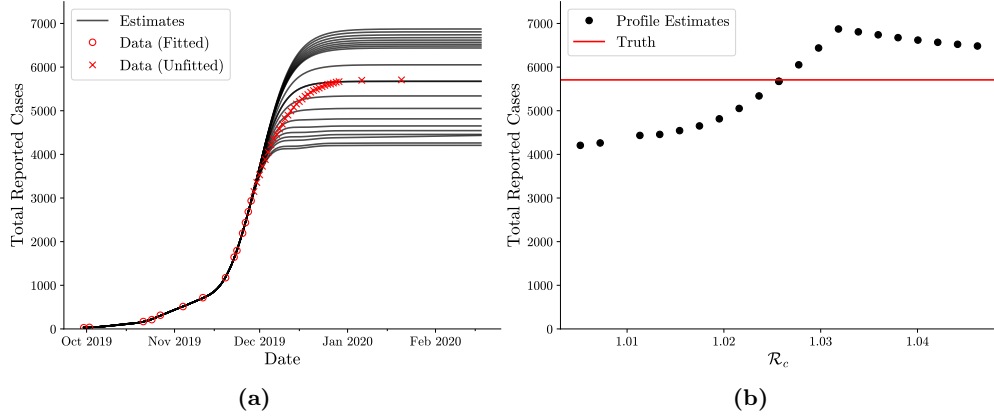
generated by profile likelihood analysis and bootstrapping for parameters such as  $\mathcal{R}_c$  with the bootstrap intervals being typically narrower (Figure 6), suggesting higher identifiability than given by profile likelihood analysis, something that Fröhlich et al. [14] also noted. However, profiling over state space values, such as the total cases and total deaths in Figures 3 and 8 shows a good agreement between profile likelihood and bootstrap methods. It is likely that the validity of the bootstrap intervals is related to the correctness of the sampling distribution of the estimator for a given parameter. We assume Gaussian zero-mean uncorrelated observational and model error which may not be correct – for example data from a system with a higher  $\mathcal{R}_c$  would have systematically higher case counts at every time point.

One other thing to note is that our bootstrap method is novel in comparison to the existing methods [14, 68] due to the explicit resampling of the model error, in addition to observational error. However, we typically see that the covariances of the model error are quite small compared to the observational error at the chosen IRLS iteration, so it does have limited effect on the samples.

### 5.3 Model Misspecification

We see elements of model misspecification, which may stem from the ill-posedness of the problem. Empirically, we see that the fit of the case fatality rate, an “out-of-sample” parameter, is biased in both the bootstrap and profile likelihood analysis. Drawing on our knowledge of the case study, we also know that our model does not capture the spatial characteristics of the spread in Samoa,



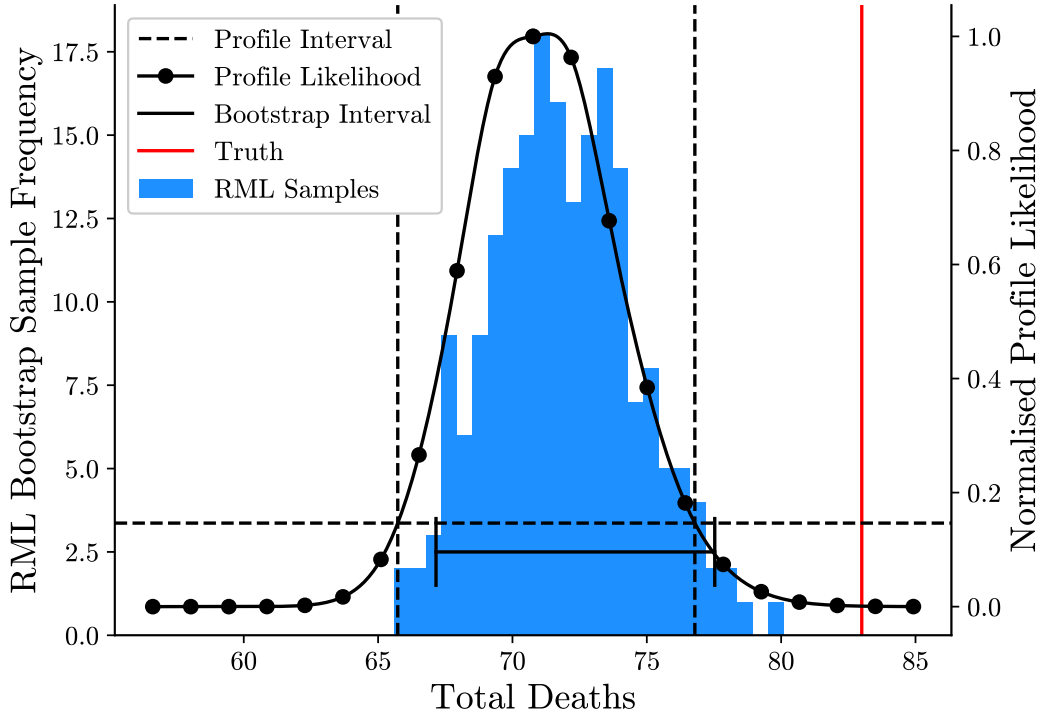


**Figure 7:** State estimates of the total number of cases along the profile of  $\mathcal{R}_c \in [1.005, 1.046]$  (a) timeseries of the cumulative reported cases (b) total number of reported cases.

or the fact that various non-pharmaceutical interventions were enacted in the form of curfews and a mass vaccination campaign. We could have allowed for this in the model, for example, by allowing for some of the parameters to be time-varying in a structured way, akin to how Hooker et al. [5] allow for a cyclical  $\beta$  value in their endemic model. Further, we have information from previous measles outbreaks about the clinical presentation and infection characteristics of the virus, which could inform poorly identified parameters such as  $\gamma$  (the latent period) and  $\alpha$  (the recovery rate). This additional information could be included as regularising terms in the likelihood function, using the same stochastic constraint arguments. In this way, this method can be seen as analogous to the Bayesian formulation, but instead of marginalising over the prior, we maximise (profile) over the prior. Additionally, our model assumes that hospitalisation and death processes compete, and are exclusive of each other. In reality, hospitalised individuals are also likely to die. This is also a likely contributing factor to the underestimation of the total number of deaths.

## 5.4 Practical Considerations

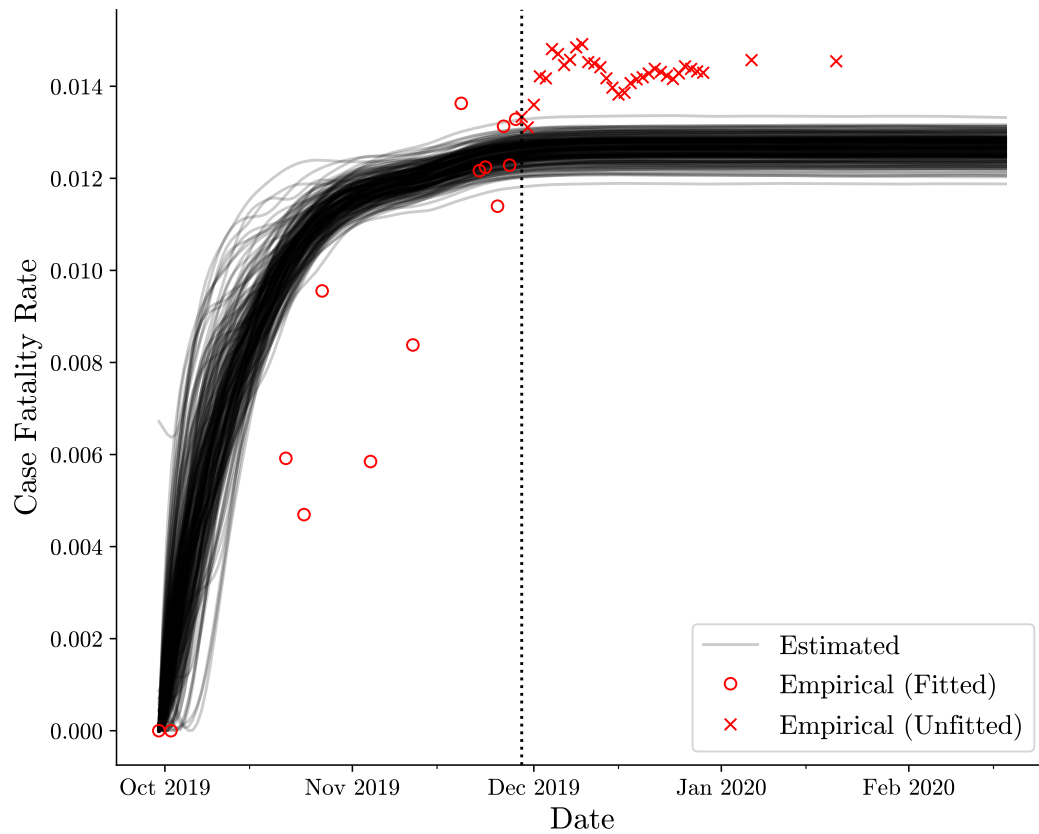
In our implementation of the method, we have used the **CasADi** framework [69] and its interface in Python. We found that the estimation of the MLE for the Samoan measles case study took around 10 minutes, with multiple IRLS iterations not converging. The profiles each take a similar amount of time to generate, and the RML bootstrap samples take around 30 minutes. Most of these timings are scale with the number of basis functions, and collocation points, though the nonconvergence of the optimisation problem in many cases is more likely attributed to poor starting iterates. Our analysis of an SIR model (in the Supplementary Material) takes significantly less time, for a similar amount of data. Additionally, we have made a simplifying choice to perform all of the procedures in serial. However, there is room for parallelisation to significantly speed up the process of bootstrapping in particular, since bootstrap samples are independent of each other. For profiling, many different values of interest can be computed in parallel to each other, but the profile itself is constructed successively away from the MLE. This is due to the limitation of nonlinear optimisation – the fact that there are limited regions of convergence for any given problem. We see nonconvergence in many of the steps in the maximum likelihood estimation and the uncertainty quantification procedures, which computationally is flagged by longer runtimes.



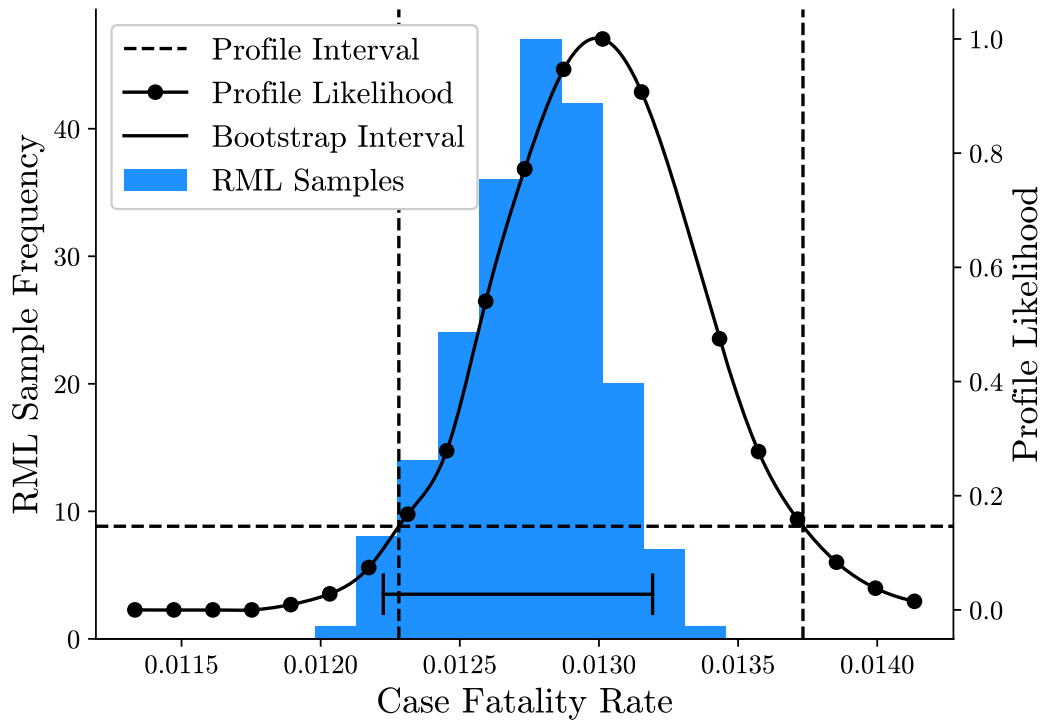
**Figure 8:** Profile likelihood and RML bootstrap samples ( $n = 200$ ) of the total deaths with associated 95% confidence intervals marked.

The successive chaining of estimates in the profile likelihood approach is particularly sensitive to this problem, since the result of the previous profile point (whether converged or not) is used as the initial iterate for the estimation of the next profile point. These convergence problems are exacerbated if the problem is relatively large, and we find that the explicit representation approach we use to save computational time has a large memory burden. We found that an RTO-style approach [53], where the Jacobian must be used in the log-likelihood when resampling for a bootstrap, is infeasible on a typical computational workstation with 16 GB of RAM.

**Acknowledgements.** We wish to acknowledge Sapeer Mayron from the Samoa Observer for providing access to press releases that were not readily available. OJM would like to thank Elvar Bjarkason and Ruanui Nicholson for helpful discussions of the RTO and RML methods in the context of inverse problems.



**Figure 9:** Observed case fatality rate over time, computed as the total deaths divided by the total reported cases, from the empirical data and the 200 RML bootstrap samples.



**Figure 10:** Profile likelihood and RML bootstrap samples ( $n = 200$ ) of the case fatality rate as computed by parameters  $\frac{\mu}{\mu + \alpha + \eta}$ , with associated 95% confidence intervals marked.

## References

1. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press; 1991.
2. Ross R. *The prevention of malaria*. 2nd ed. London: Murray; 1911.
3. Kermack WO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 1927 Aug;115(772):700–721. ([doi:10.1098/rspa.1927.0118](https://doi.org/10.1098/rspa.1927.0118)).
4. Schenzle D. An Age-Structured Model of Pre- and Post-Vaccination Measles Transmission. *Mathematical Medicine and Biology: A Journal of the IMA*. 1984 Jan;1(2):169–191. ([doi:10.1093/imammb/1.2.169](https://doi.org/10.1093/imammb/1.2.169)).
5. Hooker G, Ellner SP, Roditi LDV, Earn DJD. Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario. *Journal of the Royal Society, Interface*. 2011 Jul;8(60):961–74. ([doi:10.1098/rsif.2010.0412](https://doi.org/10.1098/rsif.2010.0412)).
6. Xia Y, Bjørnstad O, Grenfell B. Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics. *The American Naturalist*. 2004;164(2):267–281. ([doi:10.1086/422341](https://doi.org/10.1086/422341)).
7. Allen LJS. *An Introduction to Stochastic Epidemic Models*. Springer, Berlin, Heidelberg; 2008. p. 81–130. ([doi:10.1007/978-3-540-78911-6\\_3](https://doi.org/10.1007/978-3-540-78911-6_3)).
8. Allen LJS. *An introduction to stochastic processes with applications to biology*. Chapman & Hall/CRC; 2011.
9. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. London: Imperial College; 2020. ([doi:10.25561/77482](https://doi.org/10.25561/77482)).
10. Wilke CO, Bergstrom CT. Predicting an epidemic trajectory is difficult. *Proceedings of the National Academy of Sciences*. 2020 Nov;p. 28549–28551. ([doi:10.1073/pnas.2020200117](https://doi.org/10.1073/pnas.2020200117)).
11. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*. 2020 Jan;5:271–281. ([doi:10.1016/j.idm.2020.03.001](https://doi.org/10.1016/j.idm.2020.03.001)).
12. He D, Ionides EL, King AA. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society, Interface*. 2010 Feb;7(43):271–83. ([doi:10.1098/rsif.2009.0151](https://doi.org/10.1098/rsif.2009.0151)).
13. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*. 2009 Aug;25(15):1923–1929. ([doi:10.1093/bioinformatics/btp358](https://doi.org/10.1093/bioinformatics/btp358)).
14. Fröhlich F, Theis FJ, Hasenauer J. *Uncertainty Analysis for Non-identifiable Dynamical Systems: Profile Likelihoods, Bootstrapping and More*. Springer, Cham; 2014. p. 61–72. ([doi:10.1007/978-3-319-12982-2\\_5](https://doi.org/10.1007/978-3-319-12982-2_5)).
15. Kennedy MC, O’Hagan A. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001 Aug;63(3):425–464. ([doi:10.1111/1467-9868.00294](https://doi.org/10.1111/1467-9868.00294)).

16. Brynjarsdóttir J, O'Hagan A. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*. 2014 Nov;30(11):114007. (doi:10.1088/0266-5611/30/11/114007).
17. King AA, Domenech de Cellès M, Magpantay FM, Rohani P. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society B: Biological Sciences*. 2015;282(1806):20150347.
18. Chatzilena A, van Leeuwen E, Ratmann O, Baguelin M, Demiris N. Contemporary statistical inference for infectious disease models using Stan. *Epidemics*. 2019 Oct;p. 100367. (doi:10.1016/J.EPIDEM.2019.100367).
19. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007 Nov;69(5):741–796. (doi:10.1111/j.1467-9868.2007.00610.x).
20. Durbin J. A Note on Regression When There is Extraneous Information About One of the Coefficients. *J Am Stat Assoc*. 1953 Dec;48(264):799–808.
21. Theil H, Goldberger AS. On pure and mixed statistical estimation in economics. *International Economic Review*. 1961;2(1):65–78.
22. Edwards AWF. Statistical methods in scientific inference. *Nature*. 1969 Jun;222(5200):1233–1237.
23. Ramsay J, Hooker G. *Dynamic Data Analysis*. Springer Series in Statistics. New York, NY: Springer New York; 2017. (doi:10.1007/978-1-4939-7190-9).
24. Van Kampen NG. *Stochastic processes in physics and chemistry*. vol. 1. Elsevier; 1992.
25. Law K, Stuart A, Zygalakis K. *Data Assimilation: A Mathematical Introduction*. vol. 62. Springer; 2015.
26. Diggle PJ, Gratton RJ. Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1984;46(2):193–212.
27. Bretó C, He D, Ionides EL, King AA, et al. Time series analysis via mechanistic models. *The Annals of Applied Statistics*. 2009;3(1):319–348.
28. Wood SN. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*. 2010;466(7310):1102–1104.
29. Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A. Statistical inference for stochastic simulation models – theory and application. *Ecol Lett*. 2011 Aug;14(8):816–827.
30. Fasiolo M, Pya N, Wood SN. A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Statistical Science*. 2016;p. 96–118.
31. King AA, Nguyen D, Ionides EL. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software*. 2016;69(i12).
32. Ionides EL, Breto C, Park J, Smith RA, King AA. Monte Carlo profile confidence intervals for dynamic systems. *J R Soc Interface*. 2017 Jul;14(132).
33. Bretó C. Modeling and inference for infectious disease dynamics: a likelihood-based approach. *Statistical Science: A review journal of the Institute of Mathematical Statistics*. 2018;33(1):57.

34. Wilkinson DJ. Stochastic modelling for systems biology. CRC press; 2018.
35. Durbin J, Koopman SJ. Time series analysis by state space methods. Oxford university press; 2012.
36. Ramsay J, Silverman BW. Functional Data Analysis. Springer Science & Business Media; 2005.
37. Wahba G. Spline Models for Observational Data. vol. 59. SIAM; 1990.
38. Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression. Cambridge University Press; 2003. (doi:10.1017/cbo9780511755453).
39. Pawitan Y. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press; 2001.
40. Rao CR, Toutenburg H, Shalabh HC, Schomaker M. Linear model and generalization, least squares and alternatives. 3rd ed. Springer Series in statistics. Berlin, Heidelberg: Springer; 2008. (doi:10.1007/978-3-540-74227-2).
41. Lee Y, Nelder JA. Hierarchical generalized linear models. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(4):619–656.
42. Lee Y, Nelder JA, Pawitan Y. Generalized linear models with random effects: unified analysis via H-likelihood. vol. 153. CRC Press; 2018.
43. Kloeden PE, Platen E. Numerical solution of stochastic differential equations. vol. 23. Springer Science & Business Media; 2013.
44. Cao J, Ramsay JO. Parameter cascades and profiling in functional data analysis. Computational Statistics. 2007 Sep;22(3):335–351. (doi:10.1007/s00180-007-0044-1).
45. Campbell DA, Chkrebtii O. Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates. Mathematical Biosciences. 2013 Dec;246(2):283–292. (doi:10.1016/J.MBS.2013.03.011).
46. Xun X, Cao J, Mallick B, Maity A, Carroll RJ. Parameter Estimation of Partial Differential Equation Models. Journal of the American Statistical Association. 2013 Sep;108(503):1009–1020. (doi:10.1080/01621459.2013.794730).
47. Carroll RJ, Ruppert D. Transformation and Weighting in Regression. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC; 1988.
48. Simpson MJ, Baker RE, Vittadello ST, Maclaren OJ. Practical parameter identifiability for spatio-temporal models of cell invasion. Journal of The Royal Society Interface. 2020 Mar;17(164):20200055. (doi:10.1098/rsif.2020.0055).
49. Davison AC, Hinkley DV. Bootstrap Methods and their Application. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press; 1997.
50. Efron B. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics. 1979;7(1):1–26.
51. Oliver DS, He N, Reynolds AC. Conditioning permeability fields to pressure data. In: ECMOR V-5th European Conference on the Mathematics of Oil Recovery. earthdoc.org; 1996. .

52. Wang K, Bui-Thanh T, Ghattas O. A Randomized Maximum A Posteriori Method for Posterior Sampling of High Dimensional Nonlinear Bayesian Inverse Problems. *SIAM Journal on Scientific Computing*. 2018;40(1):A142–A171. (doi:10.1137/16M1060625).
53. Bardsley JM, Solonen A, Haario H, Laine M. Randomize-Then-Optimize: A Method for Sampling from Posterior Distributions in Nonlinear Inverse Problems. *SIAM Journal on Scientific Computing*. 2014;36(4):A1895–A1910. (doi:10.1137/140964023).
54. Government of Samoa. Update on the measles outbreak. OCHA; 2019. [Press Release, Multiple]. Available from: [https://reliefweb.int/updates?advanced-search=\(S4201\\_F8\\_D49263\)](https://reliefweb.int/updates?advanced-search=(S4201_F8_D49263)) [cited 22 March 2021].
55. Petousis-Harris H. Samoa’s devastating measles epidemic – why and how bad? *Sciblogs*. 28 Nov 2019. Available from: <https://sciblogs.co.nz/diplomaticimmunity/2019/11/28/samoas-devastating-measles-epidemic-why-and-how-bad/> [cited 22 March 2021].
56. Russell E. Samoan measles outbreak: 70 deaths projected and 6500 infected. *NZ Herald*. 29 Nov 2019. Available from: <https://www.nzherald.co.nz/nz/samoan-measles-outbreak-70-deaths-projected-and-6500-infected/MVSRFCFCRBRF6CO5QS6TEXSNMJ4/> [cited 22 March 2021].
57. van den Driessche P, Watmough J. Further Notes on the Basic Reproduction Number. In: Brauer F, van den Driessche P, Wu J, editors. *Mathematical Epidemiology*. Springer, Berlin, Heidelberg; 2008. p. 159–178.
58. van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*. 2002 Nov;180(1):29–48. (doi:10.1016/S0025-5564(02)00108-6).
59. Tuncer N, Le TT. Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences*. 2018 May;299:1–18. (doi:10.1016/J.MBS.2018.02.004).
60. Hodges JS. *Richly parameterized linear models: additive, time series, and spatial models using random effects*. CRC Press; 2013.
61. Berger JO, Liseo B, Wolpert RL. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*. 1999 Feb;14(1):1–28. (doi:10.1214/ss/1009211804).
62. Aitkin M, Stasinopoulos M. Likelihood Analysis of a Binomial Sample Size Problem. In: Gleser LJ, Perlman MD, Press SJ, Sampson AR, editors. *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin*. Springer; 1989. p. 399–411. (doi:10.1007/978-1-4612-3678-8\_28).
63. Aitkin M. *Statistical Inference: An Integrated Bayesian/Likelihood Approach*. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC; 2010.
64. Hooker G, Ellner SP. *On Forwards Prediction Error*. Cornell University; 2010.
65. Huang H, Handel A, Song X. A Bayesian approach to estimate parameters of ordinary differential equation. *Computational Statistics*. 2020 Feb;p. 1–19. (doi:10.1007/s00180-020-00962-8).
66. Samoan Bureau of Statistics. 2016 Census Brief No. 1; 2017. Available from: [https://www.sbs.gov.ws/images/sbs-documents/2016-Census-Brief-No\\_1%20\(Revised%20Version2\).pdf](https://www.sbs.gov.ws/images/sbs-documents/2016-Census-Brief-No_1%20(Revised%20Version2).pdf).



67. World Health Organisation, United Nations Children's Fund. Samoa: WHO and UNICEF estimates of immunization coverage: 2018 revision; 2019. Available from: [https://www.who.int/immunization/monitoring\\_surveillance/data/wsm.pdf](https://www.who.int/immunization/monitoring_surveillance/data/wsm.pdf).
68. Roosa K, Chowell G. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling*. 2019 Dec;16(1):1. (doi:10.1186/s12976-018-0097-6).
69. Andersson JAE, Gillis J, Horn G, Rawlings JB, Diehl M. CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*. 2019;11(1):1–36. (doi:10.1007/s12532-018-0139-4).

# Supplementary Material

David Wu<sup>\*1</sup>, Helen Petousis-Harris<sup>2</sup>, Janine Paynter<sup>2</sup>, Vinod Suresh<sup>1,3</sup>, and Oliver J. Maclaren<sup>1</sup>

<sup>1</sup>*Department of Engineering Science, University of Auckland, New Zealand*

<sup>2</sup>*Department of General Practice and Primary Health Care, University of Auckland, New Zealand*

<sup>3</sup>*Auckland Bioengineering Institute, University of Auckland, New Zealand*

## A Application to SIR model

We can apply the method to a synthetic dataset generated from an SIR model. We generate data  $y$  according to the model in Equation (1).

$$y = \begin{pmatrix} S \\ R \end{pmatrix} + \epsilon \quad (1a)$$

$$\frac{dx}{dt} = \frac{d}{dt} \begin{pmatrix} S \\ I \\ R \end{pmatrix} = \begin{pmatrix} -\beta SI/N \\ \beta SI/N - \alpha I \\ \alpha I \end{pmatrix} \quad (1b)$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I) \quad (1c)$$

$$x(0) = (N - 1, 1, 0)^T \quad (1d)$$

We note that the data generated has used a Gaussian error term for simplicity, although it does produce non-physical data, such as having negative values and decreasing total case trajectories.

The differential equation model of Equation (1) is also used as our model for inference, with unknown  $\theta = (\beta, \alpha, N, \text{ and } \sigma)$ , so the model is not misspecified in this case. We assume iid observation noise, and iid model error, giving the objective negative log-likelihood

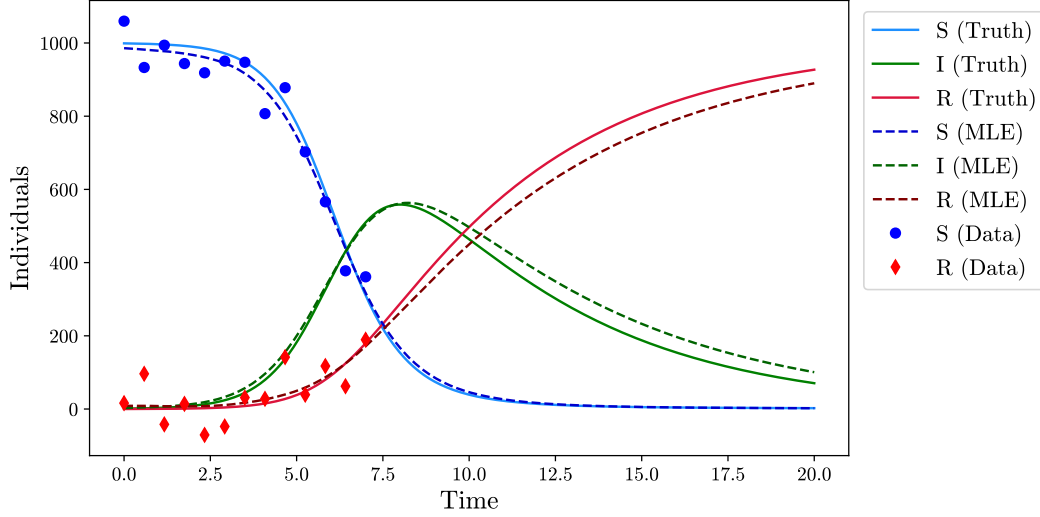
$$l(\theta, x) = \|w_0(y - g(x))\|^2 + \|w_1(\mathcal{D}x - f(x, \theta))\|^2, \quad (2)$$

for the observation function  $g(x) = \begin{pmatrix} S \\ R \end{pmatrix}$ , the derivative operator  $\mathcal{D} \cdot = \frac{d}{dt} \cdot$ , and some weights  $w_0, w_1$  that correspond to the inverse standard deviation of the observation noise and model (discrepancy) error respectively. We note that in this model that  $\mathcal{R}_c = \frac{\beta}{\alpha N} = \mathcal{R}_0$ , since there is no vaccination.

We generate data with  $x(0) = (999, 1, 0)$ ,  $\alpha = 0.2$ ,  $\beta = 1.3$ ,  $\sigma = 50$ . We then fit this using initial iterates  $x^{(0)}(t) = (1000, 1000, 1000)$ ,  $\alpha^{(0)} = \beta^{(0)} = 1$ , initial weights  $w_0 = 1$ ,  $w_1 = 1 \times 10^{-2}$ , and recover state estimates after 3 iterations plotted in Figure 1. We provide a plot of the estimated

<sup>\*</sup>Corresponding Author. dwu402@aucklanduni.ac.nz

weights at each iteration in Figure 2. The weights lie on a L-shaped curve that is analogous to the L-curve in inverse problems literature [1, 2].



**Figure 1:** Fit of the SIR model as taken at three iteration of IRLS, plotted against the data used for fitting and the underlying (non-noisy) truth.

We then profile over the parameters  $\beta$  and  $\alpha$ , as well as  $\mathcal{R}_c$  in Figures 3 to 5. We can also generate predictive uncertainties with RML bootstrapping, and also generate a bootstrap version of the parameter uncertainty. 95% intervals derived from both the profile likelihood and the bootstrap samples are listed in Table 1.

**Table 1:** Confidence intervals for the parameters of the SIR model as derived from profile likelihood and RML bootstrap samples ( $n = 200$ )

Value	Truth	Profile Likelihood 95% Interval	Bootstrap 95% Interval
$\beta$	1.3	[0.8332, 1.650]	[0.8137, 1.704]
$\alpha$	0.2	[0.05349, 0.3666]	[0.07015, 0.3528]
$\mathcal{R}_c$	6.5	[3.819, 19.73]	[3.848, 14.84]

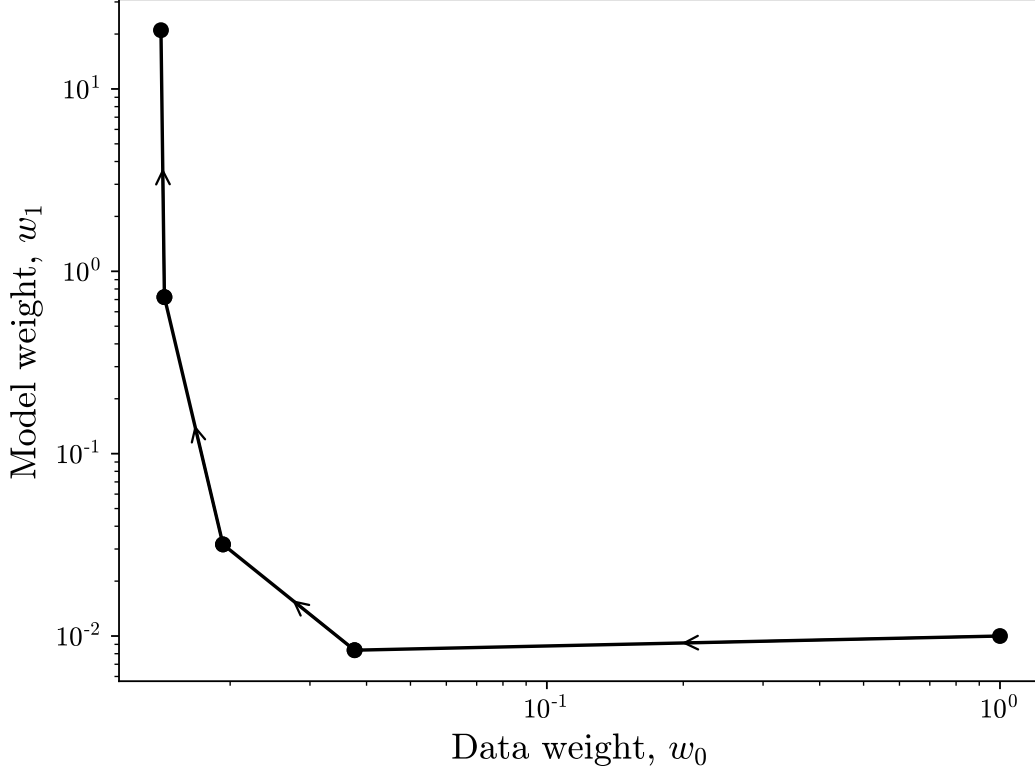
We note that the solve time for the MLE is 3 seconds, profiling takes a total of 1.5 seconds, and RML bootstrap sampling takes 37 seconds.

## B Implementation Detail

Implementation of this method can be found on GitHub at <https://github.com/dwu402/pypei>.

For the case study, we construct the objective function with the additional physical constraints

$$\begin{aligned}
0 &\leq \Phi_C = x \leq N_{max} \\
x_{I_c}(0) &\geq x_I(0) + x_E(0)
\end{aligned}$$



**Figure 2:** Log-log plot of weights  $w_0, w_1$  of the SIR model log-likelihood as estimated at each iteration.

$$\begin{aligned} x_{H_c}(0) &\geq x_G(0) \\ \theta &\geq 0 \end{aligned}$$

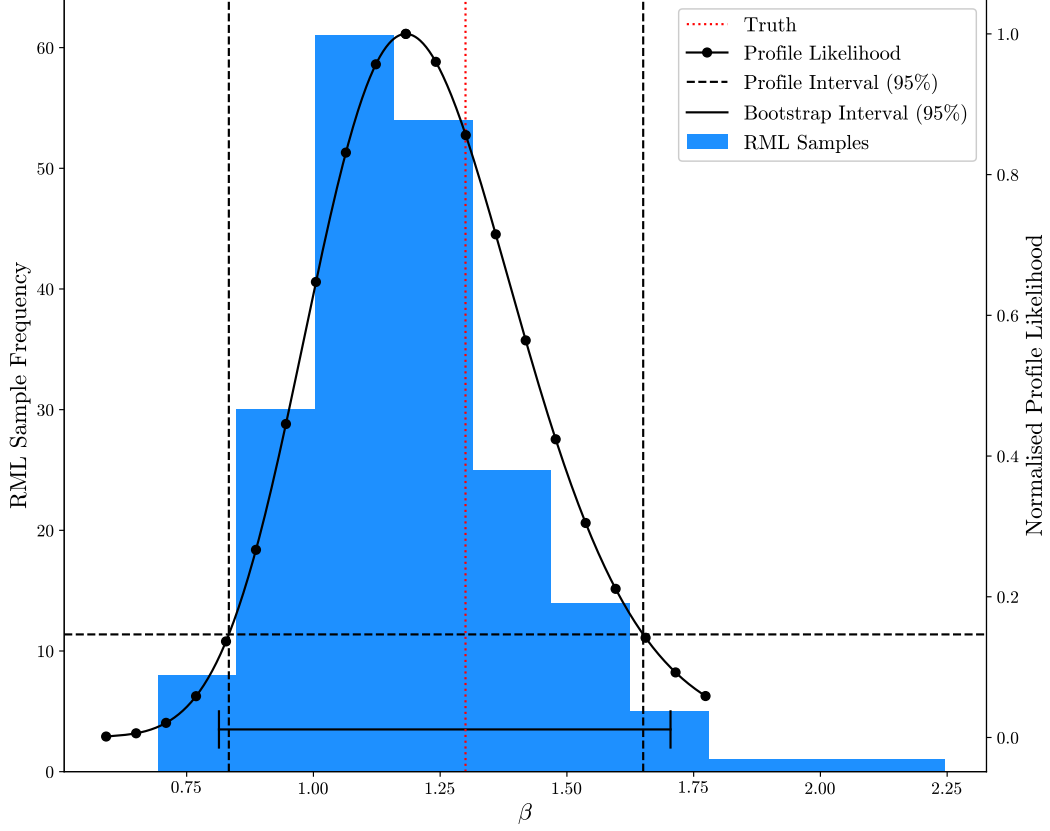
representing that the quantities in state space must be positive, and bounded by some  $N_{max}$  representing the total population; that the number of reported cases must include the number of currently infected individuals; and that the model parameters had the correct (physically motivated) sign.

The problem is represented in **CasADi** [3] by computing  $\Phi$  over a specified time grid  $\mathbb{T} = \{t_1, t_2, \dots, t_N\}$ , by evaluating the generated (MX) spline basis  $\mathbb{S} = \{\phi_1, \phi_2, \dots, \phi_K\}$  over  $\mathbb{T}$ . This has the computational advantage of reducing the calls needed when computing the objective function. The derivative basis  $D\Phi$  was computed similarly by evaluating the derivatives of  $\mathbb{S}$  over  $\mathbb{T}$ .  $c$  and  $\theta$  were represented as **SX** objects.

The objective object was constructed by constructing further **SX** objects: **y0** representing data, **W** representing the whitening matrices. The objective is then constructed as

$$\text{sum}(\text{sumsqr}(\mathbf{W} @ (\mathbf{y0} - \mathbf{y})))$$

where  $@$  is the matrix multiplication operator,  $\mathbf{y}$  represents the state- and parameter-based components. In our case,  $\mathbf{y}$  is the array  $[g(c), D\Phi c - f(\Phi c, \theta)]$ . For our case study, we implement



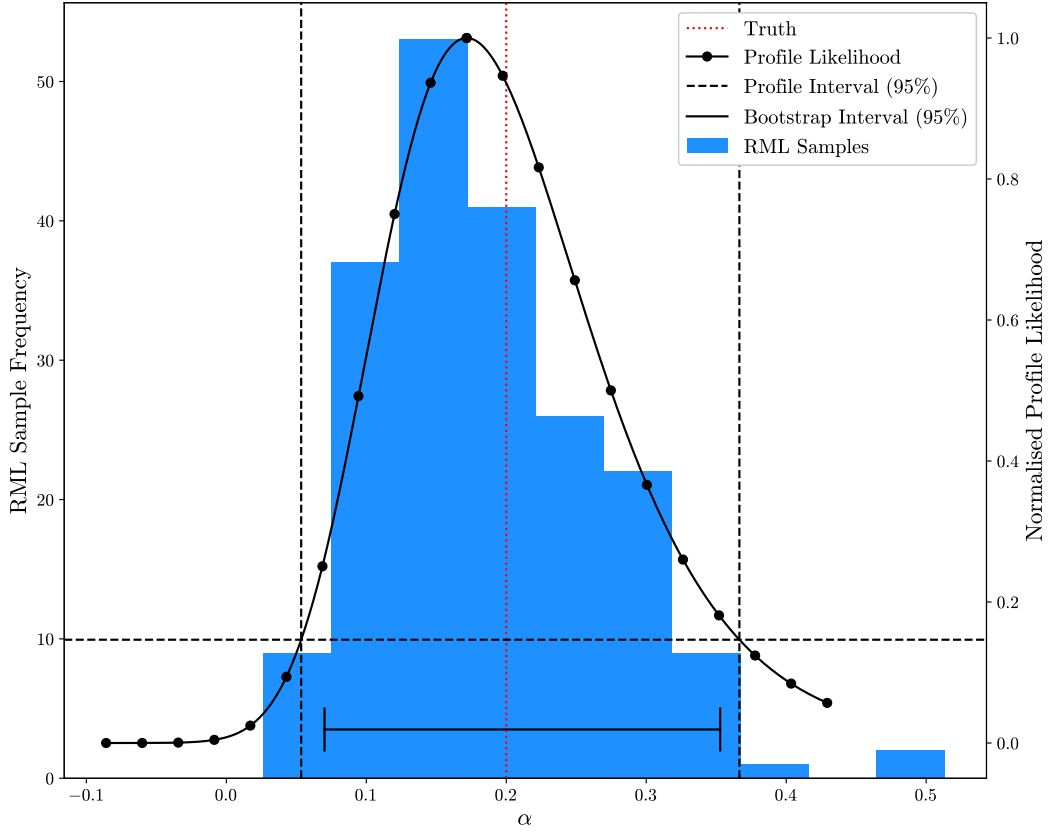
**Figure 3:** Profile likelihood and RML bootstrap samples ( $n = 200$ ) of  $\beta$  for the SIR model, with corresponding 95% intervals

$g(c)$  as  $\Psi c$ , where  $\Psi$  is constructed similarly to  $\Phi$ , by evaluating  $\mathbb{S}$  over the time points  $t_y$  of the data.  $f$  is the ODE model.

We choose  $W$  to be the Cholesky whitening matrices so that  $LL^T = \Gamma^{-1}$ ,  $MM^T = \Sigma^{-1}$ , and  $L$  and  $M$  are triangular. This allows for  $\log |\Gamma|^{-1/2} = \log |L|$  to be computed efficiently as the sum of the log of the diagonal elements of  $L$ . For the Samoan measles case study,  $\Gamma$  and  $\Sigma$  are constructed to be diagonal, with some structure that allows for different states (of the ODE model) to have different variances. We specify elements of (the diagonal)  $W$  so that states with similar sizes have the same model error. Specifically, we have  $\text{var}(S) = \text{var}(I_c) = \text{var}(R)$ ,  $\text{var}(E) = \text{var}(I)$ , and  $\text{var}(G) = \text{var}(H_c)$ . This is akin to the weights  $w_i$  that are applied in Hooker et al. [4].

We solve the problem iteratively with IRLS, using the IPOPT [5] plugin to CasADi to solve the objective, and then reweighting  $L$  and  $M$  so that they match the estimated optimal covariances at each iteration. Initial iterates for the diagonal elements of  $L$  and  $M$  were set at 1 and  $2 \times 10^{-2}$  respectively. Initial iterates of  $c$  were set at 10000, and  $\theta$  at 0.1. Iterations were repeated for  $N = 5$  iterations.

Profiling was done by gridding over  $[v\omega_{MLE}, (1+v)\omega_{MLE}]$  uniformly for some constant  $v$ , and evaluating over  $(\omega_{MLE}, v\omega_{MLE}]$  and  $(\omega_{MLE}, (1+v)\omega_{MLE}]$ , using the MLE as the initial iterate,



**Figure 4:** Profile likelihood and RML bootstrap samples ( $n = 200$ ) of  $\alpha$  for the SIR model, with corresponding 95% intervals

and computing each subsequent profile point by using the solution of the previous profile point. To compute the intervals, the profiled points are interpolated using a cubic spline implementation in Scipy. The intervals are computed using “normalised” likelihoods  $\mathcal{L}^*$  that are computed as

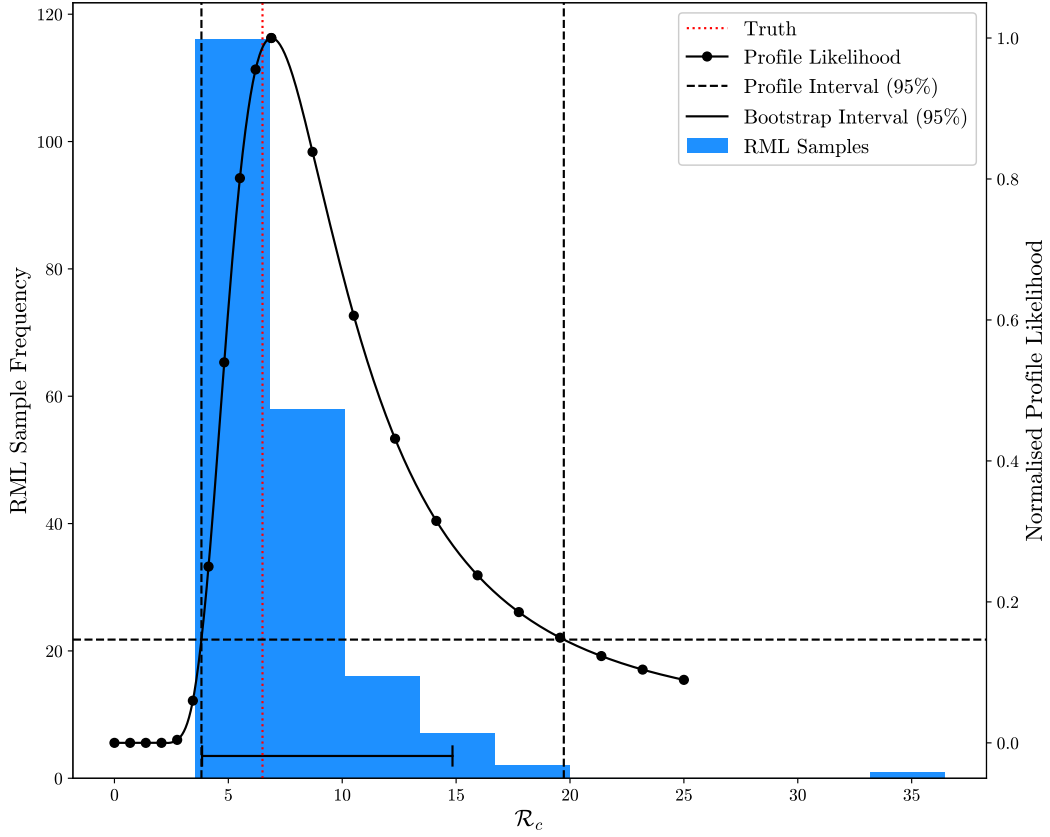
$$\mathcal{L}^*(\theta) = \frac{\mathcal{L}(\theta)}{\mathcal{L}(\hat{\theta})} = \exp\left(-\frac{1}{2}(\chi^2(\theta) - \chi^2(\hat{\theta}))\right) \quad (3)$$

and solving for

$$\mathcal{L}^*(\theta) = \exp\left(-\frac{1}{2}\chi_{(q, (1-\alpha))}^2\right) \quad (4)$$

where  $q$  is the number of degrees of freedom (1), and  $\alpha$  is the confidence level (0.95). For our values, this is  $\chi^2 = 3.84$ , or roughly  $\mathcal{L}^*(\theta) = 0.15$ .

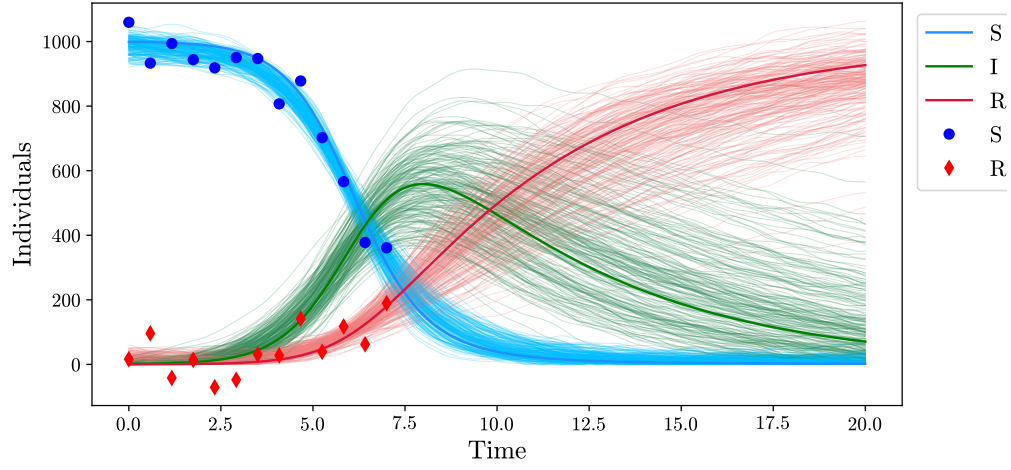
For the RML bootstrapping procedure, we use the Cholesky whitening matrices ( $W$ ) to generate the samples of  $e$  and  $\nu$ . This can be done simply by drawing  $N$  samples from the standard normal distribution  $Z \sim \mathcal{N}(0, I)$  and computing the samples  $\tilde{Y}$  by solving the linear



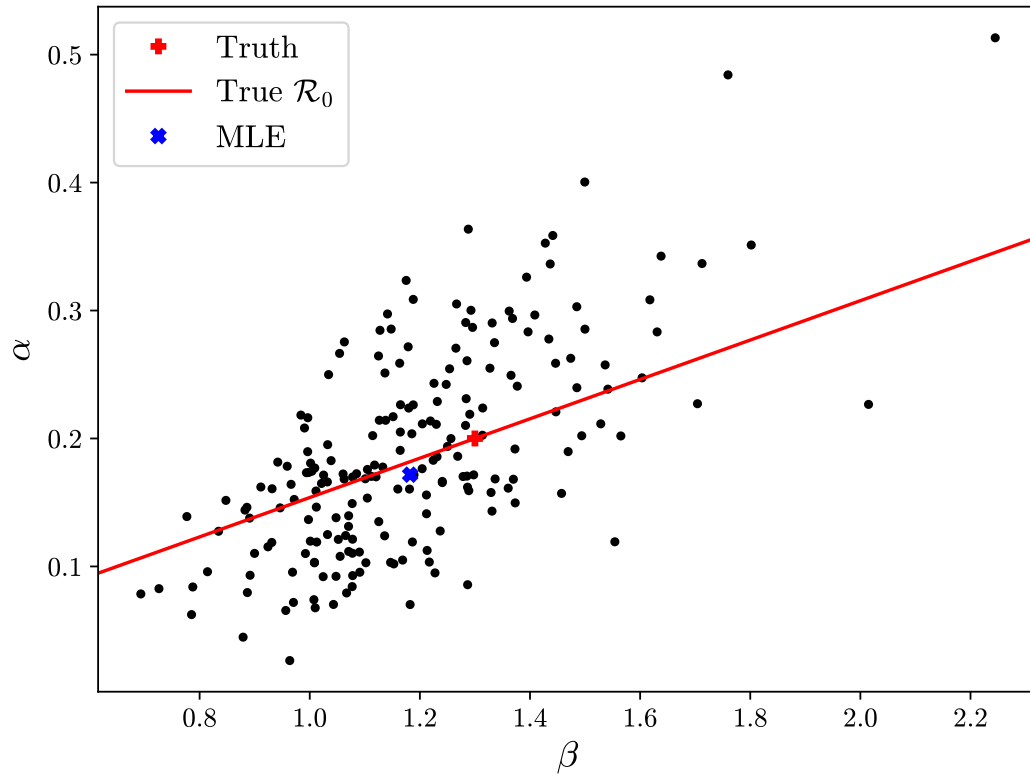
**Figure 5:** Profile likelihood and RML bootstrap samples ( $n = 200$ ) of  $\mathcal{R}_c$  for the SIR model, with corresponding 95% intervals

system  $W\tilde{Y} = Z$ . The samples are then fitted by minimising the objective function with the resampled data  $\tilde{Y}$  from the initial iterate of the MLE.

We found that implementing an RTO [6] sampling technique was not possible on our machine (Intel i7-7700 CPU, 15.5 GiB memory) due to the dense  $\bar{Q}$  matrix, which was too memory-intensive to algorithmically differentiate with the explicit **SX** representation in **CasADi**.



**Figure 6:** State estimates of the RML bootstrap samples for the SIR model ( $n = 200$ )



**Figure 7:** Parameters estimates of the RML bootstrap samples for the SIR model ( $n = 200$ ), with subset of parameters that correspond to the true  $\mathcal{R}_0$  highlighted.



## References

1. Hansen PC. Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review*. 1992 Dec;34(4):561–580. (doi:[10.1137/1034115](https://doi.org/10.1137/1034115)).
2. Hansen PC, O’Leary DP. The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems. *SIAM Journal on Scientific Computing*. 1993 Nov;14(6):1487–1503. (doi:[10.1137/0914086](https://doi.org/10.1137/0914086)).
3. Andersson JAE, Gillis J, Horn G, Rawlings JB, Diehl M. CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*. 2019;11(1):1–36. (doi:[10.1007/s12532-018-0139-4](https://doi.org/10.1007/s12532-018-0139-4)).
4. Hooker G, Ellner SP, Roditi LDV, Earn DJD. Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario. *Journal of the Royal Society, Interface*. 2011 Jul;8(60):961–74. (doi:[10.1098/rsif.2010.0412](https://doi.org/10.1098/rsif.2010.0412)).
5. Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*. 2006 Mar;106(1):25–57. (doi:[10.1007/s10107-004-0559-y](https://doi.org/10.1007/s10107-004-0559-y)).
6. Bardsley JM, Solonen A, Haario H, Laine M. Randomize-Then-Optimize: A Method for Sampling from Posterior Distributions in Nonlinear Inverse Problems. *SIAM Journal on Scientific Computing*. 2014;36(4):A1895–A1910. (doi:[10.1137/140964023](https://doi.org/10.1137/140964023)).