University of Auckland

Literature Review

Parametrisation and Identification Methods for Epidemic Systems

David Wu

supervised by

Dr. Oliver Maclaren

and

Dr. Vinod Suresh

Department of Engineering Science

November 2019

Contents

List of Figures

Nomenclature

u_x	Derivative of u with respect to x
u_t	Derivative of u with respect to time, t
$\dot{u}$	Derivative of u with respect to time, t
θ	Vector of parameters
y	Vector of observations
x	Vector of state variables
$\|\cdot\|$	The L-2 norm
inf	Infinum
min	Minimum
s/t	Subject to
$\mathcal{N}(\mu, \Gamma)$	Gaussian distribution with mean μ and covariance Γ
S	Number of Susceptibles
E	Number of Exposed
I	Numberof Infectious
R	Number of Removed
N	Size of the at-risk population
$\mathcal{R}_0$	Basic reproduction number
Φ	Spline basis
$\mathcal{D}$	A differential operator
c	Coefficient vector for spline basis
$\pi(X)$	Probability distribution of X

1 Overview

Epidemiological modelling arose from the early attempts at analysing the spread of malaria by Ross [1] and empirical modelling by Kermack and McKendrick [2] in the early 20th century. These models were simple, often analytically tractable, and hugely influential in the prediction and control of infectious diseases. In fact, the SIR models introduced by Kermack and McKendrick are still used in the analysis of infectious outbreaks to this day.

One question that has been posed recently is how well predictions can be made from these models. Studies in forecasting have been performed for epidemics of, for example, HIV [3], [4], SARS [5], influenza [6]–[8], Zika [9], [10] and Ebola [11], [12]. Other, more computationally intensive, statistical approaches have also been taken to explain the spread of the same diseases. This has led to the creation of monitoring and prediction tools such as FluSight [13] and Google Flu Trends [14], to varying degrees of success [15].

The major challenges in the field stem from the mathematical problem that is being solved. At a high level, the problem is to determine the optimal model parameters that reproduce the data. The quality of this fit is described by an objective function. The challenges with the problem are two-fold:

- the objective function is highly non-convex and discontinuous, and
- the model is likely (practically) non-identifiable.

Additionally, the need to quantify the uncertainty in the predictions further complicates the problem. Methods derived from frequentist frameworks can be used to construct confidence intervals based on local approximations of the objective, but often this approximation can break down for the limited datasets we obtain. Conversely, Bayesian methods can be used to provide more robust estimates, but their need to perform many forward model simulations mean that it is extremely computationally expensive. Both approaches also have common inherent assumptions about the model which can cause estimability problems.

This motivates the use a variety of methods, which we will touch on in this review, to combat different aspects of this problem.

Review Outline

This literature review will focus on the models in mathematical epidemiology, and methods for parametrising these models. The review is structured as follows. In Section 2, we will examine literature covering the models that are used in the epidemic modelling literature — including temporal (Section 2.1) and spatio-temporal (Section 2.2) mechanistic models, as well as touching on phenomenological and statistical models (Section 2.3). Next, we examine methods of parameter estimation for various forms of models, but particularly focusing on temporal mechanistic models. Section 4 will cover a sample of applications of these methods in the epidemiological literature. Finally, we discuss some future avenues of research in Section 5.

2 Epidemiological Models

Modelling in infectious disease epidemiology is varied and nuanced. Due to the complexities of the interactions between the pathogen and humans, and sometimes the biological vectors, mean that a swath of models have arisen. Different aspects of this complexity are investigated by different forms of models. Here, we describe a breadth of classes of models, with particular focus on mechanistic models.

2.1 Temporal Models

Much of the analysis of epidemiological models is done for ordinary differential equations (ODEs). These are models of the form

$$u_t = f(u; \theta). \tag{2.1}$$

For example, the (non-age-dependent) classic SIR model [2] can be represented this way:

$$\begin{aligned} u &= (S, I, R)^T, \\ \theta &= (\alpha, \beta)^T, \\ f(u; \theta) &= \begin{pmatrix} -\beta SI/(S+I+R) \\ \beta SI/(S+I+R) - \alpha I \\ \alpha I \end{pmatrix}. \end{aligned} \tag{2.2}$$

where the states are **S**usceptible, **I**nfected/Infectious, and **R**emoved/Recovered. The parameters α and β represent the rate of recovery and effective transmission rate respectively; typically $\lambda \coloneqq \beta I/(S+I+R)$ is termed the force of infection, and $N \coloneqq S+I+R$ is the total population size.

Most compartmental models are extensions on this representation, introducing additional states of u, or changing the form of f. Much analysis has been performed for this class of models; we will cover some basic concepts.

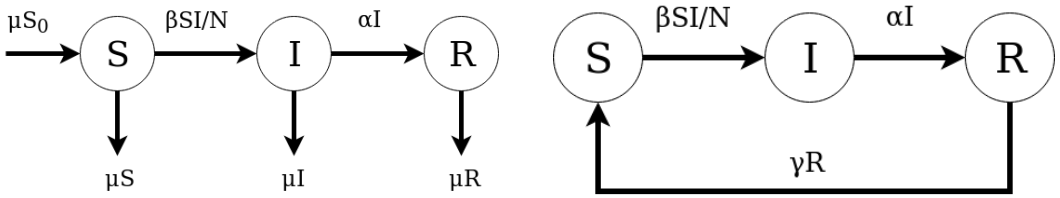
One of the most important quantities in this class of model is the basic reproduction number R_0. This represents the number of secondary infections caused by a single infected individual, and acts as a threshold value for the model. If this value is above 1 then an outbreak will occur, where a large number of individuals will become infected. For the simple SIR model, the associated R_0 would be $\frac{\beta}{\alpha}$, i.e. the ratio between rate of effective contacts and rate of recovery. The interpretation is that is if the infected manages to recover before being able to cause a secondary infection, then the disease will naturally die out.

The dynamics become more complicated with additions of elements that replenish the susceptible pool. A typical modification is the addition of vital dynamics [16], which replenishes the

susceptible pool with births. Making a constant population assumption, we can write this as:

$$\begin{aligned} \dot{S} &= -\beta SI/N - \mu S + \mu N, \\ \dot{I} &= \beta SI/N - \alpha I - \mu I, \\ \dot{R} &= \alpha I - \mu R, \\ N &= S + I + R, \end{aligned} \tag{2.3}$$

for which the quantity $R_0 = \frac{\beta N}{\mu+\alpha}$ can be computed. This form allows for the epidemic to persist endemically ($t \to \infty, I \not\to 0$), as the susceptible pool grows like μR when the infected pool approaches depletion, allowing for a recurrence of infections. A similar effect is seen if waning immunity is assumed, where recovered individuals return to the susceptible pool, either directly or via an intermediary partially susceptible state [17]. These two variants are presented diagrammatically in Figure 2.1.



(a) SIR with vital dynamics **(b)** SIRS model (waning immunity)

Figure 2.1: Diagrammatic representations of select variants of the SIR model

This endemic presence of infections and recurrence cycle can also be exacerbated by resonance, when the parameters are nonstationary [18]. A sinusoidal form of the transmission parameter $\beta = \beta_0 + \beta_1 \cos(2\pi t)$ in a waning immunity model can induce resonance for particular transmission parameters, causing large oscillations in the number of infected over time. This can interpreted as periodic epidemic outbreaks, which is what is observed for endemic epidemics such as influenza. However, this introduces more discontinuities in parameter space, as these resonance frequencies can represent suddenly well-fitting, or misfitting, trajectories. This idea of seasonal forcing is further extended by Shaman and Kohn [19] for influenza dynamics, by considering humidity as a possible driving force behind the seasonal transmission. This motivates the simulation of such recurrent epidemics for a changing climate, such as the study done in [20].

Another idea in epidemiology is modelling structures in the population. A common form is the age-structured model [21], [22], where the state variables are also partitioned by age groups. Models are typically of the form

$$\begin{aligned} \dot{S}_i &= -\sum_j \beta_{ij} S_i I_j / N, \\ \dot{I}_i &= \sum_j \beta_{ij} S_i I_j / N - \alpha I_i. \end{aligned} \tag{2.4}$$

This is useful in examining heterogeneities in the transmission properties within and between different age groups. This allows for a coarse-grain modelling of the differences in contact networks in each of the groups. They can also represent differences in the susceptibility of a group,

perhaps due to lower rates of vaccination. A similar spatial idea will be visited in Section 2.2.2, with spatial mixing heterogenities.

With such increasing complexity of epidemiological models, the recovery of the basic reproduction number, R_0, becomes more complicated. Consider the general model with m non-disease compartments and n disease compartments. Let $x \in \mathbb{R}^n$ be the disease compartments, and $y \in \mathbb{R}^m$ be the non-disease compartments, with

$$\frac{dx_i}{dt} = \mathcal{F}_i(x, y) - \mathcal{V}_i(x, y), \; i = 1, \ldots, n\,. \tag{2.5}$$

Let $(0, y_0)$ represent the disease-free equilibrium ($x \to 0, y \to y_0$ as $t \to \infty$). It can be shown that R_0 can be recovered [23] as the dominant eigenvalue of the *next generation matrix*

$$K_L = \frac{\partial \mathcal{F}_i}{\partial x_j}(0, y_0) \left(\frac{\partial \mathcal{V}_i}{\partial x_j}(0, y_0) \right)^{-1}.$$

In addition to determining the threshold relations, the eigenvector associated with the dominant eigenvalue can also be used to approximate the dominant modes of infection in an outbreak. Other model structures, such as cross-infection [24], [25] or heterogeneous mixing [26], make use of this approach in order to derive thresholds. Some work has been done on direct parametrisation on the next-generation matrix, in lieu of the rate parameters in the model [22].

The above is not an exhaustive summary of the different temporal compartmental models and related concepts in mathematical epidemiology (we have merely covered some simple examples for a few scenarios). Other variants, such as the SEIR, SITR

There also exist stochastic variants of the *deterministic* models presented above. These vary in complexity, but a simple stochastic variation is typically to consider each transition from one state to another as an event, that occurs with some probability derived from its rate [27]. Then, events are drawn and simulated consecutively with the Gillespie method or similar to generate the trajectories in time as a Markov chain. Although these models can asymptotically converge towards the trajectories of their deterministic counterparts, their stochastic nature can lead to additional behaviour, such as extinction, that is not predicted by deterministic models [28].

2.2 Spatio-temporal Models

Though purely temporal models are common in epidemiology, spatio-temporal models also have a rich history. Perhaps the first rough survey of this subject was performed by John Snow [29] on the spread of cholera in London. Spatial proliferation of epidemics is of particular interest when considering the containment and control of an emerging epidemic. Studies range in scale from local city-to-city models [30] to international models [31], [32], and remain less explored than their non-spatial counterparts. More interest has been placed in spatial models as the changes in human movement and distribution are impacting the way that diseases can spread [33].

2.2.1 Space-continuous Models

Perhaps the most natural extension to the compartmental models above is to represent spatial models as partial differential equations (PDEs), i.e. continuous in space. The earliest forms of these models was perhaps introduced by Kendall [34]. In this paper, a spatial model is derived from an averaging of the effect of nearby infected on the transmission of the infection to the susceptibles. This gives rise to the following model:

$$
\begin{aligned}
S_t &= -\beta S\left(I + kI_{xx}\right), \\
I_t &= \beta S\left(I + kI_{xx}\right) - \alpha I, \\
R_t &= \alpha I.
\end{aligned} \tag{2.6}
$$

Travelling wave solutions of this PDE were then sought, and shown to be reducible to a differential equation of I with respect to S. It is then shown that travelling wave solutions only exist when the wave velocity is above a threshold wave velocity $2\sqrt{\beta k(\beta - \alpha)}$, using arguments from the solution of the Fisher-Kolmogoroff equation [35]. A similar derivation is made in Murray [36] with a subtly different system applied to rabies, where the spatial derivative is motivated simply as a diffusion process of the infected. A thorough analysis of the wave propagation speed is presented in Rass and Radcliffe [37] for SIR and SEIR models. They prove that the propagation speed is the minimum wave speed, given radially symmetric contact distributions, and then a general result, using the saddle point method, for arbitrary contact distributions. These results are extended in [38], [39] for SIS and SIRS models, with the same result.

2.2.2 Metapopulation Models

A different, perhaps more popular, way of conceptualising spatial models is the metapopulation models. This is done by partitioning the spatial domain into different spatial compartments, and modelling the movement of people between these compartments. This means that the movement of individuals do not need to be defined over the entire domain, which is often difficult, due to the nature of human migration[40]. Further, it reduces the entire spatial domain into natural discretisations, meaning standard methods of dealing with ODE compartmental models can be used instead. A typical metapopoulation model is given by Hyman and Laforce [41] for inter-city spread of influenza:

$$
\frac{dS_i}{dt} = -\beta_S S_i \frac{I_i}{N_i} + \eta R_i + \mu(S_i^0 - S_i) + \sum_j \left(m_{ji}\frac{S_j}{N_j} - m_{ij}\frac{S_i}{N_i} \right), \tag{2.7a}
$$

$$
\frac{dI_i}{dt} = \beta_S S_i \frac{I_i}{N_i} - \alpha I_i - \mu I_i + \sum_j \left(m_{ji}\frac{I_j}{N_j} - m_{ij}\frac{I_i}{N_i} \right), \tag{2.7b}
$$

$$
\frac{dR_i}{dt} = \alpha I_i - \eta R_i - \mu R_i + \sum_j \left(m_{ji}\frac{R_j}{N_j} - m_{ij}\frac{R_i}{N_i} \right). \tag{2.7c}
$$

This includes vital dynamics with a carrying capacity in each city i of S_i^0, waning immunity at a rate of η, and explicit modelling of migration between cities at rates m_{ij} from city i to city j. The model assumes that the time scale is resolved enough so that infections are between individuals in the same city, as indicated by the form of the force of infection. This leads to a large $4 \times i$-dimensional system that can be simulated. Further, $\mathcal{R}_0$ can be computed for each of the patches, if say $\beta_{\{S,P\}}$ were different in each patch. Examples of this form of model are presented in [42], Multi-strain variants of this model are presented in [30], [43].

An alternative formulation is suggested by Bichara *et al.* [44][45], where the explicit migration

terms are removed in favour of a residence-visitor setup:

$$\frac{dS_i}{dt} = \mu(S_i^0 - S_i) + \eta R_i - \sum_j \left(\beta_i p_{ij} S_i \sum_k p_{kj} \frac{I_k}{N_k} \right), \tag{2.8a}$$

$$\frac{dI_i}{dt} = -\alpha I_i - \mu I_i + \sum_j \left(\beta_i p_{ij} S_i \sum_k p_{kj} \frac{I_k}{N_k} \right), \tag{2.8b}$$

$$\frac{dR_i}{dt} = \alpha I_i - \mu R_i - \eta R_i, \tag{2.8c}$$

where p_{ij} denotes the probability that a resident of city i is in city j. Here, the force of infection is also 'local', but x_i now tracks the states of the *residents* of city i, and the emphasis of movement is on visitations, as opposed to permanent relocation, as above. It also means that an individual's probability of migration is tied to their city of residence, as opposed to location at the time of movement.

One more common idea in metapopulations models is the gravity model of movement. This is an encapsulation of the visitation models above, but with an explicit form of p_{ij}. For example, [11] uses the equivalent of:

$$p_{ij} = \kappa_{ij} \frac{N_i N_j}{{d_{ij}}^2}, \tag{2.9}$$

where d_{ij} is the Euclidean distance between the cities i and j, and κ_{ij} is a weighting term. This model has been applied in a New Zealand-specific context in the sociology literature [46], using data generated from Census information. Particular attention was paid to the rural-urban and international migration patterns. However, there are some limitations of gravity models. Truscott and Ferguson [47] shows that infrequent, long-distance visitations are not captured, as well as movements involving nodes with small populations. Li *et al.* [32] notes that the model is valid only if the duration of the epidemic is significant; otherwise, local effects can dominate. This is congruent with observations of early-epidemic behaviour, and sparse contact networks.

2.2.3 Stochastic Spatial Models

To capture the stochasticity, sparsity and heterogeneity of contact networks, stochastic variants of the deterministic models can be used [48]. However, a more common approach is to model individuals in the population separately. This typically comes in one of two flavours:

1. Individual-based models (IBMs)
2. Network models

The individual-based, or agent-based, model is the most fine-grain level model that can be used to represent infectious disease dynamics. For typical analysis, a synthetic population of individuals is created. These individuals are then simulated forward in time by coupling models of interaction, movement and disease transmission. Typically, an infected individual will transmit the disease through interaction, which can be directly modelled, or can be approximated using a distance proxy. Individuals may also move through the spatial domain; stationary individuals generate point process models, which are popular in ecology.

Because the interaction model can be expensive to model, one simplification is to consider all interactions to occur on a network. These network models can then be simulated forward in time, with events happening with probability associated with the weights on the network arcs.

Both types of models typically model local interaction very well, and are standard appraoches for analysing the early-time behaviours of epidemic growth [49]. However, there exist problems with long-range interactions [50], which may represent, for example, visitations to family members in distant locations, or in the case of the 2014 West-African Ebola case, visitation to medical practitioners in other villages [11]. Computationally, these models are expensive to simulate, though efficient methods, such as Gillepsie's [51] are now exceeding common.

2.3 Other Models

2.3.1 Statistical Models

Another class of model is most commonly seen in statistical literature for prediction of epidemics. Statistical models are particularly popular for infections that are recurring or seasonal, such as the flu, where time series analysis techniques can be used. Statistical regression models can be constructed with no direct modelling of the underlying dynamics, and used to predict incidence of recurring epidemics. A standard regression technique is the ARIMA (autoregressive integrated moving average) model. This is a model that uses lagged differences of the dependent variables and lags of forecast errors as regressors [52]. It can be viewed as a linear nth-order differential equation model with process error, in the mechanistic framework. The drawback of this type of model is that it has limited use in predicting the behaviour of emerging infectious diseases.

2.3.2 Phenomenological Models

There are also phenomenological models that describe the incidence of disease, where standard compartmental models fail to describe specific behaviour. The most popular one used for early-epidemic behaviour [49], [53] is:

$$\dot{C} = rC^p, \; p \in [0, 1]. \tag{2.10}$$

Here, C represents the count of incidences. The bounds on p mean that the model exhibits sub-exponential ($p < 1$) or exponential growth ($p = 1$). Because early growth of epidemics is sensitive to the contact network of the initially infected, polynomial growth is often observed, for example in the 2014 West African cases of Ebola [49].

3 Parametrisation Methods

We now turn our focus to methods for parametrising these models with respect to data. The problem we are typically trying to solve is determining the the set of parameters θ for a model $\mathcal{M} : \mathcal{D}u - f(u;\theta) = 0$ such that the model outputs u match the i.e. we are interested in the inverse problem to the forward problem of examining the behaviours of the models of the previous section.

3.1 Optimisation-Based Approaches

Optimisation approaches are a natural way of constructing estimators for parameter estimation problems — the estimation problem often is reduced to finding the parameters that *best* fit the data, for some definition of best. Most optimisation approaches can also be formulated as a transformation of particular statistical estimators given certain error models, typically the maximum likelihood estimator.

3.1.1 Trajectory Matching

The typical approach is the nonlinear least-squares method, where the sum of squares error between data and model is minimised. For an ODE model, this would be minimising the error between the data and the integral (trajectory) of the model in time. We will term this trajectory matching, following Ramsay and Hooker [54]. It can be expressed as finding the optimal parameters θ_{opt} such that

$$\begin{aligned} \theta_{opt} &= \arg\min_{\theta} \sum_i \sum_n \left(y_i(t_n) - S_i(t_n)\right)^2, \\ S(t) &= \int_{t_0}^{t} f(S(\tau);\theta) d\tau, \\ S(t = t_0) &= x_0. \end{aligned} \tag{3.1}$$

where i indexes the state variables (elements of x).

One problem with trajectory matching is that the objective function is often non-convex. Further, we expect that discontinuities exist, corresponding to bifurcations in the model. This leads to a objective that is ill-suited to standard gradient-based (Newton-Raphson) methods or derivative-free methods. Despite this, there exist many common methods that attempt to tackle these problems with the aid of computational brute force. One example is the use of multi-start methods which perform multiple runs of the optimisation procedure from different initial iterates [11], [55], [56]. Methods such as the Latin hypercube sampling technique are used to generate these initial iterates. The hope is that, with sufficient coverage of the parameter space, the minima computed over all the samples will be close to the global optimum. Novel techniques have been proposed [57] to generate initial iterates that are suitable to mitigate this issue.

A secondary problem with the trajectory matching problem is that there is significant computational effort involved in calculating the model outputs, as this typically involves the numerical integration of the model. Due to problems such as stiffness, or complex time-dependent model forms, this can take an unacceptably long amount of time. Hooker *et al.* [58] notes that in their

study (described below in Section 4) that a sampling based approach would have been infeasible due to the time it would take to generate a single proposal of the state trajectories. Further, careful selection of the initial iterates, and subsequent estimates, is required to minimise the chance that the parameters estimated require a long time to integrate.

One philosophical problem with the trajectory matching method, however, is the assumption made about the correctness of the model. Because the model is integrated forward in time, the model is effectively enforced exactly, bar numerical error. This is problematic since it also implies that the only error in the data is in the observation model, and due to the form of the objective, that error is additive iid. This means that the method cannot account for any process error, which is error that accumulates and is correlated. This process error is typically biased, which biases the parameter estimates in turn. There are methods of dealing with this inadequacy of models, as discussed in Section 3.3.2.

3.1.2 Gradient Matching and Splines

An alternative approach to trajectory matching is sometimes proposed to mitigate the issues of this latter problem. Instead of evaluating the model outputs via integration, the associated derivatives of the data can be taken, and compared to the model directly. This method, termed by Ramsay and Hooker [54] as gradient matching, can be expressed:

$$\theta_{opt} = \arg\min_{\theta} \sum_{i} \sum_{n} \left(\mathcal{D}\hat{y}_i(t_n) - f(\hat{y}_i(t_n); \theta)\right)^2, \tag{3.2}$$

for some model

$$Dx = f(x; \theta),$$

where $\hat{y}$ represents an approximation to the data y. This approximation $\hat{y}$ is typically a smoothed representation of the data, and is typically expressed as a basis expansion

$$\hat{y}(t) = \sum_{k} c_k \Phi_k(t), \tag{3.3}$$

in some basis Φ. This can then be substituted in the objective function, giving the second stage:

$$\theta_{opt} = \arg\min_{\theta} \sum_{i} \sum_{n} \left(\mathcal{D}\Phi(t_n)c - f(C\Phi(t_n); \theta)\right)^2. \tag{3.4}$$

Φ is often a spline basis, and the coefficient vector c is computed from minimising the least squares error

$$\|y - \Phi c\|^2. \tag{3.5}$$

However, there exists a meta-problem with the use of splines, in the fact that the quality of the fit (when done with least-squares) is quite sensitive to the choice in number and location of the knots of the spline. In short, a low number of knots will typically lead to underfitting, and a high number of knots leads to overfitting. A now standard solution to this problem is to penalise the second derivative of the fit to enforce smoothness. This approach, presented in such texts as [59]–[61], modifies Equation 3.5 to

$$\|y - \Phi c\|^2 + \lambda \|\Phi_{tt} c\|^2, \tag{3.6}$$

with some smoothing parameter λ. Wahba [59] shows that a suitable value of λ can be computed by minimising the generalised cross validation error (weighted leave-one-out error, GCV). If an

estimate of the spline coefficients for a given the tuning parameter is denoted $\hat{g}(\lambda)$, then the GCV is given by:

$$V(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\frac{(y_k - \Phi_k c(\lambda))^2}{\left(1 - \frac{1}{n}\sum_{i=1}^{n}\Phi_i c_i(\lambda)\right)^2} \tag{3.7}$$

One of the problems with using gradient matching in a practical context is that the seconds step, computing $\mathcal{D}\Phi(t)c$, requires a fully observed system. That is, if the model has m state variables, all m states must be observed at each time point. When considering with real data, it is often impractical, if not impossible, to gather data on all states. There have been attempts at estimating these unobserved states in the econometrics literature [62] through the instrument variables method.

3.1.3 Generalised Profiling

In [63], and expanded in [54], Ramsay et al. introduce the concept of generalised profiling, which attempts to solve the (complementary) issues of both trajectory matching and gradient matching. The crux is that the optimal parameter(s) θ_{opt} is the minimiser of

$$\|y - \Phi c\|^2 + \lambda \|\mathcal{D}\Phi c - f(\Phi c; \theta)\|^2 . \tag{3.8}$$

This is extremely similar to the smoothed spline function (Equation 3.6), replacing the smoothing term with the gradient matching term. This is because the motivation is very similar. The smoothing term penalises non-smoothness (i.e. non-zero second derivatives), which generally does not hold for most models. Intuitively, we can simply replace this with our model, and penalise for that. From a different viewpoint, this can also be constructed as a relaxation of the constrained trajectory matching problem:

$$\begin{aligned} &\min \|y - \Phi c\|^2 , \\ &s/t \\ &\quad \|\mathcal{D}\Phi c - f(\Phi c; \theta)\|^2 = 0. \end{aligned} \tag{3.9}$$

In [54] (and all related works [58], [63]–[65]) the parameter estimation is split into two stages, with different objective functions. The above objective function is used in the inner optimisation procedure, where the spline coefficients c are estimated for a given θ; the standard least-squares objective is used to in an outer optimisation procedure where the optimal spline coefficients are used as a surrogate for the estimated state. We can express this as:

$$\theta_{opt} = \arg\min_{\theta} \left\{ \|y - \Phi c(\theta)\|^2 \right\} \qquad =: \arg\min_{\theta} H(\theta), \tag{3.10}$$

$$c(\theta) = \arg\min_{c} \left\{ \|y - \Phi c\|^2 + \lambda \|\mathcal{D}\Phi c - f(\Phi c; \theta)\|^2 \right\} \qquad =: \arg\min_{c} J(\theta, c). \tag{3.11}$$

It can be shown, using the implicit function theorem and the fact that $\frac{\partial J}{\partial c} = 0$ after optimising for $c(\theta)$, that the derivative of the outer objective $\frac{dH}{d\theta}$ can be expressed as

$$\frac{dH}{d\theta} = \frac{\partial H}{\partial \theta} - \frac{\partial H}{\partial c}\left(\frac{\partial^2 J}{\partial c^2}\right)^{-1}\frac{\partial^2 J}{\partial c \partial \theta}. \tag{3.12}$$

This then allows derivative methods for minimisation to be used. Further, since $c(\theta)$ is likely to be locally smooth and continuous, the inner optimisation procedure should take a minimal amount of iterations to converge.

One curiosity of the method is the form of the outer objective function. If we consider the inner objective as profiling out the structural parameters, it makes more sense for the outer objective to have the same form as the inner objective. The reasoning given by Ramsay and Hooker [54] is that $c(\theta)$ represents a regularised estimate of the state, so further regularisation in the outer objective is not needed. Perhaps a stronger (but equally non-satisfying) reason would be that the form of the outer objective lends more easily to an interpretation of the objective as the log-likelihood of a simple normally distributed error model, which allows for straightforward construction of confidence intervals for uncertainty quantification (see Section 3.1.5).

A similar method has also been explored in the geophysics literature, termed the all-at-once method[66], [67]. Similar to the generalised profiling method, the state and the parameters of the model are estimated at the same time through a minimisation algorithm. However, practitioners typically enforce the model exactly when solving the model, due to the application area having physically-defined laws that govern the models.

3.1.4 Choices of Regularisation Parameter

We see that in Equations (3.6) and (3.8) that an extra tuning parameter (λ) is introduced. This parameter describes the tradeoff between the fit to data term and another regularisation term (smoothing or model fit). It can be shown that the solution to the objective will be dependent on the choice of this parameter. That raises the question of how to choose the value of such a parameter. From the inverse problem literature, we see that there are a multitude of methods of choosing this value [68]. They can classified into *a priori* and *a posteriori* methods.

A priori methods define a strict stopping condition based on prior knowledge or estimates of the error structure and size present in the model. One of the most popular techniques in this class is the Morozov discrepancy principle Scherzer [69]. This relies on specifying an estimate of the magnitude of expected observational error in the data, and choosing the regularisation value λ that produces this magnitude of misfit error. However, unless an estimate of the size of the error is available, such methods are not well-justified.

A posteriori methods subvert the need for a prior estimate of the error by examining the properties of the estimate after it has been computed. Methods of this class include the L curve criterion and (generalised) cross-validation techniques. The L curve criterion is a form of Pareto front of the two objectives of solving the inverse problem, and regularising the solution. The argument is that the regularisation is seen to provide value, until it significantly changes the quality of the fit, thus the "optimal" choice of regularisation parameter will be at this value threshold. This corresponds geometrically to the area where the L curve turns, and consequently to where the curvature is the highest. It is, by all accounts, an ad hoc method for tuning the regularisation parameter. Cross-validation techniques [59], like the GCV (Equation (3.7)), compute the error in prediction as compared to out-of sample data bootstrapped from the given data. This is done by partitioning the provided data into two sets, one to perform the parametrisation on, and the other to validate, and compute the criterion value. They can be expensive to compute exactly, but approximations such as k-fold cross validation [70] where the computation is not performed exhaustively, are used in practice.

The problem of selecting these tuning parameters is exacerbated when more regularisation terms are included. For example, if we add a Tikhonov regularisation term to the generalised profiling objective in Equation (3.8), we would also need to add an additional tuning parameter, which would need to be selected alongside λ. Methods have emerged for such multi-parameter selection problems; we focus on L-curve methods for parameter selection. Belge *et al.* [71] introduce the

L-hypersurface method to perform multi-parameter selection, built as a natural extension of the L-curve criterion. The standard L-curve criterion can be expressed as a curvature maximisation problem:

$$\begin{aligned} &\sup_{\lambda} \texttt{curvature}(f_1(\lambda), f_2(\lambda)) \\ &\texttt{curvature}(f_1(\lambda), f_2(\lambda)) = \frac{f_1' f_2'' - f_1'' f_2'}{((f_1')^2 + (f_2')^2)^{3/2}} \end{aligned} \tag{3.13}$$

where f_2, the regularisation is plotted against f_1, the misfit, and $f_i' = \frac{df_i}{d\lambda}$. If this is extended for $f_3, f_4, \ldots$, then we can use the Gaussian curvature in place of `curvature`, but this can be expensive to compute. Belge *et al.* [72] suggest a heuristic method for a linear model, based on the eigenvalues of the operator defining an origin on f_i-space, and approximating the curvature with a simple distance surrogate.

3.1.5 Uncertainty Quantification

Much of the uncertainty estimates of the optimisation-based methods are generated in a frequentist manner. The observation is that if we specify the likelihood function as

$$\mathcal{L}(y(t)|\theta) \sim \mathcal{N}(S(t), \Sigma), \tag{3.14}$$

where $S(t)$ is defined as in Equation 3.1, and Σ is some covariance matrix; then the negative log-likelihood becomes the sum of squares function for $\Sigma = \sigma^2 I$. That is, minimising the negative log-likelihood will be identical to performing trajectory matching. This is powerful, since it allows standard asymptotic statistical theory to be applied to the estimates produced by trajectory matching. In essence, this means that the uncertainty of the estimate can be quantified. The standard approach is to compute the Fisher information matrix from the Hessian of the objective function [73]. This can be then inverted and used to construct confidence intervals of arbitrary confidence thresholds, as specified by the χ^2 distribution [74]. These confidence intervals thus represent local quadratic approximations to the log-likelihood function about the maximum likelihood estimate. This holds if the likelihood is indeed distributed as given above, but more sophisticated uncertainty quantification techniques are required for more exotic forms of error.

3.2 Bayesian Approaches

3.2.1 Bayesian Methodology

A standard framework for interpreting the uncertainty in parameter estimates is the Bayesian framework, which centres around Bayes' theorem:

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)} \propto \pi(y|\theta)\pi(\theta). \tag{3.15}$$

Here we interpret y as the data, and θ as the parameters we are interested in estimating. In many other inverse problems, the main focus of the problem is the estimation of the state of the system - this will be discussed in Section 3.2.3.

$\pi(\theta)$ is denoted the *prior distribution* of the parameters, and is a representation of prior knowledge of the parameters. It is also a source of bias in our estimate, so much effort is often made in statistical circles to construct uninformative priors that do not bias the posterior [54]. $\pi(\theta|y)$ is

the *posterior distribution*, or our estimate distribution of the target parameters conditioned on the knowledge of the data. This makes up our estimate (and representaion of uncertainty) of the model parameters that "fit" the data.

$\pi(y|\theta)$ is denoted the likelihood distribution. This is a representation of the probability that the data y would be produced given some value of the parameters θ. For model-driven problems, this is often a function of the model outputs. For example, in [75], the likelihood is constructed as a Poisson distribution, centred about the underlying state.

By computing the posterior, we can compute estimators, such as the *maximum a posteriori* or MAP estimate, which maximises the posterior distribution, as an estimate of the parameters; we can also produce a region of parameter that reflect

3.2.2 Sampling Methods

The direct computation of the posterior distribution from the likelihood and prior distributions is difficult, often infeasible. In lieu of this, sampling methods aim to generate surrogate distributions that approximate the posterior in the finite case, and converge to the posterior in the asymptotic case.

A popular method is the Monte Carlo Markov Chain (MCMC) method, where samples are generated sequentially from the previous sample in a Markov chain. There are various methods of generating these samples:

1. Metropolis-Hastings (MH-MCMC)
2. Gibbs Sampler (Gibbs-MCMC)
3. Hamiltonian Monte Carlo (HMC)

The Metropolis-Hastings algorithm [76], [77] is the most popular implementation of MCMC. It is an accept-reject algorithm, with an acceptance ratio derived from the non-normalised posterior, $\hat{\pi}(\theta|y) = \pi(y|\theta)\pi(\theta)$, and subsequent samples in the chain generated by some proposal distribution q. The method, in brief is:

1. Set $k = 1$ and choose some initial value $\theta^{(1)}$.
2. Draw a proposal $\hat{\theta}^{(k+1)}$ from the proposal distribution $q(\theta^{(k)}, \hat{\theta}^{(k)})$.
3. Compute the acceptance ratio α

$$\alpha = \min\left(1, \frac{\hat{\pi}(\hat{\theta}^{(k+1)}|y)\ q(\hat{\theta}^{(k+1)}, \theta^{(k)})}{\hat{\pi}(\theta^{(k)}|y)\ q(\theta^{(k)}, \hat{\theta}^{(k+1)})}\right)$$

   .
4. Draw $t \in [0, 1]$ from the uniform distribution $\mathcal{U}(0, 1)$.
5. If $\alpha > t$, set $\theta^{(k+1)} = \hat{\theta}^{(k+1)}$ and set $k \to k + 1$.
6. Terminate if $k > K$; else go to Step 2.

An intuitive interpretation of the algorithm is that the proposals are generated by some random traversal of the parameter space, and are accepted or rejected with probability representative of the posterior in that region of parameter space. This gives rise to a problem with such a method — that the method is only as good as the topology of the posterior. If the posterior was only significantly large on multiple non-connected regions of parameter space, it is unlikely

that all regions are explored by the proposals. More commonly, if the posterior is skewed, then the number of rejections will be large, meaning it takes a large amount of computational time to generate the desired K samples.

The Gibbs sampler [78] is an alternative sampling algorithm that samples directly from the posterior to avoid rejection. The implementation of the proposal step is much more complex, and requires the derivation of the conditional probability

$$\pi(\theta_i|\theta_{j\neq i}, y)$$

in order to generate samples. Hamiltonian Monte Carlo [79] augments the parameter vector with auxiliary momentum vectors that are used to generate derivative-informed adaptive samples of the parameters. Like MH-MCMC, the algorithm requires tuning of algorithm parameters related to step size between samples. The Stan implementation of HMC utilises a no-U-turn sampler (NUTS) to automatically tune the algorithm parameters in the warmup phase. Variational inference [80] belongs to a different class of Bayesian inference techniques, and comes from the field of machine learning. Instead of sampling the posterior distribution, the algorithm attempts to find an adequate approximation to the posterior distribution by minimising the KL-divergence, or other surrogate distance measure, between the proposed approximation and the true posterior. By using automatic differentiation, the optimisation problem can be solved using gradient-based techniques.

3.2.3 Data Assimilation Methods

The data assimilation methods typically solve a *smoothing* or *filtering* problem. The smoothing problem is the problem of estimating the true (noiseless) state of data, and the filtering problem is the problem of estimating future states from given data. These problems can be modified by augmenting the state vector with parameters, allowing for the parameter estimation problem to be answered.

The standard filtering method (for linear systems) is the Kalman filter [81]. This uses the error residual between the estimated state and the observation at a point to compute an update based on the estimated covariance of the error in the observations. Notably, the method is recursive, and only requires the previous iterate's state estimate, as opposed to the entire history, allowing for on-line assimilation of data. An extension for nonlinear systems is achieved by the linearisation of the model about the state estimate.

One other popular method is the iterated filtering method [82], from the study of partially observed Markov processes. This is to model stochastic processes, where the observations are a single realisation of the model. The method uses multiple chains as a representative sample of the realisations of the model with candidate parameters. At each iteration of the parameter estimation process, the chains are simulated forward in time, and averaged to produce an average state that is used as the state estimate. Then, the parameter estimates are updated in a similar manner to classical filtering techniques.

Classical filtering techniques are used in Pei *et al.* [42] for influenza in the United States, and particle filtering techniques are used in [82], [83] for cholera and rotavirus epidemics respectively.

3.2.4 Relation to Optimisation Methods

The trajectory matching objective is the log of the likelihood model if independent, normal, additive observation error is assumed:

$$\begin{aligned} y(t) &= \int_0^t f(x(\tau);\theta)\,d\tau + \epsilon, \\ \epsilon &\sim \mathcal{N}(0,\sigma^2\mathbb{I}). \\ \pi(y(t)|\theta) &= \mathcal{N}\left(\int_0^t f(x(\tau);\theta)\,d\tau,\ \sigma^2\mathbb{I}\right). \end{aligned} \tag{3.16}$$

A standard result in statistical inverse problem theory [84] is that the minimser of the Tikhonov-regularised form of the trajectory matching objective is the MAP estimate of the posterior density. Thus, the inferences made using least-squares optimisation can be seen as specific cases of a more general problem structure.

3.3 Estimability Problems

Despite these numerous approaches to parameter estimation, there exist some underlying problems that aren't directly answered, or overlooked by practitioners. Often, this is benign, but in certain cases, these can cause serious problems in the estimation methods.

3.3.1 Identifiability

The most significant one is the problem of identifiability [74]. This is where, regardless of the quantity of data provided, the quality of the data or the structure of the model prevents the parameters to be estimated uniquely. That is, there exist at least two distinct set of parameters that can produce indistinguishable observations .

There are different categories of identifiability. The first, *structural identifiability* is more well-defined than the other, *practical identifiability*.

Definition 1 (Structural Identifiability) *A process model, $\mathcal{M}$, under an observation model, $\mathcal{H}$, is said to be* structurally identifiable *if: $\mathcal{H}(\mathcal{M}(\theta_1)) = \mathcal{H}(\mathcal{M}(\theta_2)) \implies \theta_1 = \theta_2$ for all parameter combinations θ_1 and θ_2, given infinite data $y = \mathcal{H}(\mathcal{M}(\theta))$.*

In particular, we will consider the form

$$\dot{x}(t) = f(x(t), u(t), \theta) \leftrightarrow x(t) = \mathcal{M}(\theta), \tag{3.17a}$$

$$y(t) = g(x(t), u(t), \theta) \leftrightarrow y(t) = \mathcal{H}(x(t)). \tag{3.17b}$$

The interpretation is that each set of observations has a unique corresponding parameter set. Identifiability that is non-trivial to solve, since it has to do with the structure of the model, or the quality of the data — both difficult things to modify in the parameter estimation process. Perhaps unfortunately, it is easy to show that even simple models can be non-identifiable. The simple turnover model is one such example [85].

$$\begin{aligned} \dot{x} &= k_1 - k_2 x, \\ y &= sx. \end{aligned} \tag{3.18}$$

Rearranging the equations, we can simplify the model to

$$\dot{y} = k_1 s - k_2 y. \tag{3.19}$$

If we consider that y is the only observable state, then it is obvious that there exist an infinite set of pairs $\{(k_1, s) : k_1 s = c \in \mathbb{R}\}$ for some constant c, and thus an infinite set of parameters that can reproduce the same data. More examples appear in the literature, for a wide range of biological applications [12], [86]–[88] where the system much more complex, and such non-identifiabilities are much more difficult to detect.

There exist algebraic methods for analysing the structural identifiability of systems a priori. Differential algebraic techniques such as Differential Algebra Identifiability of Systems (DAISY) or Exact Arithmetic Rank (EAR) methods. The former of these generates a system of algebraic nonlinear equations in θ from the algebraic elimination of the state variables, which can then be analysed for multiple roots (solutions) to detect global non-identifiability. The latter computes the Jacobian of the observation model as a function of the initial state and parameters, and applies the inverse function theorem to use the rank of the Jacobian matrix about a random point to determine local identifiability.

Practical identifiability is a subset of structural identifiability that specifies the estimability of the parameters from a finite sample of noisy data. The definitions for practical identifiability are not consistent within the literature, and are typically reverse-engineered from the method that is used to analyse it. This concept has also been termed inconsistently as practical identifiability[74], [89], sloppiness (for nonidentifiability)[90], [91], and estimability [92].

Definition 2 (Practical Identifiability) *A process model, $\mathcal{M}$, under an observation model $\mathcal{H}$ is said to be* practically identifiable *if $\exists\delta \;:\; \mathcal{H}(\mathcal{M}(\theta_1)) = \mathcal{H}(\mathcal{M}(\theta_2)) \implies |\theta_1 - \theta_2| < \delta$ for all parameter combinations θ_1 and θ_2, given finite data $y = \tilde{\mathcal{H}}(\mathcal{M}(\theta))$ for some noisy observation model $\tilde{\mathcal{H}}$.*

Practical identifiability is generally not tractable as structural identifiability a priori. Standard methods will detect practical identifiability a posteriori by analysing the qualities of the parameter estimates recovered. Popular methods include Fisher Information Matrix (or sensitivity analysis) methods, Monte Carlo (or bootstrapping) methods, and profile likelihood methods.

Fisher Information Matrix (FIM) methods approximate the local behaviour of the objective function/(log-)likelihood function about the recovered parameter estimate. The inverse FIM can be used to compute parameter correlation coefficients, or to produce bounds on the parameter precision (roughly equivalent δ in Definition 2). Because the FIM is expectation of the negative Hessian [73], computational algebra techniques can be used to compute this approximation for arbitrary sets of parameters. Typically, this is done for the maximum likelihood estimate, or the minimiser of the objective function, i.e. computing the empirical FIM. This method is also closely related to the uncertainty quantification procedures performed in Section 3.1.5.

Another method of analysing non-identifiability is the profile likelihood method as introduced by Raue *et al.* [74]. The profile likelihood analyses practical identifiability of systems by computing the change in the objective function of the optima as one constrained parameter, or more generally a function of the parameters $\Omega(\theta)$, is varied, also termed the profile likelihood, $\mathcal{L}_p$:

$$\mathcal{L}_p(\omega) = \min_{\theta|\Omega(\theta)=\omega} \{\mathcal{L}(\theta|y)\}. \tag{3.20}$$

Here, Ω represents a function over the parameters, and has its value set and varied to 'profile out' the function.

One simple form of Ω, and the one that is explored in [74], is having $\theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_p\}$ and $\Omega(\theta) = \theta_i$. This corresponds to profiling for each parameter separately, and is used to detect if any given parameter is identifiable. However, a more powerful

A similar idea appears in [89], where Monte Carlo methods are used to generate a profile of the parameter estimates as the (known synthetic) data is perturbed with observational noise. Synthetic data is generated, and noisy samples are generated with a known magnitude of additive error. Parameters are then estimated from these samples, constructing an approximation of the behaviour of the objective in parameter space about the true parameters. The samples generated are used to compute the sample covariance, and thus allow parameter covariance to be approximated. This allows the detection of unidentifiable parameters, and the plots of the recovered parameters can indicate combinations of parameters that may be identifiable. Similar methods are applied in [90], with the additional insight that the geometric alignment of the sample distribution and Fisher approximation manifest in different confidence interval properties.

3.3.2 Model Inadequacy

A second problem arises that perhaps is more prevalent in biological models than in physical models is model inadequacy or discrepancy. Because we are modelling highly complex, random processes with simple, and often deterministic, models, there exists a level of discrepancy between the real process and the model. As data is collected and assimilated for parameter estimation, these discrepancies between the model and reality can compound such that the standard error models are no longer valid. Brynjarsdóttir and O'Hagan [93] formalise this concept, and show the importance of this type of error. In their setup, they consider some true process $\dot{x} = \zeta(x;\theta)$, associated true observation model $y = x+\epsilon,\ \epsilon \sim \mathcal{N}(0,\sigma^2)$, an incorrect process model $\dot{x} = f(x;\theta)$, and three variants of observation model:

1. $\tilde{y} = x + \epsilon$
2. $\tilde{y} = x + \delta(x) + \epsilon,\ \delta(x) \sim \mathcal{GP}(0, \sigma^2 k(\cdot,\cdot|\psi))$
3. $\tilde{y} = x + \delta(x) + \epsilon,\ \delta(x) \sim \mathcal{GP}(0, \sigma^2 k(\cdot,\cdot|\psi)),\ \delta(0) = 0,\ \frac{d\delta}{dx}(0) = 0$

representing no model discrepancy, unconstrained model discrepancy, and (prior-) constrained mode discrepancy, where $k(\cdot,\cdot|\psi)$ represents some distance kernel parametrised by ψ. It is shown that for a simple process that the observation model with model discrepancy does not recover the parameters correctly, and the true parameter value for θ is not covered by the posterior, and consistently underestimates the parameter value. The unconstrained model discrepancy also displays similar behaviour, underestimating the true value of the parameter, but being able to interpolate the data correctly. However, the constrained model discrepancy observation model did manage to recover θ adequately. However, it performed poorly in the extrapolation exercise. These results were then shown to be expected through an analysis of the asymptotic behaviour of the models. The conclusion is that strong (and correct) priors on the model discrepancy term are required in order to recover the parameters correctly.

4 Applications

Many of these parameter estimation techniques have been applied in retrospective studies of epidemiological events, such as SARS, influenza pandemics, as well as modern outbreaks such as Ebola and Zika. Though the problems they tackle are often very similar, there are many approaches taken, and compromises made.

Many studies tackle the nonconvexity and nonidentifiability of the trajectory matching methods head-on. For example, Chowell *et al.* [5] uses the trajectory fitting method to parametrise a subset of parameters in an SEIR model with a partitioning of S into high-risk (S_1) and low-risk (S_2) subpopulations, as well as having partial infectivity from the exposed/latent class. The aim of the project was to determine an appropriate value of $\mathcal{R}_0$ for outbreaks of SARS in the 2002-2003, across different isolated populations (Hong Kong, Singapore, Toronto), and the effect of isolating infected individuals on $\mathcal{R}_0$ as a control strategy. Problems due to obtaining local optima were mitigated by simply choosing the best fit among 10 restarts. The uncertainty of the parameter estimates is never addressed. However, uncertainty analysis is performed on $\mathcal{R}_0$ by Monte Carlo sampling of arbitrary probability distributions assigned to model parameters, and then computing $\mathcal{R}_0$ directly. It is likely that the lack of statistical rigour is due to the computational limitations of the time, meaning that trajectory fitting methods took significant amount of timeto run.

A more recent study that also utilises optimisation-based trajectory matching was done by D'Silva and Eisenberg [11], where the spatial spread of Ebola in West Africa in 2014-2016 is modelled with a gravity model, as a variant to the models in section 2.2.2. Instead of modelling movement explicitly, like in eq. (2.8), it instead modifies the force of infection based on a gravity term. For their 3-patch country-level model, the force of infection λ_n on patch n can be written

$$\begin{aligned} \lambda_n &= \sum_{i=\{1,2,3\}} \theta_{n,i}(\beta_{1,n} I_{1,i} + \beta_{2,n} I_{2,i} + \beta_{F,n} F_i), \\ \theta_{n,i\neq n} &= \kappa_n \frac{N_i N_n}{(d_{n,i})^2}, \\ \theta_{n,n} &= 1, \end{aligned} \tag{4.1}$$

for a progression of states $S \to E \to I_1 \to I_2 \to F \to$, $I_1 \to R$, $I_2 \to R$, where I_1, I_2 represent distinct infection stages and F is the funeral stage. A similar model is used for the 63-patch district level model. The models were parametrised against WHO reports of incidence, using a trajectory matching approach, solved by multi-start Nelder-Mead optimisation. Initial iterates of the optimisation were generated with Latin Hypercube sampling over a relatively narrow range, and the high number of recovered estimates were used to generate empirical confidence intervals. The underdetermination of the system was alleviated by setting certain parameters from the literature, which may have alleviated some identifiability issues. One potential flaw that is not noted in the paper is that data collection for the Ebola outbreak in West Africa suffered from non-reporting that was both time-varying and spatially heterogeneous. This has been shown [94] to potentially cause bias in the estimates of important epidemiological parameters, such as $\mathcal{R}_0$. This is partially addressed in the D'Silva's model with a constant k_{norm} reporting rate that is set a priori.

In the realm of generalised profiling, Hooker *et al.* [58] applied the technique to the Ontario

measles cases over a period of fifty years from 1939 to 1989. In this paper, a variant of the SEIR model with time-dependent transmissivity parameter is fitted.

$$\dot{S} = -\beta(t)S(I+\nu) + \mu(t), \tag{4.2a}$$

$$\dot{E} = \beta(t)S(I+\nu) - \gamma E, \tag{4.2b}$$

$$\dot{I} = \gamma E - \alpha I, \tag{4.2c}$$

where ν represents a visitation impact, and $\mu(t)$ a recruitment (births, migration) rate. To fit $\beta(t)$, the seasonal effect is projected onto a basis of periodic B-splines

$$\beta(t) = \beta_0 + \beta_1(t - \bar{t}) + \sum_{i=1}^{k} \phi_i(\mathrm{mod}(t,1))\beta_{2,i}, \tag{4.3}$$

such that a linear effect is captured by β_0 and β_1, and $\bar{t}$ represents a reference time (1952 is used). An observation model is also imposed as

$$y(t) = p(t)I(t) \tag{4.4}$$

for some reporting rate $p(t)$. A simplification is made by observing that in Ontario, the reporting rate was relatively constant, i.e. $p(t) = p_0$. The concept of forward cross-validation as an option of determining the model regularisation parameter, λ, is also (re-)introduced. This method requires the computation of the error between the recovered (spline) trajectories at time t_i and the trajectories predicted by integrating the model with the recovered parameters from some previous time $t_i - h$. Within the paper, they also very briefly touch on aspects of practical identifiability, noting the flat profile of the objective function over a parameter of a forcing term; other parameters are not discussed. Further analysis was done by performing stochastic simulation studies and refitting, to analyse biases in the generalised profiling objective to standard trajectory matching, as well as biases due to the collocation methods (spline basis projection).

Bayesian approaches are perhaps more popular in the epidemiological parameter estimation space. One example is the work of Chatzilena *et al.* [75] using Stan, a state-of-the-art implementation of Hamiltonian Monte Carlo (HMC) and Variation Inference (ADVI) methods, to analyse the parameters of various models describing a 1978 influenza outbreak in an English boarding school. Data was collected daily on the number of cases of infection. A number of different models were used and compared, of which two are of immediate interest: a deterministic and a stochastic model. The deterministic model used the classical SIR model as the process model, and the observation model was constructed as a Poisson process

$$y \sim Poisson(\lambda), \tag{4.5a}$$

$$\lambda(t) = \int_{t_0}^{t} \frac{dI(\tau)}{dt} d\tau = \int_{t_0}^{t} \left(\beta S(\tau)I(\tau) - \alpha I(\tau)\right) d\tau. \tag{4.5b}$$

The stochastic model consisted of modelling process error in the form of an Ornstein-Uhlenbeck

process

$$y(t) \sim Poisson(\lambda(t)), \tag{4.6a}$$
$$\lambda(t) = \exp(\kappa(t)), \tag{4.6b}$$
$$d\kappa = \phi(\mu(t) - \kappa(t))dt + \sigma dB, \tag{4.6c}$$
$$\mu(t) = \log\left(\int_{t_0}^{t} (\beta S(\tau)I(\tau) - \alpha I(\tau))\, d\tau\right), \tag{4.6d}$$
$$\kappa(t+1)|\kappa(t) \sim \mathcal{N}\left(\mu(t) + (\kappa(t) - \mu(t))e^{-\phi}, \frac{\sigma^2}{2\phi}(1 - e^{-2\phi})\right), \tag{4.6e}$$

where B is standard Brownian motion, ϕ is a speed of reversion hyperparameter, and σ an instantaneous diffusion term. This formulation roughly corresponds to a process model that drifts about the true state, and has multiplicative error, observed as a Poisson process of the number of infected. Both formulations are utilising a trajectory matching approach, but both models do give adequate estimates of the data. However, the deterministic model underestimates the uncertainty of the estimate, which is similar to the result in Brynjarsdóttir and O'Hagan [93] when no process error is considered, as well as being a hallmark of non-identifiability. It is also noted that the ADVI method is computationally more efficient, but both ADVI and HMC produce almost identical results, in terms of parameter estimates.

The profile likelihood technique has also been applied by Tönsing *et al.* [10] to both the influenza outbreak used by Chatzilena *et al.* [75] on the deterministic model (Equation (4.5)), and to a dataset of Zika outbreak in Colombia. For the influenza case, they surprisingly suggest that the parameters are all identifiable, in contrast to the analytical results on structural identifiability by Tuncer and Le [89]. Interestingly, their plots of the trajectories in the unobserved state show a large amount of uncertainty in the magnitude of the outbreak. For example, the recovered (95%) profile likelihood-based interval for the initial at-risk population ranged from 565 to 2324 while reporting a literature-consistent mean of 854. For the Zika case, they fit weekly data on new infection cases to a SEIR-SEI host-vector model of Zika. The human I compartment is split into symptomatic and asymptomatic infectious classes to reflect the nature of the presentation of Zika in humans. Fitting is done by performing trajectory matching on the cumulative cases, with an observation model

$$y(t) = x(t) + \epsilon(t), \tag{4.7a}$$
$$\epsilon \sim \mathcal{N}(0, \sigma_{rel}^2 x(t) + \sigma_{abs}^2), \tag{4.7b}$$

where ϵ acts as a multiplicative and additive error term. x is taken as the cumulative number of newly symptomatic infected human cases. It is found that nearly every parameter in the model is non-identifiable, many exhibiting one-sided non-identifiability. However, by reducing the model by setting certain parameters to fixed values, motivated by the flat profile likelihoods and prior knowledge, the remaining parameters become identifiable. What is not done is profiling of combinations of parameters, which may have guided a more structured and well-motivated model reduction system.

5 Future Directions

Here, we discuss some future directions in epidemic forecasting and parameter estimation. These are drawn from both the literature directly, and as critical observations of the literature.

Characterising Contact Networks

As mentioned in Section 2.2, the modelling of contact networks is still not a mature field. Data on contact networks is difficult to acquire, as it requires relatively invasive methods to acquire fine-grain data.

New technology has made it possible to perform detailed tracking of people, and this has been used in one detailed simulation study [95] to model the spatial spread of a novel infection in the United Kingdom. However, practical use of this type of tracking for response is currently not possible due to limitations of current health information systems [96].

Without the detailed tracking information, it is often difficult to define parametrisable models that truly capture the detailed contact networks. Though hypothetical modelling is possible, it is difficult to state how such a formulation could lead to an identifiable parameter estimation problem. The fall-back is often homogenisation of the contact network, which have been done for limited cases through moment approximation [97], [98]. However, such approximations have been shown to perform poorly when the network contains long-range contacts [50].

Model Inadequacy

Inadequacies in biological models are rarely addressed explicitly, even though this neglect can severely bias parameter estimation. Though there have been attempts to formalise methods for modelling these inadequacies (Section 3.3.2), it can be argued that they are ad hoc, and require sensitive user-informed tuning. Stochastic processes can be directly modelled [75], [82] and fitted using various Bayesian techniques, but they have not been used to indicate the quality of the mechanistic model as a representation of the true process. Further, these stochastic processes are typically strongly parametric, and misspecification of the noise model can also cause problems in the parameter estimation process.

Tools For Identifiability

As noted in previous sections, there exist a wide arsenal of tools for performing identifiability analysis. However, they all suffer from their own deficiencies, which limit their applicability for analysing real systems. The DAISY tool, for example is platform constrained due to its implementation as a modification of the REDUCE program, and can take large amounts of time due to the nature of the algorithms used. However, methods are emerging for exploiting structural identifiability results, particularly from sensitivity analysis methods, to reduce models to identifiable forms [99], [100]. On the other hand, ready-to-use tools for profile likelihood are not yet widely accessible, even though their implementation is often quite straightforward. Further, they require user intervention to determine suitable reductions of the model, due to the lack of . An automated plug-and-play tool for performing practical identifiability analysis and

model reduction would be extremely beneficial to modellers in the epidemiological field, and also further afield.

A further question is how identifiability analysis, for practical identifiability in particular, extends to partial differential equations and stochastic systems. Structural identifiability analysis of PDEs have been performed for limited nonlinear cases [101], [102]. A similar extension has been made for structural identifiability of stochastic process models [103]. However, practical identifiability, where the analysis is done a posteriori in tandem with data, has not been done for either class of model.

Real-Time Prediction

Real-time prediction is a difficult venture. This is attested to by the fact that all of the cited literature in this review has been of either

- hypothetical simulation or
- retroactive prediction (*"retro-casting"*).

As mentioned above, health information infrastructure is not yet at a point where real-time data collection and reporting is possible, even in developed countries such as the United States. Some attempts have been made to assimilate auxiliary data in order to build regression-style models for prediction. One famous attempt is Google Flu Trends, that attempted to predict flu incidence by considering search data collated from their users. After missing the 2009 A/H1N1 out-of-season pandemic, the engine also overestimated incidence rates in the 2011-2012 season [14]. The engine has since been retired.

Disregarding attempts at increasing the amount of information available, there is also the promise of utilising prior knowledge as auxiliary information. This is an extremely common method in many scientific studies, where parameters in the literature are used directly, but there are indications that parameter estimates in literature have a large degree of disagreement [104]. This poses the problem of how to best synthesise existing results into emerging predictions and cases.

Further, there is a lack of a true standard for comparison of model fitting [15]. Because the field is plagued with identifiability problems and naïve methodology, direct comparison of quality of fit cannot be made. From a methodological standpoint, it can be easy to state that the least-squares objective function is the best unbiased measure of how well a model can fit the data. However, this validation technique would only be valid for checking out-of-sample data, which could be viewed as ethically objectionable, since it would require the practitioner to estimate without using all data available, and producing a poorer estimate.

References

[1] R. Ross, *The prevention of malaria*, 2nd. London: Murray, 1911.

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

[3] J. Drylewicz, D. Commenges, and R. Thiebaut, "Maximum a Posteriori Estimation in Dynamical Models of Primary HIV Infection," *Statistical Communications in Infectious Diseases*, vol. 4, no. 1, Jan. 2012, ISSN: 1948-4690. DOI: 10.1515/1948-4690.1040.

[4] C. Viboud, L. Simonsen, and G. Chowell, "A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks.," *Epidemics*, vol. 15, pp. 27–37, 2016, ISSN: 1878-0067. DOI: 10.1016/j.epidem.2016.01.002.

[5] G. Chowell, C. Castillo-Chavez, P. W. Fenimore, C. M. Kribs-Zaleta, L. Arriola, and J. M. Hyman, "Model Parameters and Outbreak Control for SARS," *Emerging Infectious Diseases*, vol. 10, no. 7, pp. 1258–1263, Jul. 2004, ISSN: 1080-6040. DOI: 10.3201/eid1007.030647.

[6] F. Carrat and A.-J. Valleron, "Epidemiologic Mapping using the "Kriging" Method: Application to an Influenza-like Epidemic in France," *American Journal of Epidemiology*, vol. 135, no. 11, pp. 1293–1300, Jun. 1992, ISSN: 1476-6256. DOI: 10.1093/oxfordjournals.aje.a116236.

[7] A. B. Lawson and H.-R. Song, "Bayesian hierarchical modeling of the dynamics of spatio-temporal influenza season outbreaks," *Spatial and Spatio-temporal Epidemiology*, vol. 1, no. 2-3, pp. 187–195, Jul. 2010, ISSN: 1877-5845. DOI: 10.1016/J.SSTE.2010.03.001.

[8] E. O. Nsoesie, M. Marathe, and J. S. Brownstein, "Forecasting peaks of seasonal influenza epidemics," *PLoS Currents*, vol. 5, no. Outbreaks, 2013, ISSN: 21573999. DOI: 10.1371/currents.outbreaks.bb1e879a23137022ea79a8c508b030bc.

[9] E. Dantas, M. Tosin, and A. Cunha Jr, "Calibration of a SEIR-SEI epidemic model to describe the Zika virus outbreak in Brazil," *Applied Mathematics and Computation*, vol. 338, pp. 249–259, Dec. 2018, ISSN: 0096-3003. DOI: 10.1016/J.AMC.2018.06.024.

[10] C. Tönsing, J. Timmer, and C. Kreutz, "Profile likelihood-based analyses of infectious disease models," *Statistical Methods in Medical Research*, vol. 27, no. 7, pp. 1979–1998, Jul. 2018, ISSN: 0962-2802. DOI: 10.1177/0962280217746444.

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

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

[13] CDC. "FluSight: flu Forecasting." (), [Online]. Available: https://www.cdc.gov/flu/weekly/flusight/index.html.

[14] D. Lazer, R. Kennedy, G. King, and A. Vespignani, "Big data. The parable of Google Flu: traps in big data analysis.," *Science (New York, N.Y.)*, vol. 343, no. 6176, pp. 1203–5, Mar. 2014. DOI: 10.1126/science.1248506.

[15] P. Chakraborty, B. Lewis, S. Eubank, J. S. Brownstein, M. Marathe, and N. Ramakrishnan, "What to know before forecasting the flu," *PLOS Computational Biology*, vol. 14, no. 10, M. Salathé, Ed., e1005964, Oct. 2018, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005964.

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

[17] J. Heffernan and M. Keeling, "Implications of vaccination and waning immunity," *Proceedings of the Royal Society B: Biological Sciences*, vol. 276, no. 1664, pp. 2071–2080, Jun. 2009, ISSN: 0962-8452. DOI: 10.1098/rspb.2009.0057.

[18] J. Dushoff, J. B. Plotkin, S. A. Levin, and D. J. D. Earn, "Dynamical resonance can account for seasonality of influenza epidemics.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 48, pp. 16 915–6, Nov. 2004, ISSN: 0027-8424. DOI: 10.1073/pnas.0407293101.

[19] J. Shaman and M. Kohn, "Absolute humidity modulates influenza survival, transmission, and seasonality.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 9, pp. 3243–8, Mar. 2009, ISSN: 1091-6490. DOI: 10.1073/pnas.0806852106.

[20] D. Tompkins, A. Brock, G. Jones, *et al.*, "Modelling the Impacts of Climate Change On Infectious Diseases in New Zealand Health Analysis & Information For Action (HAIFA)," Environmental Science and Research Limited, Porirua, Tech. Rep., 2012, p. 64.

[21] F. Brauer and C. Castillo-Chavez, *Mathematical Models for Communicable Diseases*, 84. 2013, ISBN: 9781611972412. DOI: 10.1137/1.9781611972429.

[22] R. Yaari, I. Dattner, and A. Huppert, "A two-stage approach for estimating the parameters of an age-group epidemic model from incidence data," *Statistical Methods in Medical Research*, vol. 27, no. 7, pp. 1999–2014, Jul. 2018, ISSN: 0962-2802. DOI: 10.1177/0962280217746443.

[23] O. Diekmann, J. Heesterbeek, and J. Metz, "On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations," *Journal of Mathematical Biology*, vol. 28, no. 4, pp. 365–382, Jun. 1990, ISSN: 0303-6812. DOI: 10.1007/BF00178324.

[24] M. T. Meehan, D. Cocks, J. M. Trauer, and E. S. McBryde, *Coupled, multi-strain SIS, SIR and SIRS epidemic models*, 2016.

[25] M. T. Meehan, D. G. Cocks, J. M. Trauer, and E. S. McBryde, "Coupled, multi-strain epidemic models of mutating pathogens," *Mathematical Biosciences*, vol. 296, pp. 82–92, Feb. 2018, ISSN: 0025-5564. DOI: 10.1016/J.MBS.2017.12.006.

[26] Z. Feng, A. N. Hill, P. J. Smith, and J. W. Glasser, "An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or preferential mixing," *Journal of Theoretical Biology*, vol. 386, pp. 177–187, Dec. 2015, ISSN: 0022-5193. DOI: 10.1016/J.JTBI.2015.09.006.

[27] L. J. S. Allen, "An introduction to stochastic epidemic models," in *Mathematical Epidemiology*, F. Brauer, P. van den Driessche, and J. Wu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 81–130, ISBN: 978-3-540-78911-6. DOI: 10.1007/978-3-540-78911-6_3.

[28] L. J. S. Allen, S. R. Jang, and L.-I. Roeger, "Predicting population extinction or disease outbreaks with stochastic models," *Letters in Biomathematics*, vol. 4, no. 1, pp. 1–22, Jan. 2017, ISSN: 2373-7867. DOI: 10.1080/23737867.2016.1264870.

[29] J. Snow, *On the Mode of Communication of Cholera - John Snow - Google Books*, Second Edition, much Enlarged. London: John Churchill, New Burlington Street, 1855.

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

[31] L. A. Rvachev and I. M. Longini, "A mathematical model for the global spread of influenza," *Mathematical Biosciences*, vol. 75, no. 1, pp. 3–22, Jul. 1985, ISSN: 0025-5564. DOI: 10.1016/0025-5564(85)90064-1.

[32] X. Li, H. Tian, D. Lai, and Z. Zhang, "Validation of the gravity model in predicting the global spread of influenza.," *International journal of environmental research and public health*, vol. 8, no. 8, pp. 3134–43, 2011, ISSN: 1660-4601. DOI: 10.3390/ijerph8083134.

[33] H. Heesterbeek, R. M. Anderson, V. Andreasen, *et al.*, "Modeling infectious disease dynamics in the complex landscape of global health.," *Science (New York, N.Y.)*, vol. 347, no. 6227, aaa4339, Mar. 2015, ISSN: 1095-9203. DOI: 10.1126/science.aaa4339.

[34] D. G. Kendall, "Mathematical Models of The Spread of Infections, Mathematics and Computer Science in Biology and Medicine," in *Mathematics and Computer Science in Biology and Medicine*, London: H.M.S.O., 1965, pp. 218–225.

[35] J. D. Murray, *Mathematical biology. I, Introduction*, 2nd. Springer, 2002, p. 551, ISBN: 9780387224374.

[36] ——, *Mathematical biology. II Spatial models and biomedical applications*, 2nd. Springer, 2003, p. 811, ISBN: 9780387224381.

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

[38] P. Weng and X.-Q. Zhao, "Spreading speed and traveling waves for a multi-type SIS epidemic model," *Journal of Differential Equations*, vol. 229, no. 1, pp. 270–296, Oct. 2006, ISSN: 0022-0396. DOI: 10.1016/J.JDE.2006.01.020.

[39] S. Ai and R. Albashaireh, "Traveling Waves in Spatial SIRS Models," *Journal of Dynamics and Differential Equations*, vol. 26, no. 1, pp. 143–164, 2014, ISSN: 10407294. DOI: 10.1007/s10884-014-9348-3.

[40] S. Cui and M. Bai, "Mathematical analysis of population migration and its effects to spread of epidemics," Mar. 2014. arXiv: 1403.5351.

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

[42] S. Pei, S. Kandula, W. Yang, and J. Shaman, "Forecasting the spatial transmission of influenza in the United States," *Proceedings of the National Academy of Sciences*, vol. 115, no. 11, pp. 2752–2757, Mar. 2018, ISSN: 0027-8424. DOI: 10.1073/pnas.1708856115.

[43] J. Arino, J. R. Davis, D. Hartley, R. Jordan, J. M. Miller, and P. van den Driessche, "A multi-species epidemic model with spatial dynamics," *Mathematical Medicine and Biology: A Journal of the IMA*, vol. 22, no. 2, pp. 129–142, Jun. 2005, ISSN: 1477-8602. DOI: 10.1093/imammb/dqi003.

[44] D. Bichara, Y. Kang, C. Castillo-Chavez, R. Horan, and C. Perrings, "SIS and SIR Epidemic Models Under Virtual Dispersal," *Bulletin of Mathematical Biology*, vol. 77, no. 11, pp. 2004–2034, Nov. 2015, ISSN: 0092-8240. DOI: 10.1007/s11538-015-0113-5.

[45] C. Castillo-Chavez, D. Bichara, and B. R. Morin, "Perspectives on the role of mobility, behavior, and time scales in the spread of diseases," *Proceedings of the National Academy of Sciences*, vol. 113, no. 51, pp. 14 582–14 588, Dec. 2016, ISSN: 0027-8424. DOI: 10.1073/PNAS.1604994113.

[46] J. Poot, O. Alimi, M. P. Cameron, and D. C. Maré, "The gravity model of migration: the successful comeback of an ageing superstar in regional science," *Investigaciones Regionales*, no. 36, pp. 63–86, 2016.

[47] J. Truscott and N. M. Ferguson, "Evaluating the Adequacy of Gravity Models as a Description of Human Mobility for Epidemic Modelling," *PLoS Computational Biology*, vol. 8, no. 10, M. Pascual, Ed., e1002699, Oct. 2012, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002699.

[48] D. L. Smith, B. Lucey, L. A. Waller, J. E. Childs, and L. A. Real, "Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 6, pp. 3668–72, Mar. 2002, ISSN: 0027-8424. DOI: 10.1073/pnas.042400799.

[49] G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud, "Mathematical models to characterize early epidemic growth: A review.," *Physics of life reviews*, vol. 18, pp. 66–97, 2016, ISSN: 1873-1457. DOI: 10.1016/j.plrev.2016.07.005.

[50] N. Bifolchi, R. Deardon, and Z. Feng, "Spatial approximations of network-based individual level infectious disease models," *Spatial and Spatio-temporal Epidemiology*, vol. 6, pp. 59–70, Sep. 2013, ISSN: 1877-5845. DOI: 10.1016/J.SSTE.2013.07.001.

[51] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The journal of physical chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[52] X. Zhang, T. Zhang, A. A. Young, and X. Li, "Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data," *PLoS ONE*, vol. 9, no. 2, Y. Yang, Ed., e88075, Feb. 2014, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0088075.

[53] G. Chowell, C. Viboud, L. Simonsen, and S. M. Moghadas, "Characterizing the reproduction number of epidemics with early subexponential growth dynamics," *Journal of The Royal Society Interface*, vol. 13, no. 123, p. 20 160 659, Oct. 2016, ISSN: 1742-5689. DOI: 10.1098/rsif.2016.0659.

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

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

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

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

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

[59] G. Wahba, *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Jan. 1990, ISBN: 978-0-89871-244-5. DOI: 10.1137/1.9781611970128.

[60] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B -splines and penalties," *Statistical Science*, vol. 11, no. 2, pp. 89–121, May 1996, ISSN: 0883-4237. DOI: 10.1214/ss/1038425655.

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

[62] M. Mogstad and M. Wiswall, "Instrumental variables estimation with partially missing instruments," *Economics Letters*, vol. 114, no. 2, pp. 186–189, Feb. 2012, ISSN: 0165-1765. DOI: 10.1016/J.ECONLET.2011.10.013.

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

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

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

[66] E. Haber and U. M. Ascher, "Preconditioned all-at-once methods for large, sparse parameter estimation problems," *Inverse Problems*, vol. 17, no. 6, pp. 1847–1864, Dec. 2001, ISSN: 0266-5611. DOI: 10.1088/0266-5611/17/6/319.

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

[68] M. Benning and M. Burger, "Modern regularization methods for inverse problems," *Acta Numerica*, vol. 27, pp. 1–111, May 2018, ISSN: 0962-4929. DOI: 10.1017/S0962492918000016.

[69] O. Scherzer, "The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems," *Computing*, vol. 51, no. 1, pp. 45–60, Mar. 1993, ISSN: 0010-485X. DOI: 10.1007/BF02243828.

[70] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY: Springer New York, 2009, ISBN: 978-0-387-84857-0. DOI: 10.1007/978-0-387-84858-7.

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

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

[73] Y. Pawitan, *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, 2001, ISBN: 978-0199671229.

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

[75] A. Chatzilena, E. van Leeuwen, O. Ratmann, M. Baguelin, and N. Demiris, "Contemporary statistical inference for infectious disease models using Stan," *Epidemics*, p. 100 367, Oct. 2019, ISSN: 1755-4365. DOI: 10.1016/J.EPIDEM.2019.100367. arXiv: 1903.00423.

[76] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, *et al.*, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953, ISSN: 0021-9606. DOI: 10.1063/1.1699114.

[77] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, vol. 57, no. 1, p. 97, Apr. 1970, ISSN: 00063444. DOI: 10.2307/2334940.

[78] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984, ISSN: 0162-8828. DOI: 10.1109/TPAMI.1984.4767596.

[79] M. Betancourt, "A Conceptual Introduction to Hamiltonian Monte Carlo," Jan. 2017. arXiv: 1701.02434.

[80] A. Kucukelbir, R. Ranganath, A. Gelman, and D. M. Blei, "Automatic Variational Inference in Stan," Jun. 2015. arXiv: 1506.03431.

[81] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation*, ser. Texts in Applied Mathematics. Cham: Springer International Publishing, 2015, vol. 62, ISBN: 978-3-319-20324-9. DOI: 10.1007/978-3-319-20325-6.

[82] E. L. Ionides, C. Bretó, and A. A. King, "Inference for nonlinear dynamical systems.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 49, pp. 18 438–43, Dec. 2006, ISSN: 0027-8424. DOI: 10.1073/pnas.0603181103.

[83] T. Stocks, T. Britton, and M. Höhle, "Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany," *Biostatistics*, Sep. 2018, ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxy057.

[84] J. P. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, ser. Applied Mathematical Sciences. New York: Springer-Verlag, 2005, vol. 160, pp. i–339, ISBN: 0-387-22073-9. DOI: 10.1007/b138659.

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

[86] I. Swameye, T. G. Muller, J. Timmer, O. Sandra, and U. Klingmuller, "Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 3, pp. 1028–33, Feb. 2003, ISSN: 0027-8424. DOI: 10.1073/pnas.0237333100.

[87] V. Raia, M. Schilling, M. Bohm, *et al.*, "Dynamic Mathematical Modeling of IL13-Induced Signaling in Hodgkin and Primary Mediastinal B-Cell Lymphoma Allows Prediction of Therapeutic Targets," *Cancer Research*, vol. 71, no. 3, pp. 693–704, Feb. 2011, ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-10-2987.

[88] R. Muñoz-Tamayo, L. Puillet, J. B. Daniel, *et al.*, "Review: To be or not to be an identifiable model. Is this a relevant question in animal science modelling?" *Animal*, vol. 12, no. 4, pp. 701–712, Apr. 2018, ISSN: 1751732X. DOI: 10.1017/S1751731117002774.

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

[90] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, "Universally Sloppy Parameter Sensitivities in Systems Biology Models," *PLoS Computational Biology*, vol. 3, no. 10, e189, 2007, ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.0030189.

[91] M. K. Transtrum, B. B. Machta, and J. P. Sethna, "Geometry of nonlinear least squares with applications to sloppy models and optimization," *Physical Review E*, vol. 83, no. 3, p. 036 701, Mar. 2011, ISSN: 1539-3755. DOI: 10.1103/PhysRevE.83.036701.

[92] O. J. Maclaren and R. Nicholson, "What can be estimated? Identifiability, estimability, causal inference and ill-posed inverse problems," Apr. 2019. arXiv: 1904.02826.

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

[94] B. D. Dalziel, M. S. Y. Lau, A. Tiffany, *et al.*, "Unreported cases in the 2014-2016 Ebola epidemic: Spatiotemporal variation, and implications for estimating transmission.," *PLoS neglected tropical diseases*, vol. 12, no. 1, e0006161, 2018, ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0006161.

[95] P. Klepac, S. Kissler, and J. Gog, "Contagion! The BBC Four Pandemic — The model behind the documentary," *Epidemics*, vol. 24, pp. 49–59, Sep. 2018, ISSN: 1755-4365. DOI: 10.1016/J.EPIDEM.2018.03.003.

[96] Centre for Strategic and International Studies, "Can Digital Health Help Stop the Next Epidemic? | Center for Strategic and International Studies," Centre for Strategic and International Studies, Tech. Rep., Oct. 2019.

[97] B. Bonté, J.-D. Mathias, and R. Duboz, "Moment Approximation of Infection Dynamics in a Population of Moving Hosts," *PLoS ONE*, vol. 7, no. 12, R. Huerta-Quintanilla, Ed., e51760, Dec. 2012, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0051760.

[98] M. J. Keeling, T. House, A. J. Cooper, and L. Pellis, "Systematic Approximations to Susceptible-Infectious-Susceptible Dynamics on Networks," *PLOS Computational Biology*, vol. 12, no. 12, K. Koelle, Ed., e1005296, Dec. 2016, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005296.

[99] T. J. Snowden, P. H. van der Graaf, and M. J. Tindall, "Methods of Model Reduction for Large-Scale Biological Systems: A Survey of Current Methods and Trends," *Bulletin of Mathematical Biology*, vol. 79, no. 7, pp. 1449–1486, Jul. 2017, ISSN: 0092-8240. DOI: 10.1007/s11538-017-0277-2.

[100] A. Pandey and R. M. Murray, "An automated model reduction tool to guide the design and analysis of synthetic biological circuits," *bioRxiv*, p. 640 276, May 2019. DOI: 10.1101/640276.

[101] A. Perasso, B. Laroche, Y. Chitour, and S. Touzeau, "Identifiability analysis of an epidemiological model in a structured population," *Journal of Mathematical Analysis and Applications*, vol. 374, no. 1, pp. 154–165, Feb. 2011, ISSN: 0022-247X. DOI: 10.1016/J.JMAA.2010.08.072.

[102] S. Zhu, N. Verdière, L. Denis-Vidal, and D. Kateb, "Identifiability analysis and parameter estimation of a chikungunya model in a spatially continuous domain," *Ecological Complexity*, vol. 34, pp. 80–88, May 2018, ISSN: 1476-945X. DOI: 10.1016/J.ECOCOM.2017.12.004.

[103] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf, "Sensitivity, robustness, and identifiability in stochastic chemical kinetics models.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 21, pp. 8645–50, May 2011, ISSN: 1091-6490. DOI: 10.1073/pnas.1015814108.

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