Likelihood-based estimation and prediction for a measles outbreak in Samoa

David Wu [a, *], Helen Petousis-Harris [b], Janine Paynter [b], Vinod Suresh [a, c], Oliver J. Maclaren [a]

[a] *Department of Engineering Science, University of Auckland, Grafton, Auckland, 1010, New Zealand*
[b] *Department of General Practice and Primary Health Care, University of Auckland, Grafton, Auckland, 1023, New Zealand*
[c] *Auckland Bioengineering Institute, University of Auckland, Grafton, Auckland, 1010, New Zealand*

ARTICLE INFO

ABSTRACT

Prediction of the progression of an infectious disease outbreak is important for planning and coordinating a response. Differential equations are often used to model an epidemic outbreak's behaviour but are challenging to parameterise. Furthermore, these models can suffer from misspecification, which biases predictions and parameter 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 justification for this approach by introducing a new interpretation of the model approximation component as a stochastic constraint. This preserves the rationale for using profiling rather than integration to remove nuisance parameters while also 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 it achieved relatively fast, accurate predictions. Here we present the most recent version of our method and its application to this measles outbreak, along with additional validation.

© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

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.

The models used to characterise the spread of infectious diseases are usually mechanistic models, and these are well-established in the mathematical epidemiological community (Anderson & May 1991). From the early deterministic models of Ross (1911) and Kermack and McKendrick (1927), to recent developments in both deterministic (Schenzle, 1984; Hooker et al., 2011; Xia et al., 2004) and stochastic models (Allen, 2008, 2011; Ferguson et al., 2020), 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 (Roda et al., 2020; Wilke & Bergstrom, 2020, pp. 28549–28551).

More widely, there are many methods for performing parameter inference for mechanistic models. However, 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 (He et al., 2010). Additionally, limitations in data collection, combined with complex model structures, may mean parameter estimates are not uniquely determinable, or *nonidentifiable* (Raue et al., 2009; Fröhlich et al., 2014). There is also the subtle problem of *model misspecification* (Kennedy & O'Hagan, 2001; Brynjarsdóttir & O'Hagan, 2014; King et al., 2015), where there are significant discrepancies between the model and the true data generating process. Such discrepancies 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 (King et al., 2015) and Bayesian frameworks (Chatzilena et al., 2019). 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 at least five times longer than deterministic counterparts (Chatzilena et al., 2019).

An alternative to explicit error modelling 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* (Ramsay et al., 2007). We introduce a variant of this approach under a new statistical interpretation and describe how this assists in 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 (Durbin, 1953; Theil & Goldberger, 1961). This can be thought of as a form of mixed estimation, or as a *prior likelihood* (Edwards, 1969) 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:

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

where f is some vector field, and θ 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:

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

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. However, Ramsay and Hooker (2017) point out that this formulation has a few difficulties. 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.

This formulation also 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 (Brynjarsdóttir & O'Hagan, 2014). One way to model this is to use an explicit stochastic process, for example a Gaussian process, to capture the misspecification (Kennedy & O'Hagan, 2001). 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:

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

where W_t represents a Wiener process (Law et al., 2015; Van Kampen, 2007). 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 (Van Kampen, 2007, §9.5).

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 probability distribution used in 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 (Bretó, 2018; Bretó et al., 2009; Chatzilena et al., 2019; Diggle & Gratton, 1984; Fasiolo et al., 2016; Hartig et al., 2011; Ionides et al., 2017; King et al., 2016; Wood, 2010). Discrete-time, and/or discrete-state, stochastic models are often used for modelling complex systems in biology and statistics (Durbin & Koopman, 2012; Wilkinson, 2018), 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. (2007). The generalised profiling method, also known as the parameter cascade method, was built on the methods of functional data analysis (Ramsay & Silverman, 2005), 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 (Ramsay et al., 2007; Ramsay & Hooker, 2017), 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 θ.

$$\hat{c}(\theta) = \operatorname*{arg\,min}_{c|\theta} \quad \|y - g(x)\|_2^2 + \lambda \int \left(\frac{dx}{dt} - f(x; \theta)\right)^2 dt, \tag{4}$$
$$\text{s.t.} \qquad x = \Phi c.$$

Here the state x is represented in terms of some basis comprising the columns of Φ, which is typically a basis of B-splines. This inner problem is tuned with the hyperparameter λ, which acts as a tradeoff between interpolation (when λ is low) and model matching (when λ is high). This produces some optimal value for c for the given θ, 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 at a chosen set of points in the time domain. The derivatives in the integral term can be directly determined by differentiating the spline representation. In contrast to standard derivative matching, this hybrid approach allows for unobserved states to be dealt with and 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 (Ramsay & Hooker, 2017; Ramsay & Silverman, 2005). As discussed by Wahba (1990); Ruppert et al. (2003), 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 (Ramsay et al., 2007). Ramsay et al. (2007); Ramsay and Hooker (2017), however, stress that they do not view their estimates as

arising from an explicit stochastic process model. One aim of the present work is to further develop the interpretation of the work done by Ramsay et al. (2007).

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* (Pawitan, 2001). Hence, it appears that the outer objective is intended as the profile likelihood of the inner objective, with respect to θ, though this is not made explicit in the original paper by Ramsay et al. (2007). However, if we are to take this interpretation, questions arise as to why the outer objective is different to the inner objective. 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

$$y = G\Phi c + e,$$
$$\mathbb{E}[e] = 0, \ \text{Cov}[e] = \Gamma,$$

(6)

where G is a linear observation operator, Φ consists of a basis with a large number of basis functions ('columns' of Φ) relative to the number of observations ('rows' of G), and e is observational noise with mean 0 and covariance Γ. 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 (Wahba, 1990; Ruppert et al., 2003) or, in non-Bayesian inference, in terms of random-effects models (Ruppert et al., 2003). 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 (1961), building on the work of Durbin (1953). This amounts to considering, in addition to Equation (6), the additional linear regression equation for another observable random variable r:

$$r = H\Psi c + v,$$
$$\mathbb{E}[v] = 0, \ \text{Cov}[v] = \Sigma$$

(7)

Here H is a linear observation operator, Ψ consists of a potentially different set of basis functions, and v is observational noise with mean 0 and covariance Σ. 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 (1961) 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 (1961) 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 (1969), 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 (1961) can be found in Rao et al. (2008). H-likelihood (Lee et al., 2018; Lee & Nelder, 1996) is another related, though distinct, idea.

The above equations can be 'stacked' to give

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} G\Phi \\ H\Psi \end{bmatrix} c + \begin{bmatrix} e \\ v \end{bmatrix}$$

(8)

This can be solved using the generalised least squares method, using the notation $\|f\|_A^2 := f^T A f$ to represent the (square of the) Mahalanobis distance (Mahalanobis, 1936):

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

(9)

which has the usual explicit solution

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

(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 Γ and Σ respectively ($L^T L = \Gamma^{-1}, M^T M = \Sigma^{-1}$). $\tilde{G}^T \tilde{G}$ is invertible if the intersection of the nullspaces of $LG\Phi$ and $MH\Psi$ is the empty set (Engl et al., 1996).

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

$$r = L(\theta)x \tag{11}$$

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 θ, 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

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

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 $m(t)$.

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

$$r = L(\theta)x + \Sigma DW, \tag{13}$$

where DW stands for the (formal) time derivative of a multi-dimensional Wiener process (Law et al., 2015). 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 (Kloeden & Platen, 2013) for the derivatives. This amounts to taking, for the ith 'observation' (discretisation grid point) of the equation,

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

Multiplying this by Δt to put it in differential form gives the standard Euler-Maruyama discretisation. This interpretation of the stochastic constraint can be captured in the least squares problem by setting the weighting function $m(t)$ from Equation (12) as

$$\begin{aligned}\int (m(r - L(\theta)\Phi c))^2 dt &\approx (r - L(\theta)\Phi c)\left(\frac{1}{\Delta t}\Sigma^{-1}\right)(r - L(\theta)\Phi c)^T \\ &= \|r - L(\theta)\Phi c\|_{(\Sigma\Delta t)^{-1}}^2.\end{aligned} \tag{15}$$

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 θ 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, $l(\theta, c)$:

$$\min_{\theta, c} \quad l(\theta, c) = \|y - g(x)\|_{\Gamma^{-1}}^2 + \int (m(Dx - f(x; \theta) - r))^2 dt$$
$$\text{s.t.} \qquad x = \Phi c \tag{16}$$

We will in general take $r = 0$ and hence drop this from now. As discussed above, we can approximate the second term using an Euler-Maruyama discretisation, for some choice of time step Δt, and write:

$$\min_{\theta, c} \quad l(\theta, c) = \|y - g(x)\|_{\Gamma^{-1}}^2 + \|Dx - f(x; \theta)\|_{\hat{\Sigma}^{-1}}^2$$
$$\text{s.t.} \qquad x = \Phi c \tag{17}$$

where $\hat{\Sigma} = \Sigma\Delta t$.

We can also add additional prior likelihood terms in the same manner, alongside the differential equation term. We consider additional covariance terms in more detail in the following subsection. First we relate the above version to the standard generalised profiling approach.

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 (Pawitan, 2001). For example, denoting the overall objective by $l(\theta, c; \Gamma, \Sigma)$, for fixed Γ, Σ and for the constraint $x = \Phi c$ imposed, our problem can be decomposed into two analogous stages:

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

While the inner optimisation step is the same here as the standard formulation Equation (4), our outer optimisation retains the model misfit term. By defining

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

our outer optimisation then corresponds to

$$\min_{\theta} \ \ l_p(\theta, \hat{c}(\theta); \Gamma, \Sigma). \tag{20}$$

Existing literature using the generalised profiling method (Campbell & Chkrebtii, 2013; Cao & Ramsay, 2007; Hooker et al., 2011; Ramsay et al., 2007; Xun et al., 2013) appears to focus on parameter inference, and neglect the predictive properties of the method. Some work has been done in prediction in (Hooker et al., 2011), but in the context of tuning the smoothing hyperparameter λ 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 λ, 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 the discussants of Ramsay et al. (2007). 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 prediction can be done as part of the inference process, with no further decisions required. 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 λ can be done efficiently in the linear case, through the application of the generalised cross-validation criterion (Wahba, 1990), 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 (Campbell & Chkrebtii, 2013; Ramsay & Hooker, 2017).

In our reformulation, the estimation of the covariances of the errors are analogous to the tuning of λ. Hence, unlike the standard least squares form of generalised profiling, here covariance-related terms in the negative log-likelihood are not dropped from the final form of our objective. Following the above assumptions about the errors in the model, we get, for Gaussian errors, the following negative log-likelihood objective function:

$$l(\theta, x) = \|y - g(x)\|_{\Gamma^{-1}}^2 + \|Dx - f(x; \theta)\|_{\hat{\Sigma}^{-1}}^2 - \frac{1}{2}\log |\Gamma| - \frac{1}{2}\log |\hat{\Sigma}|. \tag{21}$$

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 (Carroll & Ruppert, 1988), which alternate between estimation of the parameters and covariances. These are early-stopping versions of iteratively-reweighted least squares, but the optimal number of iterations N to take are hard to determine in most cases, and empirical experiments done by Carroll and Ruppert (1988) show that $N \geq 2$ is advised. We also observe that in our formulation, there is a potential degeneracy in the model fit term, due to it being able to be exactly satisfied, leading the estimated model fit covariance to unboundedly decrease. We counter this behaviour by setting restrictions on the maximum or minimum estimated covariances of each term. 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 co-variances use (typically normal-theory) maximum likelihood or related methods (Carroll & Ruppert, 1988; Ruppert et al., 2003). 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, and we expand on the details of the implementation in the Supplementary Material. For brevity, and alignment with the objects in the code, we introduce the decision variables $\xi = [c, \theta]$ as the stacked spline coefficients and model parameters, and the weights $w = [vec(L), vec(M)]$ which are the stacked

vectorised lower triangular entries of the whitening matrices L and M associated with the covariance matrices ($L^T L = \Gamma^{-1}$, $M^T M = \Sigma^{-1}$).

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 ϵ,
    minimum weight w_{min}, maximum weight w_{max}

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

1 **begin**
2   initialisation
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         $w_{prop} \leftarrow \arg\min_w l(x_i, w)$
7         $w_i \leftarrow \max(w_{min}, \min(w_{prop}, w_{max}))$
8         $\Delta f_i \leftarrow f_i - f_{i-1}$
9   **if** $\Delta f_i < -\epsilon$ **then**
             // 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 θ. 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 (Pawitan, 2001).

An alternative way to perform analysis on the uncertainty of quantities of interest is using the profile likelihood (Pawitan, 2001; Ionides et al., 2017). We can apply typical frequentist sampling theory methods to form approximate 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

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

where $\chi^2_{1,(1-a)}$ 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 (Pawitan, 2001)):

$$\left\{ \omega, \frac{\mathcal{L}_p(\omega)}{\mathcal{L}_p(\hat{\omega})} > k \right\} \tag{22}$$

where ω is the quantity of interest, $\mathcal{L}_p(\omega)$ is the profile likelihood of ω, and $\mathcal{L}_p(\hat{\omega})$ is the profile likelihood of ω evaluated at the MLE, for which $\omega = \hat{\omega}$. This confidence interval can also inform the identifiability of the quantity of interest (Raue et al., 2009; Simpson et al., 2020). This process can be carried out by re-solving the maximum likelihood estimation procedure over a grid of values for ω, with an additional equality constraint for those values of ω. We stress here that ω is not necessarily a model parameter, but can be any function of the model parameters θ or state variables c (Kreutz et al., 2012). This means that we can capture uncertainty in the trajectories by profiling over state quantities, for example. Although this process can be slow, as the profiling procedure requires solving a number of least-squares problems, we can assume that value of the profile at ω_i is close to the value of the profile at ω_{i+1}, and thus can initialise the problem at ω_{i+1} at the solution at ω_i.

Another approach is to use a bootstrap-style sampling distribution (Davison & Hinkley, 1997; Efron, 1979), determined by repeatedly solving an augmented version of the optimisation problem of Equation (21):

$$\min_{\theta,c} \quad \|y - g(x) + e\|^2_{\Gamma^{-1}} + \|Dx - f(x;\theta) + \nu\|_{\hat{\Sigma}^{-1}}^2, \text{s.t.} x = \Phi c, \tag{23}$$

for a series of realisations of e and ν, which have zero mean and covariances Γ and $\hat{\Sigma}$. The values of these covariances can be extracted from the weights of solution at the maximum likelihood estimate. This bootstrapping procedure can be done efficiently by solving from the maximum likelihood estimate and the problems induced by each realisation of e and ν can be carried out in perfect parallel. In the inverse problems literature, the above approach falls under the umbrella of Randomised Maximum Likelihood (RML) (Oliver et al., 1996) methods that include randomised Maximum a Posteriori (rMAP) (Wang et al., 2018) and Randomize-then-Optimize (RTO) (Bardsley et al., 2014). 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 (1961), or a prior likelihood in the style of Edwards (1969). 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 (Efron, 1979):

$$\left[t^*_a, t^*_{(1-a)} \right] \tag{24}$$

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 ~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 (Government of Samoa, 2019). Initially these were published by the Ministry of Health, but during the emergency period the function 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 Fig. 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 (Petousis-Harris, 2019; Russell, 2019). For our analysis here, we use data up to the 3rd of December 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?
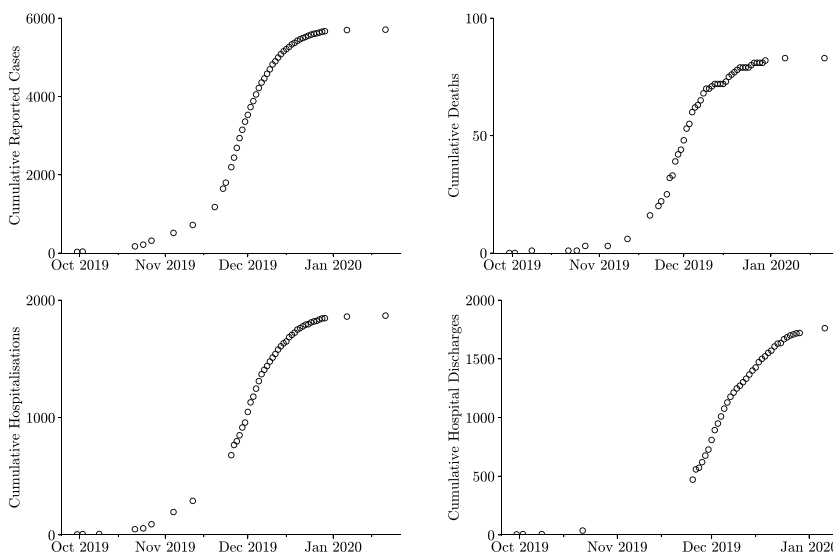
4.1. Model

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

$$\dot{S} = -\beta S \frac{I}{N}, \tag{25a}$$

$$\dot{E} = \beta S \frac{I}{N} - \gamma E, \tag{25b}$$

$$\dot{I} = \gamma E - (\alpha + \eta)I, \tag{25c}$$

Fig. 1. Plots of the available data for the 2019 Samoan measles outbreak. Top-left: cumulative reported cases; top-right: cumulative reported deaths; bottom-left: cumulative reported hospitalisations; bottom-right: cumulative reported hospital discharges. Current hospitalisation data is also available, but not shown.

$$\dot{R} = \alpha I, \tag{25d}$$

$$\dot{H} = \eta I - (\delta + \mu)H, \tag{25e}$$

$$\dot{G} = \delta H, \tag{25f}$$

$$\dot{D} = \mu H, \tag{25g}$$

$$\dot{I}_c = \gamma E, \tag{25h}$$

$$\dot{H}_c = \eta I. \tag{25i}$$

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 γ, and report their infection (tracked by I_c). Infectious individuals can then recover to R at rate α, become hospitalised (H) at rate η. Hospitalisation is tracked by H_c, and hospitalised individuals recover and are discharged (G) at rate δ, or die (D) at rate μ. 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 (van den Driessche & Watmough, 2002, 2008) as

$$\mathcal{R}_c = \frac{\beta}{\alpha + \eta} \frac{S(t = 0)}{N(t = 0)} \tag{26}$$

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

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

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 (Tuncer & Le, 2018). In this analysis, we also relax the implicit assumption in (Tuncer & Le, 2018; Roosa & Chowell, 2019) that the state at $t = 0$ is known, in particular the ratio $S(t = 0)/N(t = 0)$.

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 December 3, 2019, is presented in Fig. 2.

We add two prior likelihood/regularisation terms that penalise the magnitudes of the parameters β and γ. We tune the hyperparameters on the two penalty terms by analysis of synthetic data generated from the model (see the Supplementary Material for tuning and regularisation details). These terms penalise pathological behaviour where the E or I state estimates converge to small values where their derivative are small, as encouraged by the model misfit term.

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. The total reported cases and deaths are taken at the final time point of the state estimate, which is set to $1.25 \times$ the final time point (when the epidemic was declared over by the Samoan Government), where dates are zeroed with respect to the date of discovery of suspected first case (September 30, 2019). We draw 200 samples for the RML bootstrap method, and reject any samples that result in non-convergent optimisation behaviour ($n = 22$). Intervals are computed for a 95% confidence level, and a summary of results is presented in Table 2.

We see that we capture the truth for the total number of cases with both profile likelihood and bootstrap intervals. The bootstrap sample distribution and profile likelilhood are presented in Fig. 3, and Fig. 4 plots the RML state estimates for the total reported cases over time.

We see that the intervals for $\mathcal{R}_c$ by both methods roughly agree, though the bootstrap interval is slightly narrower, as shown in Fig. 5.

We also construct a joint profile of total cases and $\mathcal{R}_c$ in Fig. 6. We see that the RML samples and the 95% confidence region agree. The univariate profiles of $\mathcal{R}_c$ and total cases follow the likelihood surface of the bivariate profile, as expected (Bolker, 2008).

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 Fig. 7, and derive an 95% bootstrap confidence interval of [104–120] and a profile interval of [102–120]. We note that these bounds do not capture the true value of 83 deaths. This is likely due to the drop-off in deaths just outside of range of the provided data (see Fig. 1), which cannot be captured by a single mortality rate parameter.

5. Discussion and conclusion

The estimation and prediction of epidemics is a difficult task, in both theory and practice. In this paper, we have presented a likelihood-based reformulation of the standard generalised profiling method introduced and developed by Ramsay et al. (2007). This is based on ideas of mixed estimation in the sense of Theil and Goldberger (1961) and prior likelihood in the sense of Edwards (1969). This 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 θ-profile of the inner objective, albeit evaluated at $\lambda \to \infty$, where the model is exactly enforced. Our reformulation can thus be seen as more natural
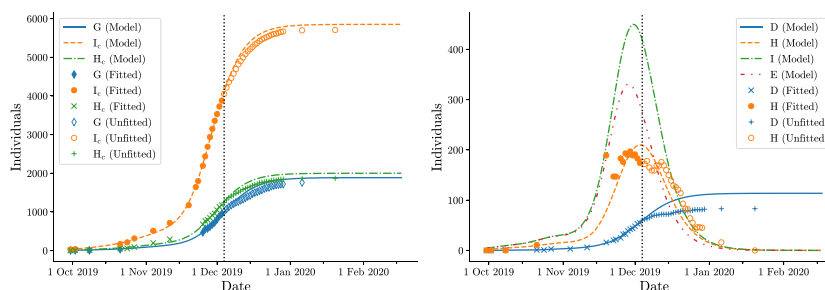

(a) Cumulative reports (I_c), cumulative hospitalisations (H_c) and discharges (G)

(b) Deaths (D), incidence of non-infectious latent individuals(E) and infectious individuals (I), and hospitalisations (H)

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

Table 2
Point and interval estimates of quantities of interest for the Samoan measles case study.

Type	Quantity		MLE	Profile	Bootstrap	Truth
	Interpretation	Symbol				
Parameters	Infectivity	β	0.870	–	[0.813–0.920]	–
	Incubation	γ	0.717	–	[0.657–0.772]	–
	Recovery	α	0.350	–	[0.316–0.380]	–
	Hospitalisation	η	0.176	–	[0.162–0.186]	–
	Discharge	δ	0.306	–	[0.295–0.318]	–
	Death	μ	0.0186	–	[0.0173–0.0200]	–
State Values	Total Cases		5785	[5593–6006]	[5621–6043]	5707
	Total Deaths		111	[102–120]	[104–120]	83
Quantities	Number	$\mathcal{R}_c$	1.648	[1.586–1.717]	[1.596–1.712]	–
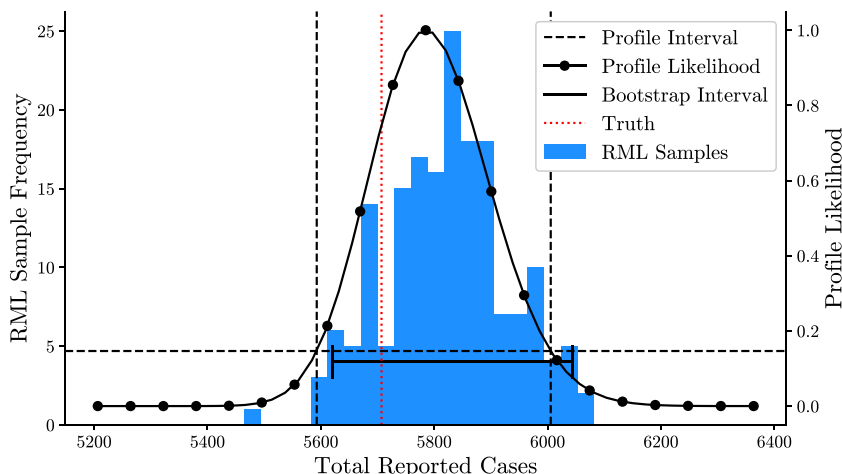
Fig. 3. Profile likelihood and RML bootstrap samples ($n = 178$) of the total reported cases, with corresponding 95% confidence intervals marked.

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 the profile likelihood (Pawitan, 2001) and RML bootstrapping that allow us to perform uncertainty quantification 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 quantities of interest, like $\mathcal{R}_c$ and the total number of cases.

5.1. Profile likelihood 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 (Ruppert et al., 2003; Hodges, 2013), which embed spline models in mixed effect models using random effects. These approaches also relate closely to Bayesian interpretations and the work of Wahba (1990) 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 statistical rationale for using profiling (maximisation) rather than marginalisation for eliminating nuisance parameters. Even in the fixed parameter setting, there is ongoing debate whether profiling or marginalisation better captures properties of likelihood when eliminating nuisance parameters (Aitkin, 2010; Aitkin & Stasinopoulos, 1989; Berger et al., 1999).

When estimating covariances with our approach, there is still the open question of an optimal stopping criterion, or optimal number of cycles, to take in the IRLS algorithm. We find that even if the objective function terms are balanced, the solution tends to diverge without regularisation. In this case study, we opted to apply bound constraints on the covariance matrices Γ and Σ as a regularisation, preventing the solution from diverging towards pure data interpolation or degenerate solutions of the underspecified differential equation model. Currently the specification of these bounds is done heuristically − through analysing the behaviour of the weights at successive iterations and performing synthetic data studies − but there needs to be further work done into characterising appropriate methods for determining these values. Furthermore, it is unclear how well covariance matrices can be estimated in this setting, reflecting similar difficulties noted in the mixed effects literature (Hodges, 2013). Validation methods have been used with generalised profiling, such as the forwards prediction
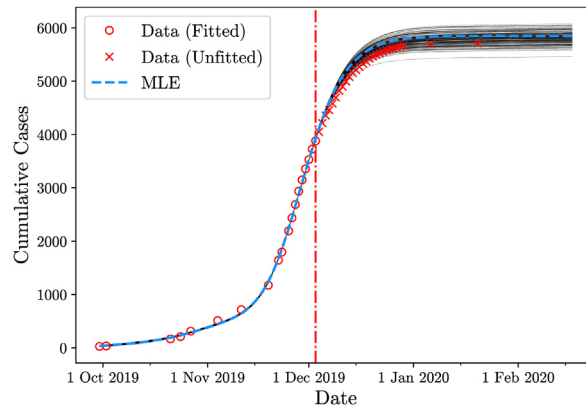
Fig. 4. Timeseries of RML bootstrap samples of total reported cases ($n = 178$).
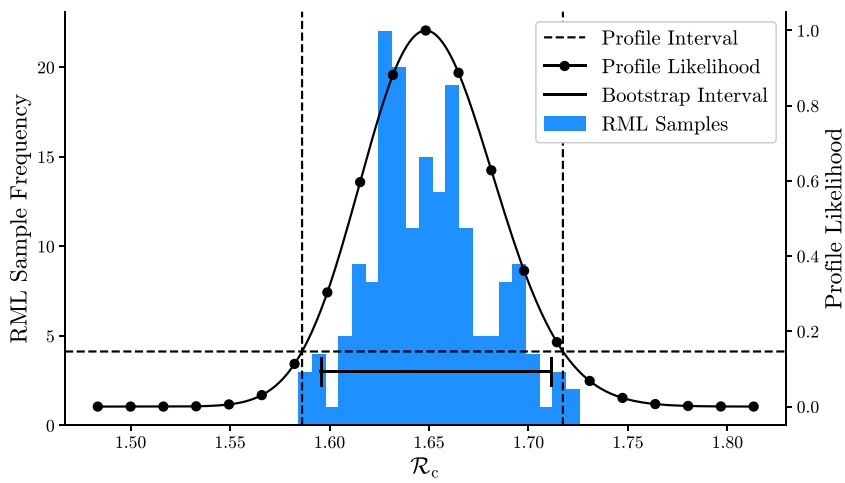

Fig. 5. Profile likelihood and RML bootstrap samples ($n = 178$) of $\mathcal{R}_c$, with associated 95% confidence intervals marked.
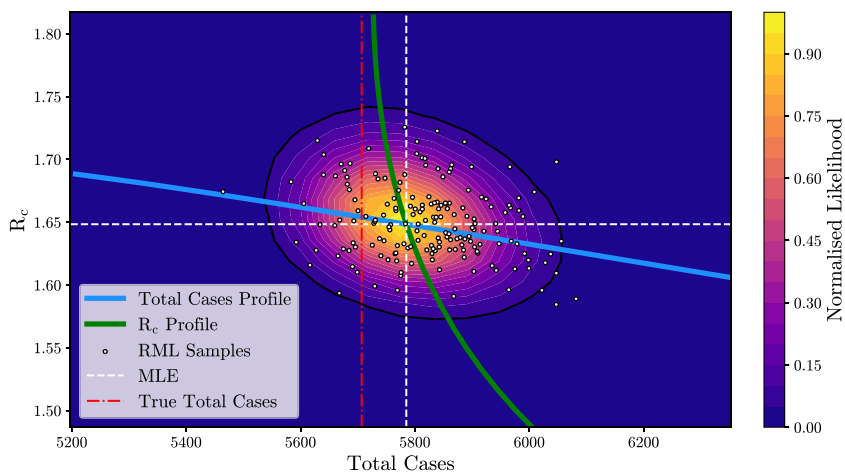

Fig. 6. Joint profile of total cases and $\mathcal{R}_c$, with independent profiles of each variable and the RML samples overlaid. The contour highlighted in black represents the boundary of the 95% confidence region, where the normalised likelihood value is approximately 0.15.
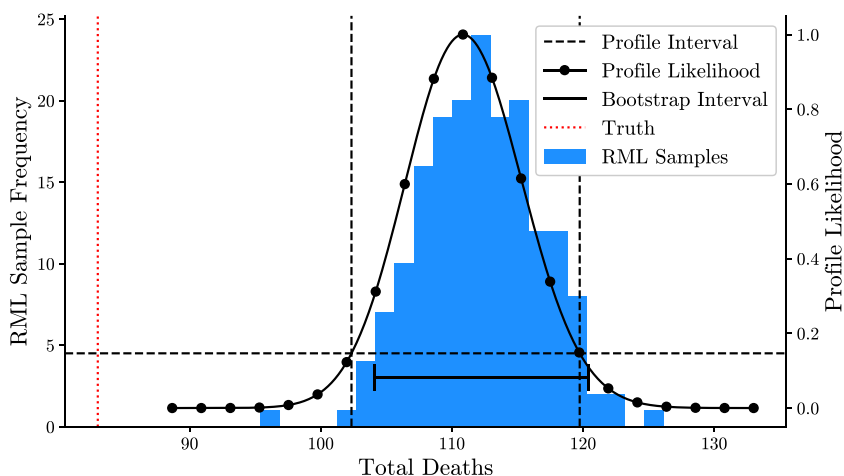
Fig. 7. Profile likelihood and RML bootstrap samples ($n = 178$) of the total deaths with associated 95% confidence intervals marked.

error suggested by (Hooker & Ellner, 2010), but there are problems when using this for ongoing or 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. (2020) have considered this for their Bayesian reinterpretation of the standard formulation of generalised profiling, setting priors on the tuning parameter λ. 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 (Hodges, 2013). In general, estimating these hyperparameters remains a key challenge for all approaches. In this particular study, we choose bounds that lead to stable iterates in the maximum likelihood estimation procedure, and use a synthetic data study (see Supplementary Materials) to determine an appropriate magnitude for the parameter regularisation strength.

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 that $\mathcal{R}_c$ is identifiable in our analysis, which is consistent with the findings of Roosa and Chowell (2019) that show that $\mathcal{R}_0$ is identifiable even if there are identifiability issues with the other parameters. We do note however, that the identifiability of the individual parameters is only possible due to the regularisation terms added, as prior likelihoods, to the likelihood function, and without them, there is evidence that there is practical nonidentifiability of model parameters β and γ. This is likely related to the relaxation of the assumption that the initial state is known, since both are related to $\mathcal{R}_c$, where the initial condition appears (through $\dot{S}$ and $\dot{E}$).

Further, the coverage properties of the intervals we have constructed with regularisation are not well-known. Work in post-selection inference is formalising the properties of estimators that are regularised, or data-informed in some ways, in particular for methods like lasso (Hastie et al., 2015). There have also been alternative formulations that rely on imposing physically-motivated constraints on the problem to construct confidence intervals with appropriate coverage properties (Mikael Kuusela & Philip B. Stark, 2017). In our work, we interpret the regularisation terms as prior likelihoods, or additional data. This means the properties of our confidence intervals can be more directly justified in terms of the relevant joint sampling distribution involving both the present study data, and our prior regularisation data. This introduces a dependence on the quality of the prior data, similar to the dependence of Bayesian credible intervals on the prior distribution. Thus, there still remains the question of whether the approximate intervals we compute have close to the correct coverage, and how the sampling distribution should depend on the auxiliary data. The simple SIR experiment (in the Supplementary Material) suggests that the $\mathcal{R}_c$ profile should be skewed without the presence of prior regularisation data, whereas we see a more symmetrical profile when that additional regularisation data is added in the case study.

We find that the estimation procedure is sensitive to initial guesses of the weightings (covariance matrices), especially for this data observation window, which ends around the peak of the epidemic. This manifests itself numerically as non-convergence after a large number of optimisation iterations. In the Supplementary Material, we perform a synthetic study of observation windows for a standard SEIR model, and see that there is a sudden regime change in the estimation errors (for both state and parameters) once the observation window passes the turning point of the outbreak curve. This is consistent with the findings of Wilke and Bergstrom (2020, pp. 28549–28551) and Roda et al. (2020), which show that the predictive uncertainty window shrinks rapidly as data is collected past the turning point of the epidemic curve. This is probably linked to

the inability to determine the rate of recovery in the early stages of the epidemic, since the effect of recovery is small on the overall behaviour. However, as it is often difficult to determine this turning point, we would suggest that practitioners be wary of these observability-induced errors when performing prediction.

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 total number of deaths, 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, 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. (2011) allow for a cyclical β 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 parameters such as γ (the latent period) and α (the recovery rate). Although we have regularised β and γ towards 0 to make the estimation procedure better-behaved, this additional information could be used instead. In this way, this method can be seen as analogous to the Bayesian formulation, but instead of marginalising over the prior data, we maximise (profile) over the prior.

5.4. Practical considerations

In our implementation of the method, we have used the CasADi framework (Andersson et al., 2019) and its interface in Python. We found that the estimation of the MLE for the Samoan measles case study took around 3 min. The profiles each take around 10 min to generate, dependent on the number of points to evaluate, and the 200 RML bootstrap samples take around 30 min. Most of these timings 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 another. 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 there being limited regions of convergence for any given optimisation problem, requiring chaining to generate suitable profiles. 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. 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. We have tried using restarts from surrounding iterates, and see that in some cases this can mitigate the problem, but this is not guaranteed. 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. For example, we found that an RTO-style approach (Bardsley et al., 2014), where the Jacobian must be used in the log-likelihood when resampling for a bootstrap, is infeasible on a workstation with 16 GB of RAM. Alternative implementations of these methods may significantly reduce these memory issues, with a less harsh runtime penalty than we found using CasADi.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2023.01.007.

References

Aitkin, M. (2010). Statistical inference: An integrated Bayesian/likelihood approach. In *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC.

Aitkin, M., & Stasinopoulos, M. (1989). Likelihood analysis of a binomial sample size problem. In L. J. Gleser, M. D. Perlman, S. J. Press, & A. R. Sampson (Eds.), *Contributions to probability and statistics: Essays in honor of ingram olkin* (pp. 399–411). Springer. https://doi.org/10.1007/978-1-4612-3678-8_28.

Allen, L. J. S. (2008). An introduction to stochastic epidemic models. In F. Brauer, P. van den Driessche, & J. Wu (Eds.), *Mathematical epidemiology chapter 4* (pp. 81–130). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-78911-6_3.

Allen, L. J. S. (2011). *An introduction to stochastic processes with applications to biology*. Chapman & Hall/CRC.

Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: Dynamics and control*. Oxford: Oxford University Press.

Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., & Diehl, M. (2019). CasADi — a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation, 11*, 1–36. https://doi.org/10.1007/s12532-018-0139-4

Bardsley, J. M., Solonen, A., Haario, H., & Laine, M. (2014). Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Scientific Computing, 36*, A1895–A1910. https://doi.org/10.1137/140964023

Berger, J. O., Liseo, B., & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science, 14*, 1–28. https://doi.org/10.1214/ss/1009211804

Bolker, B. M. (2008). *Ecological models and data in R*. Princeton University Press.

Bretó, C. (2018). Modeling and inference for infectious disease dynamics: A likelihood-based approach. *Statistical Science: A review journal of the Institute of Mathematical Statistics, 33*, 57.

Bretó, C., He, D., Ionides, E. L., King, A. A., et al. (2009). Time series analysis via mechanistic models. *Annals of Applied Statistics, 3*, 319–348.

Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems, 30*, Article 114007. https://doi.org/10.1088/0266-5611/30/11/114007

Campbell, D., & Chkrebtii, O. (2013). Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates. *Mathematical Biosciences, 246*, 283–292. https://doi.org/10.1016/J.MBS.2013.03.011

Cao, J., & Ramsay, J. O. (2007). Parameter cascades and profiling in functional data analysis. *Computational Statistics, 22*, 335–351. https://doi.org/10.1007/s00180-007-0044-1

Carroll, R. J., & Ruppert, D. (1988). *Transformation and weighting in regression*. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC.

Chatzilena, A., van Leeuwen, E., Ratmann, O., Baguelin, M., & Demiris, N. (2019). Contemporary statistical inference for infectious disease models using Stan. *Epidemics*. , Article 100367. https://doi.org/10.1016/J.EPIDEM.2019.100367

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application. Cambridge series in statistical and probabilistic mathematics*. Cambridge University Press.

Diggle, P. J., & Gratton, R. J. (1984). Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society: Series B, 46*, 193–212.

van den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences, 180*, 29–48. https://doi.org/10.1016/S0025-5564(02)00108-6

van den Driessche, P., & Watmough, J. (2008). Further notes on the basic reproduction number. In F. Brauer, P. van den Driessche, & J. Wu (Eds.), *Mathematical epidemiology* (pp. 159–178). Berlin, Heidelberg: Springer.

Durbin, J. (1953). A note on regression when there is extraneous information about one of the coefficients. *Journal of the American Statistical Association, 48*, 799–808.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

Edwards, A. W. F. (1969). Statistical methods in scientific inference. *Nature, 222*, 1233–1237.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1–26.

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems volume 375 of Mathematics and its applications*. Kluwer Academic Publishers.

Fasiolo, M., Pya, N., & Wood, S. N. (2016). A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Statistical Science*, 96–118.

Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L., Van Elsland, S., … Ghani, A. (2020). *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*. Technical Report Imperial College London. https://doi.org/10.25561/77482

Fröhlich, F., Theis, F. J., & Hasenauer, J. (2014). Uncertainty analysis for non-identifiable dynamical systems: Profile likelihoods, bootstrapping and more. In P. Mendes, J. O. Dada, & K. Smallbone (Eds.), *Computational methods in systems biology* (pp. 61–72). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-12982-2_5.

Government of Samoa. (2019). *Update on the measles outbreak*. URL [Press Release, Multiple] https://reliefweb.int/updates?advanced-search=(S4201_F8_D49263.

Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models — theory and application. *Ecology Letters, 14*, 816–827.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations* (1st ed.). Boca Raton: Chapman and Hall/CRC.

He, D., Ionides, E. L., & King, A. A. (2010). Plug-and-play inference for disease dynamics: Measles in large and small populations as a case study. *Journal of The Royal Society Interface, 7*, 271–283. https://doi.org/10.1098/rsif.2009.0151

Hodges, J. S. (2013). *Richly parameterized linear models: Additive, time series, and spatial models using random effects*. CRC Press.

Hooker, G., & Ellner, S. P. (2010). *On forwards prediction error*. Technical Report Cornell University.

Hooker, G., Ellner, S. P., Roditi, L. D. V., & Earn, D. J. D. (2011). Parameterizing state-space models for infectious disease dynamics by generalized profiling: Measles in Ontario. *Journal of The Royal Society Interface, 8*, 961–974. https://doi.org/10.1098/rsif.2010.0412

Huang, H., Handel, A., & Song, X. (2020). A Bayesian approach to estimate parameters of ordinary differential equation. *Computational Statistics*, 1–19. https://doi.org/10.1007/s00180-020-00962-8

Ionides, E. L., Breto, C., Park, J., Smith, R. A., & King, A. A. (2017). Monte Carlo profile confidence intervals for dynamic systems. *Journal of The Royal Society Interface, 14*.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B, 63*, 425–464. https://doi.org/10.1111/1467-9868.00294

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences, 115*, 700–721. https://doi.org/10.1098/rspa.1927.0118

King, A. A., Domenech de Cellès, M., Magpantay, F. M., & Rohani, P. (2015). Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to ebola. *Proceedings of the Royal Society B: Biological Sciences, 282*, Article 20150347.

King, A. A., Nguyen, D., & Ionides, E. L. (2016). Statistical inference for partially observed markov processes via the r package pomp. *Journal of Statistical Software, 69*.

Kloeden, P. E., & Platen, E. (2013). *Numerical solution of stochastic differential equations* (Vol. 23). Springer Science & Business Media.

Kreutz, C., Raue, A., & Timmer, J. (2012). Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Systems Biology, 6*, 120. https://doi.org/10.1186/1752-0509-6-120

Kuusela, M., & Stark, P. B. (2017). Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. *Annals of Applied Statistics, 11*, 1671–1710. https://doi.org/10.1214/17-AOAS1053

Law, K., Stuart, A., & Zygalakis, K. (2015). *Data assimilation: A mathematical introduction* (Vol. 62). Springer.

Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B, 58*, 619–656.

Lee, Y., Nelder, J. A., & Pawitan, Y. (2018). *Generalized linear models with random effects: Unified analysis via H-likelihood* (Vol. 153). CRC Press.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India, 2*, 49–55.

Oliver, D. S., He, N., & Reynolds, A. C. (1996). Conditioning permeability fields to pressure data. In *ECMOR V-5th European conference on the mathematics of oil recovery* (earthdoc.org).

Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood*. Oxford University Press.

Petousis-Harris, H. (2019). *Samoa's devastating measles epidemic – why and how bad?*. URL: https://sciblogs.co.nz/diplomaticimmunity/2019/11/28/samoas-devastating-measles-epidemic-why-and-how-bad/, 28 Nov 2019.

Ramsay, J., & Hooker, G. (2017). *Dynamic data analysis*. Springer Series in Statistics. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-7190-9.

Ramsay, J. O., Hooker, G., Campbell, D., & Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society: Series B, 69*, 741–796. https://doi.org/10.1111/j.1467-9868.2007.00610.x

Ramsay, J., & Silverman, B. W. (2005). *Functional data analysis*. Springer Science & Business Media.

Rao, C. R., Toutenburg, H., Shalabh, H. C., & Schomaker, M. (2008). *Linear model and generalization, least squares and alternatives*. Springer Series in statistics (3rd ed.). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-74227-2.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics, 25*, 1923–1929. https://doi.org/10.1093/bioinformatics/btp358

Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the covid-19 epidemic? *Infectious Disease Modelling, 5*, 271–281. https://doi.org/10.1016/j.idm.2020.03.001

Roosa, K., & Chowell, G. (2019). Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models. *Theoretical Biology and Medical Modelling, 16*, 1. https://doi.org/10.1186/s12976-018-0097-6

Ross, R. (1911). *The prevention of malaria* (2nd ed.). London: Murray.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge University Press. https://doi.org/10.1017/cbo9780511755453

Russell, E. (2019). *Samoan measles outbreak: 70 deaths projected and 6500 infected*. URL: https://www.nzherald.co.nz/nz/samoan-measles-outbreak-70-deaths-projected-and-6500-infected/MVSRCFCRBRF6CO5QS6TEXSNMJ4/, 29 Nov 2019.

Schenzle, D. (1984). An age-structured model of pre- and post-vaccination measles transmission. *Mathematical Medicine and Biology: A Journal of the IMA, 1*, 169–191. https://doi.org/10.1093/imammb/1.2.169

Simpson, M. J., Baker, R. E., Vittadello, S. T., & Maclaren, O. J. (2020). Practical parameter identifiability for spatio-temporal models of cell invasion. *Journal of The Royal Society Interface, 17*, Article 20200055. https://doi.org/10.1098/rsif.2020.0055

Theil, H., & Goldberger, A. S. (1961). On pure and mixed statistical estimation in economics. *International Economic Review, 2*, 65–78.

Tuncer, N., & Le, T. T. (2018). Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences, 299*, 1–18. https://doi.org/10.1016/J.MBS.2018.02.004

Van Kampen, N. G. (2007). *Stochastic processes in physics and chemistry* (3rd ed.). Elsevier.

Wahba, G. (1990). *Spline models for observational data* (Vol. 59). SIAM.

Wang, K., Bui-Thanh, T., & Ghattas, O. (2018). A randomized maximum a posteriori method for posterior sampling of high dimensional nonlinear bayesian inverse problems. *SIAM Journal on Scientific Computing, 40*, A142–A171. https://doi.org/10.1137/16M1060625

Wilke, C. O., & Bergstrom, C. T. (2020). *Predicting an epidemic trajectory is difficult*. Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.2020200117.

Wilkinson, D. J. (2018). *Stochastic modelling for systems biology*. CRC press.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature, 466*, 1102–1104.

Xia, Y., Bjørnstad, O., & Grenfell, B. (2004). Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *The American Naturalist, 164*, 267–281. https://doi.org/10.1086/422341

Xun, X., Cao, J., Mallick, B., Maity, A., & Carroll, R. J. (2013). Parameter estimation of partial differential equation models. *Journal of the American Statistical Association, 108*, 1009–1020. https://doi.org/10.1080/01621459.2013.794730