Comparing and Examining Different Methods of Determining Structural and Practical Identifiability of a Vector-Host Model

Joel Nathe[1], Kevin Phan[2], and Alyssa Sharma[3]

[1]University of Minnesota
[2]Harvey Mudd College
[3]Appalachian State University

August 5, 2022

Abstract

In order to predict the spread of outbreaks of vector-host diseases such as dengue fever and malaria, researchers create compartmental models which employ ordinary differential equations. It is an expanding problem of interest to investigate the structural or a priori identifiability and the practical identifiability of compartmental models. We compare the accuracy and efficiency of multiple methods for computing structural and practical identifiability of an SIS (Susceptible-Infected-Susceptible) vector-host model. We performed structural identifiability analysis of the model via a differential algebra approach, an approximation of the differential algebra approach using truncated power series and modulo arithmetic, and a combination of the differential algebra approach and the Taylor series approach. We also performed practical identifiability analysis of the model via the Monte Carlo and Profile Likelihood methods. We found that there were instances where the approximation of the differential algebra approach differed from the results of the combination of the differential algebra approach and Taylor series approach, and computed the structural identifiability of models with certain fixed parameters and initial conditions. For the Monte Carlo method, we found that the choices of both the error model and the initial guess of parameters have a large impact on practical identifiability results. For profile likelihood, we derive formula for computing profile likelihood for relative normal noise distribution, and we explore how data and different noise distributions can impact parameter estimation, confidence intervals, and identifiability.

Contents

1 Introduction

The study of vector borne diseases is pivotal in mitigating the mortality rate in countries with tropical climates. According to the World Health Organization, vector borne diseases comprise at least 17% of all infectious diseases[WHO, 2020]. In 2020, there were approximately 670,000 deaths worldwide due to malaria alone, with dengue fever contributing to approximately 22,000 deaths, most of which were children [NIAID, 2016]. Vector-borne diseases are infectious diseases that spread through some other organism, called a vector, to humans, rather than directly from human to human [Brauer et al., 2008]. The Bubonic plague, which spread through rats, decimated Europe's population in the 14th century, providing an early example of a vector-borne disease [Glatter and Finkelman, 2021]. Malaria and dengue fever are both caused parasites which are transmitted through mosquitoes.

With the advent of the Severe Acute Respiratory Syndrome Corona Virus Infection (SARS-CoV-1), epidemiological data is more prevalent than ever. This great increase in the prevalence of epidemiological data in recent years may cause researchers to experience the pitfall of trying to predict the spread of a disease solely based on available data or on the basic reproductive number R_0. R_0 is a decisive tool in infectious disease modelling, and is defined as the expected number of secondary infections produced by an index case in a completely susceptible population [Brauer et al., 2008]. R_0 can be used to determine the peak infection and final size of an epidemic [Brauer et al., 2008]. Issues arise in predicting the spread of an infectious disease occurs at the beginning of a pandemic or epidemic, when epidemiological data is widely unavailable and unreliable. While an expression for R_0 can be easily created based on the model's equations, the parameters within the expression cannot be easily estimated when there is a lack of available data. The beginning of a pandemic is the time communities must implement measures to combat the disease's spread, and thus one of the most important times to be able to predict the spread disease. It is extremely difficult to accurately estimate R_0 at this point in an epidemic based on data, as the majority of the data does not exist. The lack of clinical data at this point in a pandemic also renders most data-based methods of prediction mute, calling for other methods of mathematical prediction. Compartmental models can be used to predict the spread of an epidemic or pandemic without relying on data. Compartmental models are deterministic models comprised of systems of Ordinary Differential Equations (ODEs). These models divide the population into time-based sections which represent the status of an individual with respect to the disease [Brauer et al., 2008].

Identifiability is a measure of whether the parameters of a model can be uniquely identified with the measurement data of input and output variables. Structural identifiability deals with continuous, noise-free data, while practical identifiability deals with data which contains noise. In order to estimate the true parameters of a compartmental model, that model must be both structurally and practically identifiable. Our goal is to compare different methods of calculating structural and practical identifiability of simple SIS vector-host compartmental models. To compute structural identifiability we use the differential algebra

approach, a combination of the differential algebra approach and the Taylor series approach, and the Exact Arithmetic Rank approach (EAR). We compute practical identifiability via the Monte Carlo simulation method as well as the profile likelihood approach.

2 Epidemiological Model

We focus on a Susceptible-Infected-Susceptible (SIS) vector-host model, where hosts can only be infected by vectors, and vice versa. Hosts do not gain immunity after recovering from infection, nor do vectors recover from infection at all. It is assumed that a vector's lifespan is too small to encompass recovery. The model has the following form:

$$\begin{aligned} S_h' &= \Pi_h - \mu_h S_h - \beta_h S_h I_v + \gamma I_h, \\ I_h' &= \beta_h S_h I_v - (\mu_h + \gamma) I_h, \\ S_v' &= \Pi_v - \mu_v S_v - \beta_v S_v I_h, \\ I_v' &= \beta_v S_v I_h - \mu_v I_v. \end{aligned} \tag{1}$$

with initial conditions $S_h(0) = S_h^{(0)}, S_v(0) = S_v^{(0)}, S_v(0) = S_v^{(0)}, I_v(0) = I_v^{(0)}$ [Brauer et al., 2008, p. 170]. In this model, μ_h represents the removal or death rate of the host population, while μ_v represents the removal or death rate of the vector population. Π_h represents the birth rate of the host population, while Π_v represents the birth rate of the vector population. The infection rate for hosts is β_h, while the infection rate for vectors is β_v. Notice that β_h interacts with the susceptible population of hosts and infected population of vectors, indicating that hosts can only be infected by an infected vector, and the situation is vice versa for β_v. γ represents the recovery rate of infected hosts. It is assumed that once a host is no longer infected, they return to the susceptible compartment, and do not gain immunity. It is assumed that the lifespan of a vector is too small for a vector to experience recovery, thus there is no recovery rate for vectors. Another version of this model we studied can be found by scaling our infection terms by population size, depicted in the following ODE system:

$$\begin{aligned} S_h' &= \Pi_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ S_v' &= \Pi_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v. \end{aligned} \tag{2}$$

where $N_h = S_h + I_h$ and $N_v = S_v + I_v$. We see that these satisfy the differential equations $\frac{dN_h}{dt} = \Pi_h - \mu_h N_h, \frac{dN_v}{dt} = \Pi_v - \mu_v N_v$. Additionally, when our population is fixed (Inital

Conditions $S_h^{(0)} + I_h^{(0)} = \frac{\Pi_h}{\mu_h}, S_v^{(0)} + I_v^{(0)} = \frac{\Pi_v}{\mu_v}$), these models can be made equivalent by setting the β_h and β_v in 1 to the values of $\frac{\beta_h}{N_v}$ and $\frac{\beta_v}{N_h}$ from 2.

Additionally, as long as the population size is fixed, the model 2 is invariant to what the population size is. If we consider the percentages of the population in each compartment,

$$\begin{aligned}
\left(\frac{S_h}{N_h}\right)' &= \Pi_h - \mu_h \frac{S_h}{N_h} - \frac{\beta_h S_h I_v}{N_h N_v} + \gamma \frac{I_h}{N_h}, \\
\left(\frac{I_h}{N_h}\right)' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\
\left(\frac{S_v}{N_v}\right)' &= \Pi_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\
\left(\frac{I_v}{N_v}\right)' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v.
\end{aligned} \tag{3}$$

We see that 3 is 2, with percentage of the population in each compartment substituted in.

Next, we analyze the model by calculating the R_0 value using the next generation matrix method.

$$R_0 = \sqrt{\frac{\beta_h}{\mu_v} \frac{\beta_v}{\mu_h + \gamma}} \tag{4}$$

This is the product of the new infections from an infected vector $\left(\frac{\beta_h}{\mu_v}\right)$ and the new infections from an infected host $\left(\frac{\beta_v}{\mu_h + \gamma}\right)$. Since it takes two generations for an infection to go from a host to a vector and then back to a host, the R_0 value is the square root is this product.

We then find the equilibria of the model. When $R_0 \leq 1$, the only equilibrium is the disease-free one of $S_h = \frac{\Pi_h}{\mu_h}, S_v = \frac{\Pi_v}{\mu_v}, I_h = I_v = 0$. However, when $R_0 > 1$, we have the endemic equilibrium of

$$\begin{aligned}
S_h &= N_h \frac{\mu_h + \gamma + \frac{\beta_h}{R_0^2}}{\mu_h + \gamma + \beta_h} \\
I_h &= N_h \beta_h \frac{1 - \frac{1}{R_0^2}}{\mu_h + \gamma + \beta_h} \\
S_v &= N_v \frac{\mu_v + \frac{\beta_v}{R_0^2}}{\mu_v + \beta_v} \\
I_v &= N_v \beta_v \frac{1 - \frac{1}{R_0^2}}{\mu_v + \beta_v}
\end{aligned} \tag{5}$$

At this equilibrium, we see that the ratio

$$\frac{I_h/N_h}{I_v/N_v} = \frac{\beta_h}{\beta_v} \frac{\mu_v + \beta_v}{\mu_h + \gamma + \beta_h}$$

3 Structural Identifiability

In order to estimate the true parameters of an epidemiological model via practical identifiability, it is first necessary to determine whether or not a model is structurally identifiable. Identifiability is a measure of whether or not a model can be uniquely identified with the measurement data of input-output variables. To compute structural identifiability, we assume both that our data is noise free and that our model experiences no errors. We used 3 methods to quantify structural identifiability: the differential algebra approach, a mix of the Taylor series approach and differential algebra approach, and finally the exact arithmetic rank approach. The corresponding programs we used to generate results are the Differential Algebra Identifiability of SYstems software (DAISY) in Reduce, the Systems Identifiability ANalyzer (SIAN) software in Maple, and the IdentifiabilityAnalysis package in Wolfram Alpha.

3.1 Differential Algebra Approach

DAISY stands for Differential Algebra for Identifiability of SYstems . DAISY implements a mixed differential algebra algorithm [Bellu et al., 2007]. The software implements Ritt's algorithm to eliminate state variables which happen to be non-observable in order to calculate coefficients of an input-output relation [Bellu et al., 2007]. The coefficients of the input-output relation are linearly parameterized by algebraic functions of unknown parameters, called the exhaustive summary [Bellu et al., 2007]. This input-output relation is then solved via the Buchberger algorithm, and returns a gröbner basis [Bellu et al., 2007]. This basis is used to determine the number of solutions to the system, which in tern determines the identifiability of the system.

One reason why one may choose to use DAISY over other software for computing structural identifiability is that the accuracy of DAISY is not determined via a given probability. Some cons about DAISY occur when testing the structural identifiability of a non-linear systems. When compared with the other algorithms used to compute structural identifiability, DAISY was the slowest method. Running structural identifiability on systems which were not quickly solved often took weeks, and many of the models we ran returned no answer in terms of identifiability due to a memory error. Compared to the SIAN and EAR methods, DAISY took the largest toll on the hardware it was being run on, using large amounts of RAM and memory.

3.2 Differential Algebra and Taylor Series Approach

Some of the models which we studied were too computationally heavy to determine identifiability via the differential algebra approach DAISY uses. In these cases, we decided to switch from the DAISY software in reduce to the SIAN software in Maple. SIAN stands for Structural Identifiability ANalyzer [Hong et al., 2020]. SIAN is a Monte Carlo randomized algorithm [Hong et al., 2020]. Because assessing global identifiability is an extremely

complex problem, many software first attempt to determine local identifiability, which is a much less computationally expensive problem [Hong et al., 2020]. In terms of outputs, SIAN returns whether or not each unknown parameter or initial condition is globally identifiable, locally but not globally identifiable, or non-identifiable. SIAN's algorithm is based on a combination of a differential algebra approach and a Taylor Series approach. The fact that SIAN uses user-defined probability to set the correctness of it's resulting outputs makes SIAN stand out from other identifiability programs [Hong et al., 2020].

A simple abstraction of the identifiability problem is to think of the problem as a map between unknown parameters and unknown initial conditions which are connected by fibers to the output data, which represents the functions of the corresponding solution. The next step is to formalize this mapping using differential algebra, reducing this map to a new map between finite dimensional spaces [Hong et al., 2020]. The method used to reduce this map to a map between finite dimensional spaces is to replace the output function with a truncation of its Taylor series using theorem 3.16 to provide criterion which contains enough information for the identifiability checking. This is accomplished via choosing the fiber of a randomly chosen point as opposed to a generic fiber, using the Demillo-Lipton-Schwartz-Zippel Lemma [Hong et al., 2020]. After considering a fiber at a random point, the problem turns into checking the consistency of a system of polynomial equations and inequations [Hong et al., 2020]. There are 3 approaches to this problem: symbolic, symbolic-numeric, and numeric. Researchers were surprised to find that Gröbner bases computations are efficient enough for moderate-size problems and significantly out preformed numerical algebraic geometry software [Hong et al., 2020].

3.3 Exact Arithmetic Rank Approach

The most computationally efficient approach we found for computing the structural identifiability of various SIS models was the Exact Arithmetic Rank approach (EAR) software available in the IdentifiabilityAnalysis package in Mathematica. This software uses an approximation of the differential algebra approach to compute structural identifiability [Karlsson et al., 2012]. Some reasons why the EAR approach might not be used to compute structural identifiabililty include the fact that the approach is only accurate to a specified probability, as well as the fact that the Wolfram Alpha package is only capable of identifying whether or not a parameter is locally identifiable [Karlsson et al., 2012]. This is evident in cases when the output of the Mathematica software differs from the output of the Maple software. Some reasons why many would prefer the EAR approach to other methods of computing structural identifiabililty include that EAR produces results faster than the Maple or DAISY approaches, as well as the fact that the EAR approach uses much less RAM and space in memory than the other approaches. The EAR software was created to be a more computationally efficient alternative to DAISY, specifically to combat the expression swell DAISY often encounters. Expression swell occurs when the integers in computations become too large for the device calculating the computations to handle.

DAISY usually encounters expression swell when it uses repeated Lie Derivatives to directly calculate the Jacobian matrix [Karlsson et al., 2012]. Rather than using Lie Derivatives, EAR software indirectly calculates the entries of the Jacobian using truncated power series expansion of partial derivatives of the output with respect to initial conditions and parameters [Karlsson et al., 2012]. Another difference between DAISY and EAR is that EAR software cuts down on time by doing calculations using modulo of a large prime number, rather than the exact values, which prevents the risk of switching to slow software arithmetics for large integers [Karlsson et al., 2012].

4 Practical Identifiability

4.1 Monte Carlo Simulation

1. We start by solving the model numerically at the true parameters $\boldsymbol{\theta}_0$ at the discrete points in time $\{t_i\}_{i=0}^n$. We used MATLAB `ode45`, and chose t_j to be the data for day j. This gives $n+1$ evenly spaced points, one for every day between day 0 and day n. We then take our observations of this data, defined as the function $\mathbf{g}(t_i, \boldsymbol{\theta}_0)$.

   We generally have these observations be prevalence data, the size of the population currently in the infected compartments. But they can also be incidence data, the number of new infections since the last observation, or cumulative data, the current total of infections in the epidemic.

2. We create M (generally 1000) sets of data with noise using the following model

$$y_i = \mathbf{g}(t_i, \boldsymbol{\theta}_0) + (\mathbf{g}(t_i, \boldsymbol{\theta}_0))^{\xi} \epsilon_i \tag{6}$$

   where ϵ_i are error distributions. We chose ϵ_i to be normally distributed with variance σ^2, which makes them i.i.d.

   The amount in which noise scales with the size of our observations is determined by ξ. When $\xi = 0$(absolute noise), the noise does not scale with the size of our observations. When $\xi = 1$(relative noise), the noise is percentage error, completely scaled with the magnitude of the observations[Capaldi et al., 2012]. We chose relative noise, following [Tuncer and Le, 2018],

3. Next, we define an objective function to measure how parameters fit the noisy data. We used the least squares method, which results in 7[Capaldi et al., 2012]. In addition, in order to decrease the computation time of `ode45`, we set our objective function to return 10^{10} if any parameter was either negative or greater than 100.

$$\boldsymbol{\theta}_j = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \frac{1}{(\mathbf{g}(t_i, \boldsymbol{\theta}))^{2\xi}} \left| y_i - \mathbf{g}(t_i, \boldsymbol{\theta}) \right|^2 \tag{7}$$

For our objective function, we chose $\xi = 0$, ordinary least squares(OLS). To solve for $\boldsymbol{\theta}_j$ in 7, we used the Nelder-Mead algorithm implemented by the MATLAB function `fminsearch`[Lagarias et al., 1998].

4. Finally, we find the average relative error value for each parameter

$$\text{ARE}(\theta^{(k)}) = 100\% \frac{1}{M} \sum_{j=1}^{M} \frac{\left|\theta_j^{(k)} - \theta_0^{(k)}\right|}{\theta_0^{(k)}} \tag{8}$$

We consider $\theta^{(k)}$ practically identifiable if $ARE(\theta^{(k)}) \leq \sigma$[Tuncer and Le, 2018].

4.1.1 Factors Independent of Model Structure

When running the Monte Carlo method on a model, there are many factors which influence results but are independent of the model structure. Some of these are from the data used when fitting parameters, and others are from the methods of fitting parameters to the data. But they each have an impact on the practical identifiability results.

These factors include the time span that observations occur over, as well as how frequently they occur. They also include the type of data we use, and which populations we have data for. For the vector-host model, it is possible that there is data for only one of the populations, as obtaining accurate vector data may be difficult. For each of these possibilities, we investigated the identifiability of prevalence, incidence, and cumulative data.

For cumulative data, we can have our usual error model, which scales noise based on cumulative data. Or we can view cumulative data as the sum of noisy incidence data. Let $I(t_i, \boldsymbol{\theta})$ be incidence observations, and $C(t_i, \boldsymbol{\theta})$ be cumulative observations. When our error is in cumulative observations, our noisy data is

$$\begin{aligned} y_i &= C(t_i, \boldsymbol{\theta}) + C(t_i, \boldsymbol{\theta})\mathcal{N}(0, \sigma^2) \\ &= C(t_i, \boldsymbol{\theta}) + \mathcal{N}\left(0, \sigma^2 \left(\sum_{j=0}^{i} I(t_j, \boldsymbol{\theta})\right)^2\right) \end{aligned} \tag{9}$$

But if it's in incidence observations,

$$\begin{aligned} y_i &= \sum_{j=0}^{i} I(t_i, \boldsymbol{\theta}) + I(t_i, \boldsymbol{\theta})\mathcal{N}(0, \sigma^2) \\ &= C(t_i, \boldsymbol{\theta}) + \sum_{j=1}^{i} \mathcal{N}(0, \sigma^2 (I(t_i, \boldsymbol{\theta}))^2) \\ &= C(t_i, \boldsymbol{\theta}) + \mathcal{N}\left(0, \sigma^2 \sum_{j=0}^{i} (I(t_j, \boldsymbol{\theta}))^2\right) \end{aligned} \tag{10}$$

Since $\sum_{j=0}^{i}(I(t_j,\boldsymbol{\theta}))^2 \leq \left(\sum_{j=0}^{i} I(t_j,\boldsymbol{\theta})\right)^2$, if the noise is in incidence data, the variance is less than if it is in cumulative data.

In the process of fitting parameters to the data, there are also the choices of error model and objective function. We can choose between relative noise and absolute noise, as well as between ordinary and generalized least squares. And for the vector-host model, we must choose if the absolute noise and ordinary least squares scale with population size.

Finally, we must choose the optimization method. This includes the choice of optimization algorithm, and if the algorithm takes in bounds, the choice of bounds. It also includes the initial guess for the value of parameters.

4.2 Profile Likelihood

The maximum likelihood approach to estimating parameters view the ODE model as a statistical model and try to answer the question, "What parameters of the ODE model best generate the observed data?" Rather than viewing the observed data as random, the data is fixed and the parameters are unknown. The maximum likelihood approach exploits the fact that information is known about the distribution that the observed data come from to better estimate parameters. For instance, if the noisy data come from a Poisson distribution, then the objective function should use this fact to better fit the parameters. The maximum likelihood approach give us a natural way to incorporate knowledge of the noise distribution into the objective function. Furthermore, pointwise and simultaneous likelihood-based confidence intervals can be easily calculated for each parameter. Lastly, profile likelihood allows for a graphical method to check for identifiability through confidence intervals.

4.2.1 Maximum Likelihood Approach

The following two definitions are attributed to Larson's "An Introduction to Mathematical Statistics and Its Applications" [Larsen and Marx, 2006].

Definition 4.1. Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be a random sample of size n from the probability distribution function/probability mass function $f_Y(y_i;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters. The likelihood function is

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} f_Y(y_i;\boldsymbol{\theta}). \tag{11}$$

Definition 4.2. Let $L(\boldsymbol{\theta}|\mathbf{y}) = f_Y(y_i;\boldsymbol{\theta})$ be the likelihood function corresponding to the random samples $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ sampled from a probability distribution, where $\boldsymbol{\theta}$ is an unknown parameter. Let $\hat{\boldsymbol{\theta}}$ be a value of the parameter such that $L(\hat{\boldsymbol{\theta}}) \geq L(\boldsymbol{\theta})$ for all possible values of $\boldsymbol{\theta}$. Then, $\hat{\boldsymbol{\theta}}$ is called a maximum likelihood estimate for $L(\boldsymbol{\theta}|\mathbf{y})$.

For parameter estimation, we need to maximize the likelihood function. We will now reframe the system of ordinary differential equations and the observed data as a statistical model [Raue et al., 2009].

Definition 4.3. A model can be described by

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \boldsymbol{\theta}) \tag{12}$$

where $\mathbf{x}$ denotes the state variables and $\boldsymbol{\theta}$ denotes the parameters. Furthermore, let

$$y(t) = g(\mathbf{x}(t), \boldsymbol{\theta}) + \varepsilon(t) \tag{13}$$

where g is the observation function and ε is the measurement noise.

We will now find the maximum likelihood estimate when the measurement error $\varepsilon(t) \sim \mathcal{N}(0, \sigma^2)$, the observed data $\mathbf{y} \sim \text{Poi}(g(t_i, \boldsymbol{\theta}))$, and the measurement error $\varepsilon(t) \sim \mathcal{N}(0, (\eta \cdot g(t_i, \boldsymbol{\theta}))^2)$. Note that instead of writing $g(\mathbf{x}(t), \boldsymbol{\theta})$, we will write $g(t, \boldsymbol{\theta})$.

Example 4.4 (Measurement error follow a normal distribution with known variance σ^2 [Raue et al., 2009].)**.** Suppose that the measurement errors follow the distribution $\mathcal{N}(g(t_i, \boldsymbol{\theta}), \sigma^2)$ with unknown parameter $\boldsymbol{\theta}$ and known mean $\mu = 0$ and variance σ^2. We wish to find the parameter vector $\hat{\boldsymbol{\theta}}$ that maximize the likelihood. The likelihood is

$$L(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2)\sum_{i=1}^{n}(y_i - g(t_i, \boldsymbol{\theta}))^2/\sigma^2}. \tag{14}$$

We will find the negative log-likelihood as this transform the maximization problem to a minimization problem. The negative log-likelihood is

$$-2\log L(\boldsymbol{\theta}|\mathbf{y}) = n\log 2\pi + n\log\sigma^2 + \sum_{i=1}^{n}(y_i - g(t_i, \boldsymbol{\theta}))^2/\sigma^2 \tag{15}$$

or

$$-2\log L(\boldsymbol{\theta}|\mathbf{y}) = \text{const} + \sum_{i=1}^{n}(y_i - g(t_i, u, \boldsymbol{\theta}))^2/\sigma^2. \tag{16}$$

Thus, the maximum likelihood estimate is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n}(y_i - g(t_i, u, \boldsymbol{\theta}))^2/\sigma^2. \tag{17}$$

Note 4.5. Example 4.4 can be extended to for arbitrary number of observations. It can be shown that

$$\chi^2(\theta) = \sum_{j=1}^{m}\sum_{i=1}^{n}\left(\frac{y_{ij} - g(t_{ij}, \boldsymbol{\theta})}{\sigma_{ij}}\right)^2 \tag{18}$$

where y_{ij} is the ith data point of the jth observable and $g(t_{ij}, \boldsymbol{\theta})$ is the ith data point of the jth observable predicted by the parameter $\boldsymbol{\theta}$ at time point t_{ij}. [Raue et al., 2009].

Note 4.6. For n observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$, the negative log-likelihood function is

$$-2\log L(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n) = \sum_{i=1}^{n} -2\log L(\boldsymbol{\theta}_i|\mathbf{y}_i)$$

where $-2\log L(\boldsymbol{\theta}_i|\mathbf{y}_i)$ is the negative log-likelihood function for the observation $\mathbf{y}_i$. The maximum likelihood estimation of $-2\log L(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ is

$$\hat{\boldsymbol{\theta}} = \sum_{i=1}^{n} \hat{\boldsymbol{\theta}}_i$$

where $\hat{\boldsymbol{\theta}}_i$ is the maximum likelihood estimate of the negative log-likelihood $-2\log L(\boldsymbol{\theta}_i|\mathbf{y}_i)$.

In the paper *Identifiability and observability analysis for experimental design in nonlinear dynamical models* by Raue et. al., they were studying biochemical reactions in the field of system biology and specifically, the erythropoietin (Epo) and Epo receptor interaction. Due to a systematic error in their measurement device, there choose for their measurement error to follow normal distribution with mean $\mu = 0$ and known variance σ^2. The concentration of Epo is expected to follow the measurement error sampled by $\mathcal{N}(0, \sigma^2)$ at each time step [Raue et al., 2010]. However, while this assumption might be reasonable in the context of system biology, it might not be reasonable to choose a measurement error that follow the normal distribution $\mathcal{N}(0, \sigma^2)$. It have been shown and justified that epidemiological data can follow a Poisson distribution [Capaldi et al., 2012, Flanders and Kleinbaum, 1995]. Furthermore, if the error is relative to the number of infected cases, then it would also be reasonable to consider a relative error model. As such, we will find the maximum likelihood estimate when the observed data follow the Poisson distribution and the measurement error follow a relative error distribution.

Example 4.7 (Observed data follow a Poisson distribution [Eisenberg, 2009].)**.** Suppose that the observed data are given by Poisson random variables with mean $g(t_i, u, \theta)$ for $1 \leq i \leq n$. Let the observed data be denoted by $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and assume that the data is independent. Then, the likelihood is

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} \frac{g(t_i, \boldsymbol{\theta})^{y_i} e^{-g(t_i, \boldsymbol{\theta})}}{y_i!}. \tag{19}$$

The negative log-likelihood function is

$$-2\log L(\boldsymbol{\theta}|\mathbf{y}) = -2\log\left(\prod_{i=1}^{n}\frac{g(t_i,\boldsymbol{\theta})^{y_i}e^{-g(t_i,\boldsymbol{\theta})}}{y_i!}\right) \tag{20}$$

$$= -2\sum_{i=1}^{n}\log\left(\frac{g(t_i,\boldsymbol{\theta})^{y_i}e^{-g(t_i,\boldsymbol{\theta})}}{y_i!}\right) \tag{21}$$

$$= -2\sum_{i=1}^{n}\Big(\log(g(t_i,u,\boldsymbol{\theta})^{y_i}) + \log(e^{-g(t_i,\boldsymbol{\theta})}) - \log(y_i!)\Big) \tag{22}$$

$$= -2\sum_{i=1}^{n}\log(g(t_i,u,\boldsymbol{\theta})^{y_i}) + 2\sum_{i=1}^{n}g(t_i,\boldsymbol{\theta}) + 2\sum_{i=1}^{n}\log(y_i!). \tag{23}$$

Since the last term is a constant, the maximum likelihood estimate is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -2\sum_{i=1}^{n}\log(g(t_i,u,\boldsymbol{\theta})^{y_i}) + 2\sum_{i=1}^{n}g(t_i,\boldsymbol{\theta}). \tag{24}$$

Example 4.8 (Measurement error follow a relative error distribution.)**.** Suppose that the measurement errors follow a normal distribution with mean $\mu = 0$ and variance $(\eta\cdot g(t_i,\boldsymbol{\theta}))^2$. We wish to find the parameter vector $\hat{\boldsymbol{\theta}}$ that maximize the likelihood. Then, the likelihood is

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n}\frac{1}{\eta\cdot g(t_i,\boldsymbol{\theta})\sqrt{2\pi}}e^{-\frac{1}{2}(y_i-g(t_i,\boldsymbol{\theta})/\eta g(t_i,\boldsymbol{\theta}))^2} \tag{25}$$

$$= \frac{1}{(\eta^2 2\pi)^{n/2}}e^{-\frac{1}{2\eta^2}\sum_{i=1}^{n}(y_i-g(t_i,\boldsymbol{\theta})/g(t_i,\boldsymbol{\theta}))^2}\prod_{i=1}^{n}\frac{1}{g(t_i,\boldsymbol{\theta})}. \tag{26}$$

The negative-log likelihood is

$$-2\log(\boldsymbol{\theta}|\mathbf{y}) = n\log\eta^2 + n\log 2\pi + \frac{1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i-g(t_i,\boldsymbol{\theta})}{g(t_i,\boldsymbol{\theta})}\right)^2 + 2\sum_{i=1}^{n}\log g(t_i,\boldsymbol{\theta}) \tag{27}$$

or

$$-2\log(\boldsymbol{\theta}|\mathbf{y}) = \text{const} + \frac{1}{\eta^2}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\frac{\mathrm{y_i}-\mathrm{g(t_i,\boldsymbol{\theta})}}{\mathrm{g(t_i,\boldsymbol{\theta})}}\right)^2 + 2\sum_{\mathrm{i}=1}^{\mathrm{n}}\log\mathrm{g(t_i,\boldsymbol{\theta})}. \tag{28}$$

Thus, the maximum likelihood estimate is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\frac{1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i-g(t_i,\boldsymbol{\theta})}{g(t_i,\boldsymbol{\theta})}\right)^2 + 2\sum_{i=1}^{n}\log g(t_i,\boldsymbol{\theta}). \tag{29}$$

Likelihood-based confidence intervals can be constructed for each parameter using the negative log-likelihood function and a threshold Δ_α. The following definitions for confidence intervals, profile likelihood, and identifiability is attributed to Raue et. al.'s *Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood* [Raue et al., 2009].

Definition 4.9. The likelihood-based confidence interval is

$$CI = \{\boldsymbol{\theta} | \chi^2(\boldsymbol{\theta}) - \chi^2(\hat{\boldsymbol{\theta}}) < \Delta_\alpha\} \tag{30}$$

where χ^2 is the negative-log likelihood, $\hat{\boldsymbol{\theta}}$ is the fitted parameters, and Δ_α is the α-quantile of the Chi-squared distribution χ^2 of degree of freedom df. We choose $df = 1$ for pointwise confidence interval and df = number of unknown parameters for simultaneous confidence intervals.

4.2.2 Profile Likelihood

The idea behind profile likelihood is that we explore a one-dimensional space of the negative log-likelihood function by keeping one parameter fixed while optimizing the rest of the unknown parameters.

Definition 4.10. Profile likelihood is defined as

$$\chi^2_{\text{PL}}(\theta_i) = \min_{\theta_{j \neq i}} \chi^2(\boldsymbol{\theta}) \tag{31}$$

where χ^2 is the negative log-likelihood.

The likelihood-based confidence region is determined by the intercepts of the χ^2_{PL} and the threshold.

Definition 4.11. If the likelihood-based confidence region is infinitely extending in one or both directions, then the parameter estimate $\hat{\theta}_i$ is practically non-identifiable.

If a parameter estimate $\hat{\theta}_i$ is practically non-identifiable, then any other parameter within the neighborhood of $\hat{\theta}_i$ is as likely as $\hat{\theta}_i$ to generate the observed data.

Definition 4.12. If a parameter θ_i has a finite confidence interval *and* a unique minimum, then the parameter θ_i is practically identifiable.

If an unique minimum does not exists, then there is more than one parameter values that is equally as likely to generate the observed data. Then, the parameter θ_i is *locally non-identifiable*.

5 Results

5.1 Structural Identifiability

We computed structural identifiability, and recorded the results per parameter according to the key below. The outputs of each model is depicted on the y axis, while the inputs are depicted on the x axis. Initial conditions are depicted with $I_h(0)$, $I_v(0)$, $S_h(0)$, $S_v(0)$. Fixed parameter values are represented in the chart by the value itself. Locally identifiable parameters are represented in blue, while globally identifiable parameters are represented in purple, non identifiable parameters are represented in red, and finally, errors are represented in orange.

Key	
Black Text	Fixed parameter
	Not applicable
	Locally Identifiable
	Globally identifiable
	Non-identifiable
	Error

5.1.1 DAISY

The structural identifiability results from DAISY are depicted in the chart below. The DAISY codes used to generate these results are available in the appendix. Because DAISY took approximately 3 to 7 days to reach error conclusions, we stopped using DAISY for structural identifiability results.

DAISY All Results													
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)
Ih, Iv													
Ih													
Iv													
Ch, Cv													
Ch													
Cv													

5.1.2 SIAN

The structural identifiability results from SIAN are depicted in the chart below. The SIAN codes used to generate these results are available in the appendix. The first model we tested the structural identifiability of is depicted below.

$$
\begin{aligned}
S_h' &= \Pi_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= \Pi_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v
\end{aligned}
\tag{32}
$$

Below are the structural identifiability results of the above model. SIAN gave Kernel Errors for the outputs of I_h and I_v.

SIAN No Fixed Parameter Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ih, Iv														.922 s
Ih														
Iv														

This is the model used to generate cumulative cases with no fixed parameters.

$$
\begin{aligned}
S_h' &= \Pi_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= \Pi_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v \\
C_h' &= \frac{\beta_h S_h I_v}{N_v} \\
C_v' &= \frac{\beta_v S_v I_h}{N_h}
\end{aligned}
\tag{33}
$$

Below are the structural identifiability results of the above model. SIAN gave Kernel Errors for the outputs of C_h and C_v.

SIAN Cumulative Case No Fixed Parameter Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ch, Cv														269.092 s
Ch														
Cv														

Here is the fixed demographic model. By setting the birth rates equal to the death rates, we keep the population constant.

$$\begin{aligned} S_h' &= \mu_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= \mu_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v \end{aligned} \tag{34}$$

Below are the structural identifiability results of the above model. SIAN gave Kernel Errors for the outputs of I_h and I_v.

SIAN Fixed Demographic Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ih, Iv			μ_h	μ_v										1.440 s
Ih			μ_h	μ_v										
Iv			μ_h	μ_v										

Here is the cumulative case model for fixed demographics.

$$\begin{aligned} S_h' &= \mu_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= \mu_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v \\ C_h' &= \frac{\beta_h S_h I_v}{N_v} \\ C_v' &= \frac{\beta_v S_v I_h}{N_h} \end{aligned} \tag{35}$$

Below are the structural identifiability results of the above model. SIAN gave Kernel Errors for the outputs of C_h and C_v.

SIAN Cumulative Cases Fixed Demographic Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ch, Cv			μ_h	μ_v										31.727 s
Ch			μ_h	μ_v										
Cv			μ_h	μ_v										

Here is the model with fixed μ_h and μ_v values.

$$\begin{aligned} S_h' &= \Pi_h - \frac{1}{25000}S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= \Pi_v - \frac{9}{100}S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma)I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100}I_v \end{aligned} \tag{36}$$

Below are the structural identifiability results of the above model. It is of interest to note that these parameter results differ from the outputs given by the exact arithmetic rank software. The EAR software ran the same model with I_v as the output, and found Π_v and $S_v(0)$ to be locally structurally identifiable rather than unidentifiable, and β_v being non-identifiable. This also happened with the I_h output, with β_v being locally structurally identifiable according to the EAR software.

SIAN Fixed μ_h, μ_v														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ih, Iv	$\frac{1}{25000}$	$\frac{9}{100}$												11.7 s
Ih	$\frac{1}{25000}$	$\frac{9}{100}$												27.835 s
Iv	$\frac{1}{25000}$	$\frac{9}{100}$												27.720 s

Here is the model used to generate cumulative case results with fixed μ_h and μ_v values.

$$\begin{aligned} S_h' &= \Pi_h - \frac{1}{25000}S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= \Pi_v - \frac{9}{100}S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma)I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100}I_v \\ C_h' &= \frac{\beta_h S_h I_v}{N_v} \\ C_v' &= \frac{\beta_v S_v I_h}{N_h} \end{aligned} \tag{37}$$

Below are the structural identifiability results of the above model.

SIAN Cumulative Case Fixed μ														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ch, Cv	1/25000	9/100												3.454 s
Ch	1/25000	9/100												
Cv	1/25000	9/100												

Here are the equations with 3 fixed parameters: μ_h, μ_v, and γ.

$$\begin{aligned} S_h' &= \Pi_h - \frac{1}{72.6 \cdot 365}S_h - \frac{\beta_h S_h I_v}{N_v} + \frac{1}{10}I_h \\ S_v' &= \Pi_v - \frac{1}{46}S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{72.6 \cdot 365} + \frac{1}{10})I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{1}{46}I_v \end{aligned} \tag{38}$$

Below are the structural identifiability results of the above model. It is of interest to note that these parameter results differ from the outputs given by the exact arithmetic rank software. The EAR software ran the same model with I_v as the output, and found Π_v and $S_v(0)$ to be locally structurally identifiable rather than unidentifiable, and β_v being non-identifiable. This also happened with the I_h output, with β_v being locally structurally identifiable according to the EAR software.

SIAN Fixed μ_h, μ_v, γ														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ih, Iv	$1/(72.6 \cdot 365)$	1/46			0.1									1.297 s
Ih	$1/(72.6 \cdot 365)$	1/46			0.1									1.372 s
Iv	$1/(72.6 \cdot 365)$	1/46			0.1									1.291 s

Here is the cumulative case system of equations with 3 fixed parameters: μ_h, μ_v, and γ.

$$\begin{aligned}
S_h' &= \Pi_h - \frac{1}{72.6\cdots 365}S_h - \frac{\beta_h S_h I_v}{N_v} + \frac{1}{10}I_h \\
S_v' &= \Pi_v - \frac{1}{46}S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{72.6\cdots 365} + \frac{1}{10})I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{1}{46}I_v \\
C_h' &= \frac{\beta_h S_h I_v}{N_v} \\
C_v' &= \frac{\beta_v S_v I_h}{N_h}
\end{aligned} \tag{39}$$

Below are the structural identifiability results of the above model. It is of interest to note that these parameter results differ from the outputs given by the exact arithmetic rank software. The EAR software ran the same model with C_v as the output, and found Π_v and $S_v(0)$ to be locally structurally identifiable rather than unidentifiable, and β_v being non-identifiable. This also happened with the C_h output, with β_v being locally structurally identifiable according to the EAR software.

SIAN Cumulative Case Fixed μ_h, μ_v, γ														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ch, Cv	$\frac{1}{72.6\cdot 365}$	$\frac{1}{46}$			$\frac{1}{10}$									1.677 s
Ch	$\frac{1}{72.6\cdot 365}$	$\frac{1}{46}$			$\frac{1}{10}$									65.985 s
Cv	$\frac{1}{72.6\cdot 365}$	$\frac{1}{46}$			$\frac{1}{10}$									57.212 s

Below is the model with fixed values for μ_h, μ_v, Π_h, and Π_v.

$$\begin{aligned}
S_h' &= \frac{100}{25000} - \frac{1}{25000}S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= 90 - \frac{9}{100}S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma)I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100}I_v
\end{aligned} \tag{40}$$

Below are the structural identifiability results of the above model, which agree with the EAR results.

SIAN Fixed $\mu_h, \mu_v, \Pi_h, \Pi_v$														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ih, Iv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										.683 s
Ih	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										1.677 s
Iv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										1.001 s

Below is the cumulative case model with fixed values for μ_h, μ_v, Π_h, and Π_v.

$$\begin{aligned}
S_h' &= \frac{100}{25000} - \frac{1}{25000}S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= 90 - \frac{9}{100}S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma)I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100}I_v, \\
C_h' &= \frac{\beta_h S_h I_v}{N_v}, \\
C_v' &= \frac{\beta_v S_v I_h}{N_h}
\end{aligned} \tag{41}$$

Below are the structural identifiability results of the above model, which agree with the EAR results.

SIAN Cumulative Case Fixed $\mu_h, \mu_v, \Pi_h, \Pi_v$														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ch, Cv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										1.206 s
Ch	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										557.188 s
Cv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										15.245

Below is the model with fixed values for Π_h, and Π_v.

$$\begin{aligned}
S_h' &= \frac{100}{25000} - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= 90 - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma)I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v
\end{aligned} \tag{42}$$

Below are the structural identifiability results of the above model, which agree with the EAR results.

SIAN Fixed Π_h, Π_v														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ih, Iv			100/25000	90										1.322 s
Ih			100/25000	90										
Iv			100/25000	90										3343.054

Below is the cumulative case model for fixed values of Π_h and Π_v.

$$\begin{aligned} S_h' &= \frac{100}{25000} - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= 90 - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v, \\ C_h' &= \frac{\beta_h S_h I_v}{N_v}, \\ C_h' &= \frac{\beta_v S_v I_h}{N_h} \end{aligned} \tag{43}$$

Below are the structural identifiability results of the above model, which agree with the EAR results.

SIAN Cumulative Case Fixed Π_h, Π_v														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ch, Cv			100/25000	90										11.618
Ch			100/25000	90										
Cv			100/25000	90										

5.1.3 EAR

The structural identifiability results from the Wolfram Alpha IdentifiabilityAnalysis package are depicted in the charts below. These results treat initial conditions as variables. We also ran structural identifiability for all of the models below in EAR for fixed initial conditions, and found all of the parameters for all of the models to be structurally identifiable. and found The EAR codes used to generate these results are available in the appendix. The first model we tested the structural identifiability of in Wolfram Alpha is depicted below.

$$
\begin{aligned}
S_h' &= \Pi_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= \Pi_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v
\end{aligned}
\tag{44}
$$

Below are the structural identifiability results of the above model. No errors were given for the outputs of I_h and I_v.

EAR No Fixed Parameter Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ih, Iv														0.640625 s
Ih														0.5625 s
Iv														0.546875 s

This is the model used to generate cumulative cases with no fixed parameters.

$$
\begin{aligned}
S_h' &= \Pi_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= \Pi_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v \\
C_h' &= \frac{\beta_h S_h I_v}{N_v} \\
C_v' &= \frac{\beta_v S_v I_h}{N_h}
\end{aligned}
\tag{45}
$$

Below are the structural identifiability results of the above model. This method gave no errors for the outputs of C_h and C_v.

EAR Cumulative Case No Fixed Parameter Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ch, Cv														0.8125 s
Ch														1.34375 s
Cv														1.45313 s

Here is the fixed demographic model. By setting the birth rates equal to the death rates, we keep the population constant.

$$\begin{aligned} S_h' &= \mu_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= \mu_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v \end{aligned} \tag{46}$$

Below are the structural identifiability results of the above model. EAR gave no errors for the outputs of I_h and I_v.

EAR Fixed Demographic Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ih, Iv			μ_h	μ_v										0.421875 s
Ih			μ_h	μ_v										0.421875 s
Iv			μ_h	μ_v										0.453125 s

Here is the cumulative case model for fixed demographics.

$$\begin{aligned} S_h' &= \mu_h - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= \mu_v - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v \\ C_h' &= \frac{\beta_h S_h I_v}{N_v} \\ C_v' &= \frac{\beta_v S_v I_h}{N_h} \end{aligned} \tag{47}$$

Below are the structural identifiability results of the above model. EAR gave no errors for the outputs of C_h and C_v.

EAR Fixed Cumulative Case Demographic Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ch, Cv			μ_h	μ_v										.75 s
Ch			μ_h	μ_v										.5625 s
Cv			μ_h	μ_v										0.53125 s

Here is the model with fixed μ_h and μ_v values.

$$\begin{aligned} S_h' &= \Pi_h - \frac{1}{25000}S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= \Pi_v - \frac{9}{100}S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma)I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100}I_v \end{aligned} \tag{48}$$

Below are the structural identifiability results of the above model. It is of interest to note that these parameter results differ from the outputs given by the SIAN software. The EAR software ran the model with I_v as the output, and found Π_v and $S_v(0)$ to be locally structurally identifiable rather than unidentifiable, and β_v being non-identifiable. This also happened with the I_h output, with β_v being locally structurally identifiable according to the EAR software.

EAR Fixed μ_h, μ_v Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ih, Iv	1/25000	9/100												0.640625 s
Ih	1/25000	9/100												1.125 s
Iv	1/25000	9/100												0.578125 s

Here is the model used to generate cumulative case results with fixed μ_h and μ_v values.

$$
\begin{aligned}
S_h' &= \Pi_h - \frac{1}{25000} S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= \Pi_v - \frac{9}{100} S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma) I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100} I_v \\
C_h' &= \frac{\beta_h S_h I_v}{N_v} \\
C_v' &= \frac{\beta_v S_v I_h}{N_h}
\end{aligned}
\tag{49}
$$

Below are the structural identifiability results of the above model.

EAR Cumulative Fixed μ_h, μ_v Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ch, Cv	1/25000	9/100												.8125 s
Ch	1/25000	9/100												0.5625 s
Cv	1/25000	9/100												.625 s

Here are the equations with 3 fixed parameters: μ_h, μ_v, and γ.

$$
\begin{aligned}
S_h' &= \Pi_h - \frac{1}{72.6 \cdot 365} S_h - \frac{\beta_h S_h I_v}{N_v} + \frac{1}{10} I_h \\
S_v' &= \Pi_v - \frac{1}{46} S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - \left(\frac{1}{72.6 \cdot 365} + \frac{1}{10} \right) I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{1}{46} I_v
\end{aligned}
\tag{50}
$$

Below are the structural identifiability results of the above model. It is of interest to note that these parameter results differ from the outputs given by the SIAN software. The EAR software ran the model with I_v as the output, and found Π_v and $S_v(0)$ to be locally structurally identifiable rather than unidentifiable, and β_v being non-identifiable. This also happened with the I_h output, with β_v being locally structurally identifiable according to the EAR software.

EAR Fixed μ_v, μ_h, γ Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ih, Iv	$1/(72.6 \cdot 365)$	$1/46$			0.1									0.20312 s5
Ih	$1/(72.6{\cdot}365)$	$1/46$			0.1									0.828125 s
Iv	$1/(72.6{\cdot}365)$	$1/46$			0.1									0.296875 s

Here is the cumulative case system of equations with 3 fixed parameters: μ_h, μ_v, and γ.

$$
\begin{aligned}
S_h' &= \Pi_h - \frac{1}{72.6 \cdot 365} S_h - \frac{\beta_h S_h I_v}{N_v} + \frac{1}{10} I_h \\
S_v' &= \Pi_v - \frac{1}{46} S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{72.6 \cdot 365} + \frac{1}{10}) I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{1}{46} I_v \\
C_h' &= \frac{\beta_h S_h I_v}{N_v} \\
C_v' &= \frac{\beta_v S_v I_h}{N_h}
\end{aligned}
\tag{51}
$$

Below are the structural identifiability results of the above model. It is of interest to note that these parameter results differ from the outputs given by the SIAN. The EAR software ran the model with C_v as the output, and found Π_v and $S_v(0)$ to be locally structurally identifiable rather than unidentifiable, and β_v being non-identifiable. This also happened with the C_h output, with β_v being locally structurally identifiable according to the EAR software.

EAR Cumulative Case Fixed μ_v, μ_h, γ														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ch, Cv	$1/(72.6{\cdot}365)$	$1/46$			0.1									0.84375 s
Ch	$1/(72.6{\cdot}365)$	$1/46$			0.1									0.734375 s
Cv	$1/(72.6{\cdot}365)$	$1/46$			0.1									0.546875 s

Below is the model with fixed values for μ_h, μ_v, Π_h, and Π_v.

$$
\begin{aligned}
S_h' &= \frac{100}{25000} - \frac{1}{25000}S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= 90 - \frac{9}{100}S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma)I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100}I_v
\end{aligned}
\tag{52}
$$

Below are the structural identifiability results of the above model, which agree with the SIAN results.

EAR Fixed μ_h, μ_v, Π_h and Π_v Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ih, Iv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										0.21875 s
Ih	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										0.328125 s
Iv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										0.359375 s

Below is the cumulative case model with fixed values for μ_h, μ_v, Π_h, and Π_v.

$$
\begin{aligned}
S_h' &= \frac{100}{25000} - \frac{1}{25000}S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\
S_v' &= 90 - \frac{9}{100}S_v - \frac{\beta_v S_v I_h}{N_h}, \\
I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\frac{1}{25000} + \gamma)I_h, \\
I_v' &= \frac{\beta_v S_v I_h}{N_h} - \frac{9}{100}I_v, \\
C_h' &= \frac{\beta_h S_h I_v}{N_v}, \\
C_v' &= \frac{\beta_v S_v I_h}{N_h}
\end{aligned}
\tag{53}
$$

Below are the structural identifiability results of the above model, which agree with the SIAN results.

EAR Cumulative Cases Fixed μ_h, μ_v, Π_h and Π_v Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ch, Cv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										0.90625 s
Ch	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										0.90625 s
Cv	$\frac{1}{25000}$	$\frac{9}{100}$	$\frac{100}{25000}$	90										.75 s

Below is the model with fixed values for Π_h, and Π_v.

$$\begin{aligned} S_h' &= \frac{100}{25000} - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= 90 - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v \end{aligned} \tag{54}$$

Below are the structural identifiability results of the above model, which agree with the SIAN results.

EAR Fixed Π_h and Π_v Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	Ch(0)	Cv(0)	Timing
Ih, Iv			$\frac{100}{25000}$	90										0.640625 s
Ih			$\frac{100}{25000}$	90										0.421875 s
Iv			$\frac{100}{25000}$	90										0.4375 s

Below is the cumulative case model for fixed values of Π_h and Π_v.

$$\begin{aligned} S_h' &= \frac{100}{25000} - \mu_h S_h - \frac{\beta_h S_h I_v}{N_v} + \gamma I_h \\ S_v' &= 90 - \mu_v S_v - \frac{\beta_v S_v I_h}{N_h}, \\ I_h' &= \frac{\beta_h S_h I_v}{N_v} - (\mu_h + \gamma) I_h, \\ I_v' &= \frac{\beta_v S_v I_h}{N_h} - \mu_v I_v, \\ C_h' &= \frac{\beta_h S_h I_v}{N_v}, \\ C_h' &= \frac{\beta_v S_v I_h}{N_h} \end{aligned} \tag{55}$$

Below are the structural identifiability results of the above model, which agree with the SIAN results.

EAR Cumulative Case Π_h and Π_v Results														
	μ_h	μ_v	Π_h	Π_v	γ	β_h	β_v	$I_h(0)$	$I_v(0)$	$S_h(0)$	$S_v(0)$	$C_h(0)$	$C_v(0)$	Timing
Ch, Cv			$\frac{100}{25000}$	90										0.9375 s
Ch			$\frac{100}{25000}$	90										0.546875 s
Cv			$\frac{100}{25000}$	90										0.421875 s

5.2 Practical Identifiability

5.2.1 Monte Carlo Simulation

We ran the Monte Carlo simulation method for the vector-host model with parameters set to $\beta_h = 0.1, \beta_v = 0.2, \gamma = 0.09996, \mu_h = 0.00004, \mu_v = 0.1, \Pi_h = 0.00004, \Pi_v = 0.1$. We chose μ_h and μ_v to fit the range of lifespans of humans and mosquitoes, and Π_h and Π_v to set the equilibrium population sizes to 1. While the disease-specific parameters were not chosen to model any specific disease, they produce a reasonable R_0 value of 1.41, and results in the endemic equilibrium of $I_h = 0.25$, $I_v = 0.33$.

When we set the initial guess to the correct parameters, we see in 1 that except for μ_h and Π_h, the parameters are practically identifiable. It seems likely that the unidentifiability

σ	β_h	β_v	γ	μ_h	μ_v	Π_h	Π_v
0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1%	0.08	0.05	0.09	2.99	0.08	3.22	0.14
5%	0.83	0.65	0.90	19.42	0.78	18.57	1.23
10%	2.30	1.98	2.41	30.10	1.96	29.95	2.90
20%	5.14	4.55	5.54	42.11	4.02	48.59	6.05
30%	8.30	8.38	9.24	68.22	6.70	62.09	9.68

Table 1: Practical Identifiability for Vector-Host Model

of these parameters is due to their small value, as they are chosen to fit the lifespan of humans. The percentage of the host population that is born or dies over the 217 days we measure is is only a small portion of the percentage of the host population that becomes infected.

But this identifiability is fragile to the choice of initial guess given to the optimization algorithm. When we had the initial guess of $\beta_h = 0.05, \beta_v = 0.1, \gamma = 0.05$, with the correct guess for the demographic parameters $\mu_h, \Pi_h, \mu_v, \Pi_v$, we obtained the results in 2. We see that none of the parameters meet the threshold for practical identifiability.

Next, we attempted to increase identifiability by fixing Π_h to μ_h and Π_v to μ_v. This fixes the population size to 1, but allows the death rate to change freely. When we have the correct initial guess, we obtain the results in 3. Once again, we find that all parameters are practically identifiable, with the exception of the host demographic parameter μ_h. But with the incorrect initial guess, we see in 4 that our parameters are not identifiable.

σ	β_h	β_v	γ	μ_h	μ_v	Π_h	Π_v
0%	2.16	12.94	8.92	98.40	33.21	87.23	48.90
1%	33.50	19.75	24.84	87.03	28.30	65.86	45.31
5%	34.58	14.97	21.55	66.02	35.37	69.85	52.23
10%	33.40	16.13	21.43	64.14	39.60	83.16	55.04
20%	35.08	16.87	23.11	59.10	41.43	76.71	56.28
30%	35.43	25.84	23.32	58.26	54.39	87.54	71.94

Table 2: Vector-Host Model with Incorrect Initial Guess

σ	β_h	β_v	γ	μ_h	μ_v
0%	0.00	0.00	0.00	0.00	0.00
1%	0.21	0.13	0.23	3.90	0.16
5%	1.49	1.13	1.65	14.43	1.41
10%	4.13	3.51	4.57	34.33	4.25
20%	9.43	8.45	10.68	55.74	10.14
30%	14.30	12.67	16.22	96.86	15.10

Table 3: Fixed Population Sizes with Correct Initial Guess

σ	β_h	β_v	γ	μ_h	μ_v
0%	39.18	18.77	41.63	100.00	14.90
1%	33.40	29.64	36.04	99.12	28.78
5%	30.33	26.34	33.41	97.94	25.70
10%	29.92	26.77	32.84	109.24	26.61
20%	30.82	29.24	33.55	98.99	29.42
30%	31.24	28.94	34.01	98.78	29.26

Table 4: Fixed Population Sizes with Incorrect Initial Guess

Next, we fixed all host demographic parameters, with β_h, β_v, γ, and μ_v free. Our results are shown in 5 and 6 with little improvement from the previous scenario.

Finally, we fixed all demographic parameters, leaving only the disease-specific parameters of β_h, β_v, and γ free. Here, as shown in 7 and 8, our parameters remain practically identifiable even with the incorrect initial guess.

5.2.2 Impact of Factors Independent of Model Structure

For this section, we look at the impact of factors independent of the model structure. We used the Monte Carlo simulation method on both the fixed demographic version of the

σ	β_h	β_v	γ	μ_v
0%	0.00	0.00	0.00	0.00
1%	0.21	0.11	0.25	0.18
5%	1.53	1.23	1.76	1.57
10%	4.72	4.18	5.25	5.01
20%	10.37	9.76	11.69	11.42
30%	17.25	14.63	19.47	17.23

Table 5: Fixed Host Demographics, Vector Population with Correct Initial Guess

σ	β_h	β_v	γ	μ_v
0%	0.02	0.02	0.02	0.02
1%	900.27	4.74	871.20	4.87
5%	601.01	98.16	573.87	92.82
10%	669.73	17.41	650.05	18.30
20%	354.24	31.84	340.94	33.46
30%	254.14	31.84	245.48	33.88

Table 6: Fixed Host Demographics, Vector Population with Incorrect Initial Guess

σ	β_h	β_v	γ
0%	0.00	0.00	0.00
1%	0.31	0.12	0.30
5%	1.57	0.59	1.55
10%	3.16	1.18	3.10
20%	6.40	2.44	6.30
30%	9.53	3.51	9.24

Table 7: Fixed Demographic with Correct Initial Guess

σ	β_h	β_v	γ
0%	0.02	0.00	0.02
1%	0.31	0.12	0.30
5%	1.58	0.59	1.56
10%	3.35	1.18	3.33
20%	6.51	2.47	6.43
30%	9.66	3.54	9.39

Table 8: Fixed Demographic with Incorrect Initial Guess

vector-host model, as well as the SIR model

$$\begin{aligned} S' &= -\frac{\beta SI}{N} \\ I' &= \frac{\beta SI}{N} - \alpha I \\ R' &= \alpha I \end{aligned} \tag{56}$$

with initial conditions $S(0) = S_0, I(0) = I_0, R(0) = R_0$. We have $N = S + I + R$, which gives us $\frac{dN}{dt} = 0$.

The true parameters are $\alpha = 0.25, \beta = 0.5$. We use 1000 iterations for the Monte Carlo method, with prevalence observations daily over 51 days. The initial guess given to `fminsearch` is $\alpha = 0.1, \beta = 0.1$, and we use relative noise with ordinary least squares. For the vector-host model, the setup is the one used for 7.

We first investigated the changes in identifiability when the number of observations is decreased.

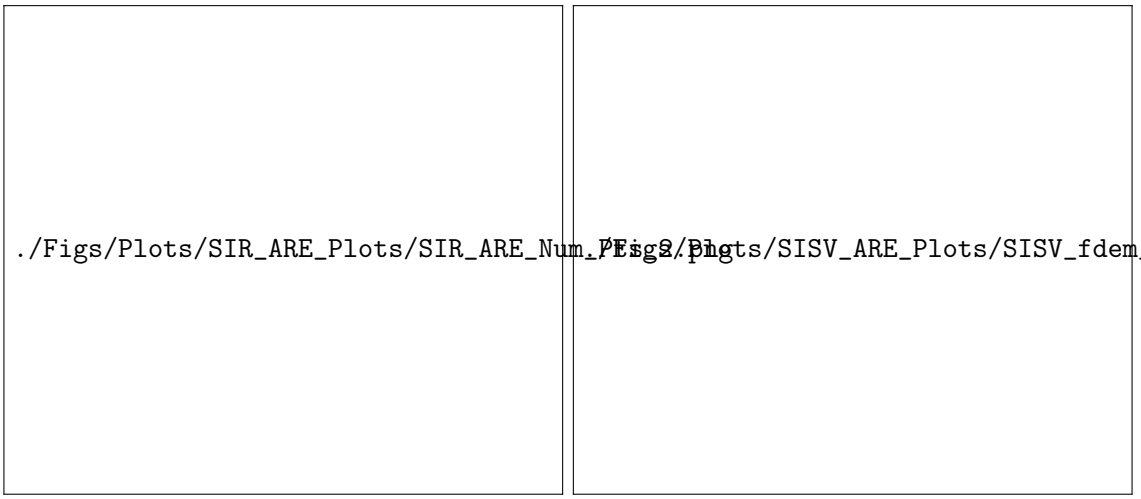

Figure 1: Changing the Number of Observations

We see in 1 that with fewer observations, we have less information about parameters, and our parameters are less identifiable. However, few observations are need for the parameters to be under the threshold for practical identifiability. We only need 21 observations over our 216 days for all of the vector-host model parameters to be identifiable. And the ARE for both parameters of the SIR model was underneath the threshold with only 3 observations.

Next, we investigated changing the final date of observations. In order to avoid the effect shown in 1, no matter the final date, we have the same number of linearly spaced observations. For the vector-host model, we set the initial guess to half of the true parameters. As seen in 2, the identifiability generally decreases as the time span decreases.

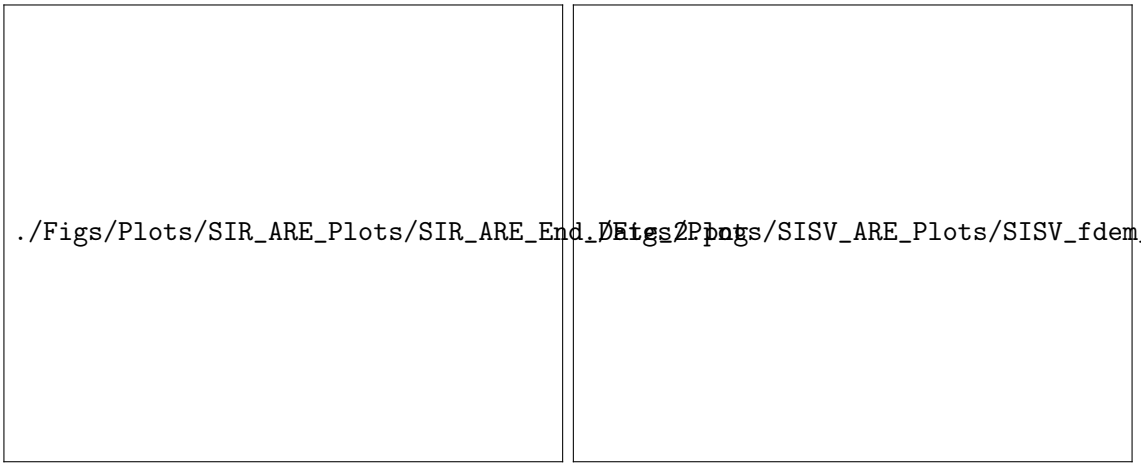

Figure 2: Changing The Final Date of Observations

Notably, before and during the rise in cases, the parameters are not practically identifiable.

For the Vector-Host Model, we investigated identifiability when only hosts, or only vectors, were observable.

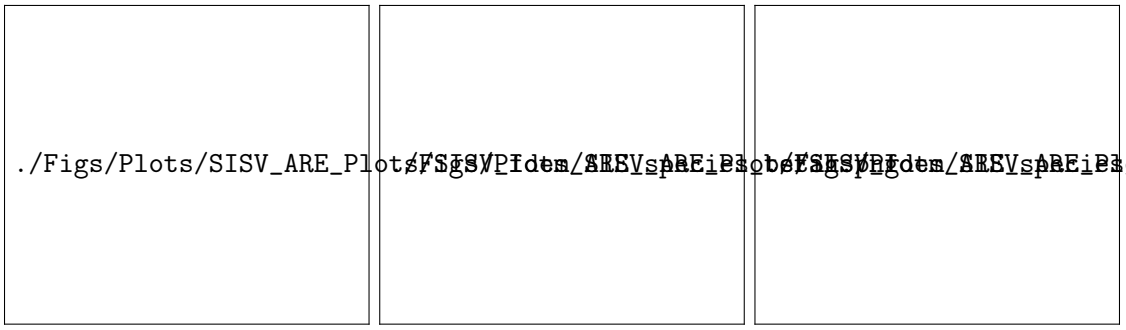

Figure 3: Changing Which Populations are Observable

As shown in 3, the ARE values were significantly less when information on both populations was known. However, the parameters were still identifiable from prevalence data of only one population.

We looked at the identifiability of these models from different types of data. We have been using prevalence data, but we can also fit using incidence or cumulative data. And our cumulative data can be treated as the sum of noisy incidence data, or the noisy data itself. As can be seen from comparing 9 and 10, we have less variance in our error when we assume that the error occurs in incidence.

This improvement is visible in our numerical experiments, shown in 4 and 5. For all parameters, when our error is in incidence data, the ARE is less than when it is in

cumulative data. Thus, when considering the practical identifiability of cumulative data, it is important to consider if error actually occurs in cumulative data, or if it occurs in incidence data.

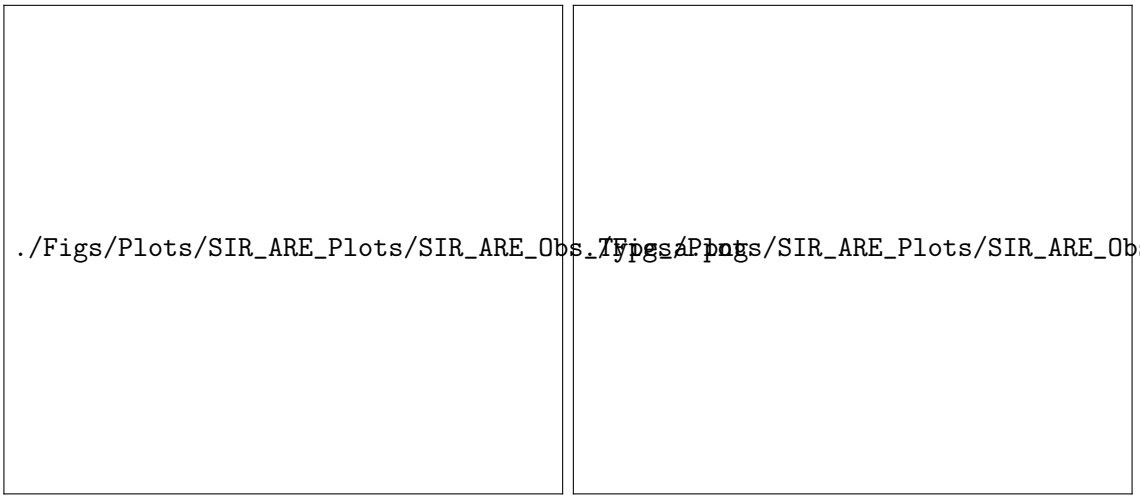

Figure 4: Observation Types for the SIR Model

Figure 5: Observation Types for the Vector-Host Model

We also tested different observation types for when we had data for only one population. The results are shown in 6, and match what our above observations for the error in cumulative data.

Between 4,5, and 6, we see that neither prevalence nor incidence consistently outperforms the other. Prevalence has lower ARE values for the SIR model, and for β_v and γ in the vector-host model. But incidence outperforms prevalence when we only have one species of observations. And cumulative data with incidence error performs similar to, if slightly worse, than incidence data.

Figure 6: Changing Observation Type for One Population

Next, we investigated changes in our error model and objective function. We tested setting $\xi = 0$(absolute noise) and $\xi = 1$(relative noise) in 6, as well as $\xi = 0$ and $\xi = 1$(Ordinary Least Squares and Generalized Least Squares) in 7. As shown in 7, for relative noise, the ARE is less when using GLS than OLS, but OLS still performs reasonably well. The same cannot be said for GLS with absolute error. In this case, OLS performs well, but our parameters are not practically identifiable when using GLS.

Now for our Vector-Host model, when considering the case where $\xi = 0$ for either 6 or 7, we are making the assumption that error is independent of the magnitude of our observations. But we leave open the dependence of the error on the size of the population.

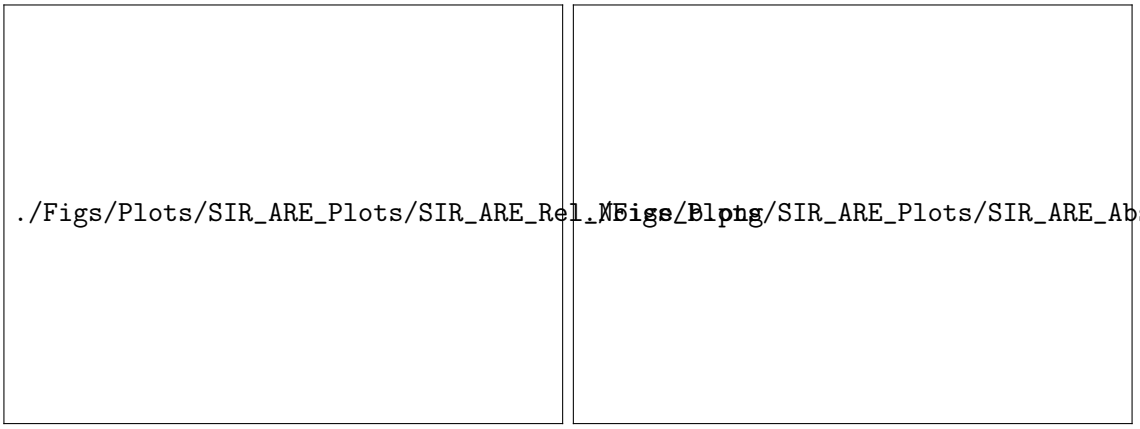

Figure 7: Changing the Error Model and Objective Function for the SIR Model

We assume that this is the case for absolute noise, which we weight to form

$$y_i = \mathbf{g}(t_i, \boldsymbol{\theta}) + \frac{N}{10}\epsilon_i$$

We can also use this weighting for ordinary least squares. We ran this for the vector-host model, with a vector population of $100,000$ and a host population of 100.

Our results can be seen in 8. We see that adjusting for population decreases the error in parameter estimates for absolute (population size dependant) noise, as expected. But the weighting also improves the estimates for relative noise, although GLS still outperforms it. With absolute noise, we once again see the using GLS as our optimization algorithm leads to our parameters being unidentifiable. So we find that a population adjusted version of ordinary least squares is the most versatile choice of objective function.

We next focus on the optimization method used for our parameter estimations. When we ran the Nelder-Mead[Lagarias et al., 1998],interior point[Byrd et al., 1996], and quasi-newton[Broyden, 1970] algorithms, we did not see any change in the ARE values, as seen in 9.

The Interior Point optimization algorithm allows bounds to be set on its parameter values. Additionally, the function `fminsearchbnd` in MATLAB allows bounds to be set for the Nelder-Mead algorithm[D'Errico, 2022]. We investigated the impact of setting bound on ARE values for cumulative data. As seen in 10, our bounds must be very close to the true parameters in order to decrease the error in parameter estimations. This suggests that with a correct initial guess, the choice of optimization algorithm and bounds does not impact practical identifiability results.

But what if the algorithm does not have a correct initial guess? We have already seen in 1 and 2 that changing the initial guess can make the ARE measurements go from practically identifiable to practically unidentifiable.

Figure 8: Changing the Error Model and Objective Function for the Vector-Host Model

Figure 9: Changing the Optimization Algorithm

For multiple optimization algorithms, we made contour plots of the ARE value at $\sigma = 10\%$ for different initial guesses. These used 100 iterations of the Monte Carlo simulation method, and were run for the Nelder-Mead[Lagarias et al., 1998], interior point[Byrd et al., 1996], sqp[Nocedal, 2006], active set[Gill, 1981], and quasi-newton[Broyden, 1970] algorithms. These plot the ARE of β_h when changing the initial guess of β_h and β_v, and can be seen in 11. We saw similar contour plots for the ARE value of γ. But our ARE values for β_v were much lower, as can be seen in 12.

Another interesting pattern is that for the algorithms besides Nelder-Mead, there is a line of unidentifiability along $\beta_h = \beta_v$.

Adding bounds to an optimization algorithm can also change the impact of the initial

Figure 10: Changing Bounds for the Optimization Algorithm

Figure 11: ARE of β_h when Changing Initial Guess

guess. We investigated this impact for the Interior Point algorithm, as well as for the MATLAB `fminsearchbnd`[D'Errico, 2022] implementation of Nelder-Mead. With 13, we see that with moderately precise bounds (10x the true parameters), nearly all initial guesses are identifiable. However, bounds of 100x the true parameters produce an entirely different effect. For the interior point algorithm, this may be due to the true parameters being much closer to the lower bound of 0 than the upper bound of 100x their true value. In

Figure 12: ARE of β_v when Changing Initial Guess

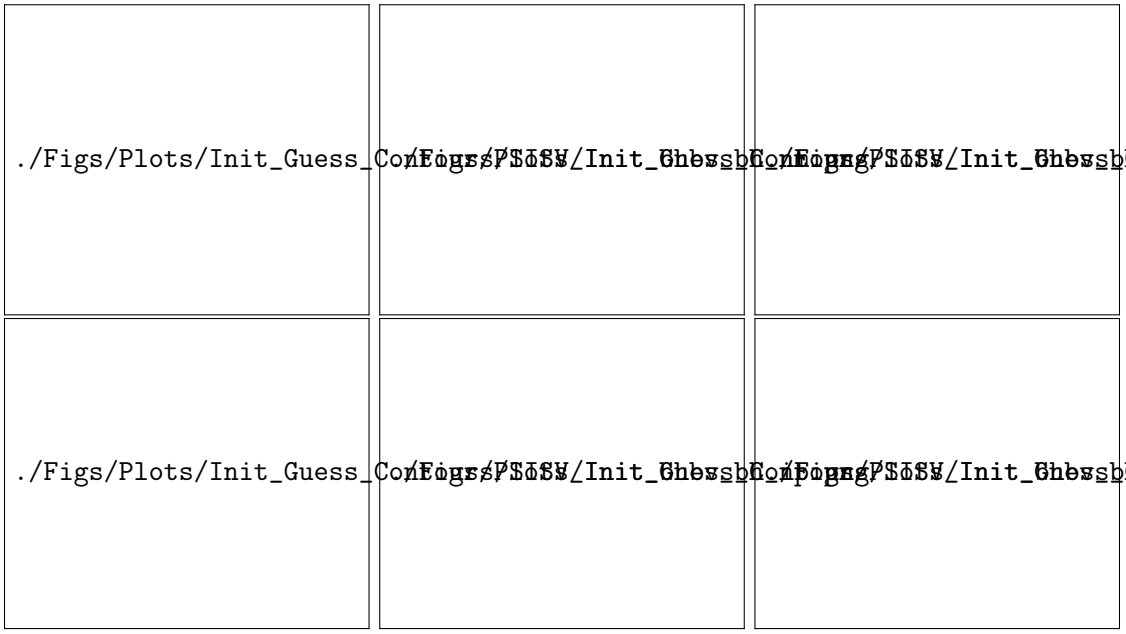

Figure 13: From Left to Right: No Bounds, 100x True Parameters, 10x True Parameters

conclusion, the choice of bounds or optimization algorithm does not have a large impact when we have a good initial guess. But with a poor initial guess, adding bounds or changing the optimization algorithm can have a large effect.

5.2.3 Profile Likelihood

Program implementation. The code to find profile likelihood is written in Julia due to its fast speed in solving differential equations which is important for parameter estimation [Bezanson et al., 2017].* We use the Tsitouras 5/4 Runge-Kutta package from the DifferentialEquations.jl package to solve the SIS Vector-Host model with a relative tolerance of 10^{-5} and an absolute tolerance of 10^{-10} [Rackauckas and Nie, 2017]. To find the maximum likelihood estimate, the global minimum must be found which is more difficult in simulation-based optimization [Kolda et al., 2003]. Information such as the continuity and differentiability of the objective function may or may not be well behaved. As such, we employ the generating set search from the BlackBoxOptim.jl package and differential evolution from the Metaheuristics.jl package [Feldt and Stukalov, 2018, Mejía et al., 2022]. The basic idea behind generating set search is that the method randomly choose a point to evaluate, search in the neighborhood of the point and expand the neighborhood till it find a better point, and then, repeat the process till a termination criteria is satisfied [Kolda et al., 2003]. Differential evolution have a population of points that evolve through mutation and crossover events to find the global minimum [Price, 2013]. Due to the difficulty of finding the global minimum, it is recommend to use two global optimization methods where one optimization method is used to find the profile likelihood and the other is used to check whether the global optimization was successfully done. For most plots, generating set search was used to create the plots and differential evolution was used to check whether the global minimum was found through generating set search. To find the confidence intervals, linear interpolation from the Interpolations.jl package was used to approximate χ^2_{PL} and the simple bisection method from the Roots.jl package was used to find the intercepts of χ^2_{PL} and the threshold. The code can be found at `https://github.com/ph-kev/ProfileLikelihood`.

Procedure A high-level description of the procedure for finding the profile likelihood is described below.

1. Gather data either experimentally or numerically.
2. Choose the probability distribution that best describes the noise of the data and find the negative log-likelihood function.
3. Find the fitted parameters by minimizing the negative log-likelihood function.
4. Compute the profile likelihood for each parameter:

$$\chi^2_{\text{PL}}(\theta_i) = \min_{\theta_{j\neq i}} \chi^2(\boldsymbol{\theta}).$$

*See appendix for work-precision benchmarks on solving differential equations.

5. Compute the threshold $\Delta_\alpha = \chi^2(\alpha, df)$.

6. Graph, find confidence intervals, and check identifiability.

Model and Parameters Recall the SIS vector-host model

$$\begin{aligned} I_h' &= \beta_h S_h I_v - (\mu_h + \gamma) I_h, \\ I_v' &= \beta_v S_v I_h - \mu_v I_v, \\ S_h' &= \Pi_h - \mu_h S_h - \beta_h S_h I_v + \gamma I_h, \\ S_v' &= \Pi_v - \mu_v S_v - \beta_v S_v I_h. \end{aligned}$$

The basic reproduction number R_0 for this model is

$$R_0 = \sqrt{\frac{\beta_h \beta_v S_h(0) S_v(0)}{(\mu_h + \gamma)\mu_v}}.$$

We fix the parameters $\Pi_h = 0.004$, $\Pi_v = 90$, $\mu_h = 0.00004$, and $\mu_v = 0.09$. These parameters are fixed because these parameters are related to the demographics as opposed to the disease itself. We will generate data numerically to use for profile likelihood. The true parameters are

$$\boldsymbol{\theta} = \begin{bmatrix} \beta_h & \beta_v & \gamma \end{bmatrix} = \begin{bmatrix} 0.0001 & 0.001 & 0.09 \end{bmatrix}$$

and the initial conditions are

$$u_0 = \begin{bmatrix} I_h & I_v & S_h & S_v \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1000 & 5000 \end{bmatrix}.$$

Additional data points of the same observation. We first consider the scenario where we add additional data points of the same observation and examine how that affect parameter estimation, confidence interval, and identifiability.

The observation will be incidence data of hosts and we will generate noisy incidence data of hosts using the Poisson distribution. Incidence data is taken at $t = 0, 1, \ldots, 46$. Since the data is generated from a Poisson distribution, the negative log-likelihood function is equation 23:

$$-2 \log L(\boldsymbol{\theta}|\mathbf{y}) = -2 \sum_{i=1}^{n} \log(g(t_i, \boldsymbol{\theta})^{y_i}) + 2 \sum_{i=1}^{n} g(t_i, \boldsymbol{\theta}) + 2 \sum_{i=1}^{n} \log(y_i!).$$

and the maximum likelihood estimate is equation 24:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -2 \sum_{i=1}^{n} \log(g(t_i, \boldsymbol{\theta})^{y_i}) + 2 \sum_{i=1}^{n} g(t_i, \boldsymbol{\theta}).$$

Figure 14: Perfect host incidence data and noisy host incidence data generated from the Poisson distribution.

Before $t \approx 15$, there is relatively very little noise compared to data at the peak and for $t > 15$. To better evaluate the profile likelihood plots, it is important to take into account of the data available as the fitted parameters might not accurately describe the incidence data perfectly for $t > 15$ due to the amount of the noise. Profile likelihood plots will be constructed for 16, 31, and 46 data points of noisy host incidence data and 95% confidence intervals will be used.

Figure 15: Profile likelihood of β_h, β_v, and γ for 16 data points of host incidence data.

For 16 data points of host incidence data, the fitted parameters are $\beta_h = 9.593 \cdot 10^{-5}$, $\beta_v = 0.00101$, and $\gamma = 0.0788$. From Figure 15, β_h, β_v, and γ are practically identifiable. Although there is no data about the peak of the host incidence data, the confidence interval of β_h is relatively smaller than the confidence interval of β_v and the confidence interval of γ. Although γ is practically identifiable by definition, the width of the confidence interval of γ is large which indicate uncertainty in the parameter estimation of γ. Also, since the profile likelihood of β_v and γ flatten for $t \gtrapprox 0.02$ and $t \gtrapprox 25$ respectively, these parameters might be practically non-identifiable if there are more unknown parameters.

Figure 16: Profile likelihood of β_h, β_v, and γ for 31 data points of host incidence data.

For 31 data points of host incidence data, the fitted parameters are $\beta_h = 8.794 \cdot 10^{-5}$, $\beta_v = 0.00115$, and $\gamma = 0.106$. Additional data points can only help better the identifiability of parameters. Data points about the peak of the host incidence data resulted in a tighter confidence intervals for all parameters which reduce the uncertainty in parameter estimation. The width of the confidence interval of β_h did not decrease as much compared to the width of the confidence intervals of β_v and γ. In other words, data about the peak of host incidence data reduce uncertainty in parameter estimation the most in β_h and γ as compared to β_h.

Figure 17: Profile likelihood of β_h, β_v, and γ for 46 data points of host incidence data.

For 46 data points of host incidence data, the fitted parameters are $\beta_h = 8.433 \cdot 10^{-5}$, $\beta_v = 0.00122$, $\gamma = 0.118$. Similar to the case of 31 data points of host incidence data, the width of the confidence intervals decrease due to the additional data. However, the width of the confidence intervals of β_h and β_v did not decrease as much as compared to the width of the confidence interval of γ. Since information is already known about the beginning and peak of the host incidence data, this could have contribute to the data not having as big as an impact in decreasing the confidence interval of the parameters. In all, this suggests that host incidence data is helpful in estimating β_h and to some extent, β_v with low uncertainty but struggle to estimate γ with low uncertainty without a large amount of host incidence data.

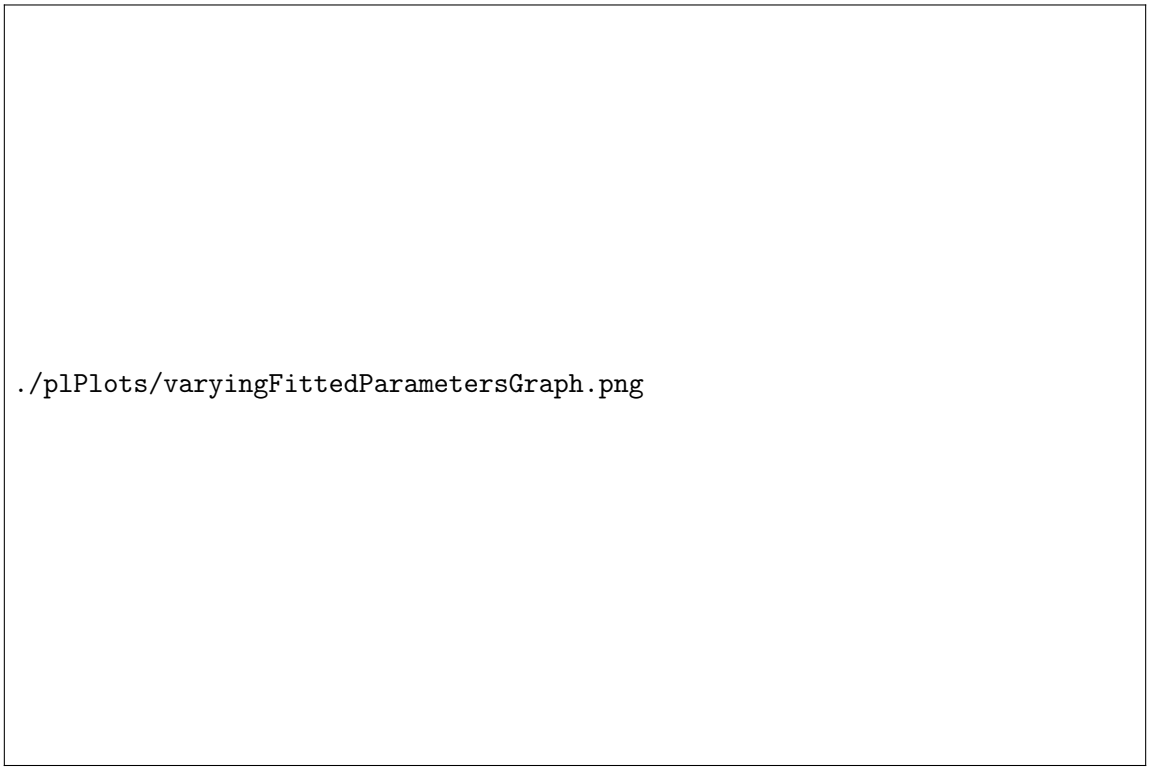

Figure 18: Incidence data plotted from fitted parameters using 16, 31, and 46 data points of host incidence data. Host incidence data from the true parameters and noisy incidence data is also plotted.

	β_h	β_v	γ
16 Data Points	$[4.737 \cdot 10^{-5}, 0.00162]$	$[0, 0.0215]$	$[0, 21.166]$
31 Data Points	$[5.019 \cdot 10^{-5}, 0.000146]$	$[0.000611, 0.00497]$	$[0, 1.183]$
46 Data Points	$[5.452 \cdot 10^{-5}, 0.000133]$	$[0.00069, 0.00309]$	$[0.0753, 0.568]$

Figure 19: Simultaneous confidence intervals for β_h, β_v, and γ for 16 data points, 31 data points, and 46 data points of host incidence data.

	R_0	Relative Error
True Parameters	7.855	0%
16 Data Points	6.973	11.2%
31 Data Points	7.265	7.5%
46 Data Points	6.974	11.2%

Figure 20: R_0 values calculated from the fitted parameters for 16 data points, 31 data points, and 46 data points of host incidence data.

In Figure 18, the host incidence data from the fitted parameters accurately describe the true host incidence data for $t \lessapprox 10$. Similarly, the host incidence data from the fitted parameters mostly capture information after the peak of the host incidence data despite the amount of noise. This contrasts with host incidence data at the peak since where none of the fitted parameters is close to describing the peak of the true host incidence data. This is due to the amount of noise and the data available at the peak of the host incidence data.

In Figure 19, since the confidence intervals are finite and there is an unique minimum for χ^2_{PL} for each parameter, each parameter is identifiable. However, the width of the confidence interval of γ is large relative to the rest of the confidence intervals. This indicate that there is a high uncertainty in estimating γ. As such, more data is required to estimate γ with low uncertainty.

In Figure 20, R_0 values was calculated for each fitted parameters which use 16, 31, and 46 data points. The closest value of R_0 was calculated using the fitted parameters that use 31 data points of host incidence data with a relative error of 7.5%. Meanwhile, the value of R_0 calculated with fitted parameters that use 16 and 46 data points of host incidence data both have a relative error of 11.2%. Even though there is not that much noise for $t \lessapprox 15$, the value of R_0 still have a relative error of 11.2% for the fitted parameters using 16 data points. As such, the addition of data may not necessarily improve the parameter estimation due to overfitting to the noisy data.

Additional data points of a different observation. Adding more data may not necessarily improve parameter estimation due to overfitting. However, adding data of a different observation could better the parameter estimation rather than adding data of the same observation. Thus, we now examine how adding data points of a different observation can affect parameter estimation, confidence interval, and identifiability.

Figure 21: Perfect host incidence data and noisy host incidence data generated from the Poisson distribution. Perfect vector incidence data and noisy vector incidence data generated from the Poisson distribution.

In Figure 21, 31 data points of both host and vector incidence data are shown. We will only analyze the case of 16 and 31 data points of incidence data for hosts and vectors. We choose this amount of data to analyze as the data include information of the beginning and peak of the incidence data. Furthermore, as seen in the last example, there is a limiting effect of improving the parameter estimation and confidence intervals through the addition of more data.

We will use the same noisy incidence data for hosts as in the previous example. The noisy incidence data for vectors is generated using a Poisson distribution and compared to the host incidence data, very little noise is present in the incidence data for vectors.

Figure 22: Profile likelihood of β_h, β_v, and γ for 16 data points of both host incidence data and vector incidence data.

For 16 data points of both host incidence data and vector incidence data, the fitted parameters are $\beta_h = 9.593 \cdot 10^{-5}$, $\beta_v = 0.00101$, $\gamma = 0.0788$. Compared to the confidence intervals constructed using 46 data points of host incidence data in Figure 19, the width of the confidence intervals from only using 16 data points of both host incidence data and vector incidence data is smaller. Also, the lower bound of the confidence interval for γ is 0 since a value of $\gamma < 0$ is meaningless to interpret in the context of this model.

Figure 23: Profile likelihood of β_h, β_v, and γ for 31 data points of both host incidence data and vector incidence data.

For 31 data points of both host incidence data and vector incidence data, the fitted parameters are $\beta_h = 9.329 \cdot 10^{-5}$, $\beta_v = 0.00107$, and $\gamma = 0.0978$. The additional data reduce the most uncertainty in estimating γ while having marginal effects in reducing uncertainty in β_h and β_v. This again illustrate that it is difficult to estimate γ with low uncertainty compared to β_v and β_h using incidence data.

Figure 24: Host incidence data and vector incidence data plotted from fitted parameters using 16 and 31 data points of both host incidence data and vector incidence data. Host incidence data and vector incidence data from the true parameters and noisy host incidence data and noisy vector incidence data is also plotted.

	β_h	β_v	γ
16 Data Points	$[7.942 \cdot 10^{-5}, 0.000119]$	$[0.000753, 0.00140]$	$[0, 0.266]$
31 Data Points	$[7.897 \cdot 10^{-5}, .000111]$	$[0.000903, 0.00128]$	$[0.0727, 0.139]$

Figure 25: Simultaneous confidence intervals for β_h, β_v, and γ for 16 and 31 data points of both host incidence data and vector incidence data.

	R_0	Relative Error
True Parameters	7.855	0%
16 Data Points	8.263	5.2%
31 Data Points	7.518	4.3%

Figure 26: R_0 values calculated from fitted parameters using 16 and 31 data points of both host incidence data and vector incidence data.

In Figure 24, the host incidence data from the fitted parameters also have trouble describing the peak of the true host incidence data. However, the fitted parameters using both types of data fare better than using only host incidence data. Furthermore, since there is very little noise in the vector incidence data, the host incidence data from the fitted parameters accurately describe the true host incidence data. Although the fitted parameters using 16 and 31 data points of both host incidence data and vector incidence data are different, the vector incidence data produced remains the same while the host incidence data differ after the peak. This possibly indicates that host incidence data is

more sensitive to parameter estimation. A parameter sensitivity analysis must be done to verify this.

In Figure 25, since all confidence intervals are finite and an unique minimum exists for the profile likelihood of each parameter, all parameters are identifiable. Furthermore, since the width of the confidence intervals are small, there is a low uncertainty in estimating the parameters.

In Figure 26, the additional vector incidence data help reduce the relative error of the R_0 value compared to the observations being only the host incidence data. This decrease in relative error is about 2% which is noticeable. Furthermore, since the R_0 value differ for 16 and 31 data points of both host incidence data and vector incidence data, this indicates that there is a large variability in the R_0 value and suggests that more data is required to reduce the uncertainty in the R_0 value.

Mischoosing probability distribution. When fitting to experimental data as opposed to simulated data, we will not know the distribution that generated the observed data. We will examine what happen when the "wrong" probability distribution is chosen for the measurement error and its consequences on the estimated parameters, confidence intervals, and identifiability.

We will only choose 31 data points of noisy host incidence data which is the same as in Figure 14. This amount of data points was chosen since the data have information about the peak which was crucial in reducing the uncertainty in estimating γ.

The distribution of the noise will be assumed to follow relative error and constant variance error. The negative log-likelihood when the measurement error follow a relative error distribution is equation 27:

$$-2\log(\boldsymbol{\theta}|\mathbf{y}) = n\log\eta^2 + n\log 2\pi + \frac{1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - g(t_i,\boldsymbol{\theta})}{g(t_i,\boldsymbol{\theta})}\right)^2 + 2\sum_{i=1}^{n}\log g(t_i,\boldsymbol{\theta})$$

and the maximum likelihood estimate is equation 29:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - g(t_i,\boldsymbol{\theta})}{g(t_i,\boldsymbol{\theta})}\right)^2 + 2\sum_{i=1}^{n}\log g(t_i,\boldsymbol{\theta}).$$

The negative log-likelihood when the measurement error follow a normal distribution $\mathcal{N}(0,\sigma^2)$ is equation 15:

$$-2\log L(\boldsymbol{\theta}|\mathbf{y}) = n\log 2\pi + n\log\sigma^2 + \sum_{i=1}^{n}(y_i - g(t_i,\boldsymbol{\theta}))^2/\sigma^2$$

and the maximum likelihood estimate is equation 17:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n}(y_i - g(t_i,\boldsymbol{\theta}))^2/\sigma^2.$$

For these noise distributions, we must choose σ and η respectively for constant variance error and relative noise error. For experimental data, this is typically done by examining the data and performing literature review. However, since this is simulated data, we cannot perform the same analysis to determine σ and η. One method to determine an appropriate value for σ and η is to fit the noise distribution using maximum likelihood estimation. The relative error distribution $\mathcal{N}(0, (\eta \cdot g(t_i, \hat{\boldsymbol{\theta}}))^2)$ should follow the distribution of $(y_i - g(t_i, \hat{\boldsymbol{\theta}}))/g(t_i, \hat{\boldsymbol{\theta}})$ where $\hat{thetab}$ is the fitted parameter using equation 29. Similarly, the normal distribution $\mathcal{N}(0, \sigma^2)$ should follow the distribution of $y_i - g(t_i, \hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ is the fitted parameter using equation 17. Using this method, we choose $\eta = 0.3$ and $\sigma = 5.7$.

Figure 27: Profile likelihood of β_h, β_v, and γ where the measurement error is assumed to follow relative error with a noise level of $\eta = 0.3$.

If the measurement error is assumed to follow relative error with a noise level of $\eta = 0.3$, then the fitted parameters are $\beta_h = 4.910 \cdot 10^{-5}$, $\beta_v = 0.00356$, and $\gamma = 0.593$. Despite the wrong distribution is chosen for the measurement error, the identifiability and confidence intervals of the parameters exhibit similar trends to the case of 31 data points of host incidence data. All parameters are identifiable and the width of the confidence interval of γ is relatively larger than the width of the confidence intervals of β_h and β_v respectively.

However, the width of the confidence intervals constructed using relative error are larger than the width of the confidence intervals in the case of 31 data points of host incidence data in Figure 19.

Figure 28: Profile likelihood of β_h, β_v, and γ where the measurement error follow $\mathcal{N}(0, 5.7^2)$.

If the measurement error is assumed to follow constant variance error with $\sigma = 5.7$, then the fitted parameters are $\beta_h = 8.700 \cdot 10^{-5}$, $\beta_v = 0.00116$, and $\gamma = 0.109$. Unlike the previous case where the measurement error is assume to follow relative error, the width of the confidence intervals are reasonably close to the actual confidence intervals constructed using the maximum likelihood estimate where the measurement error follow a Poisson distribution.

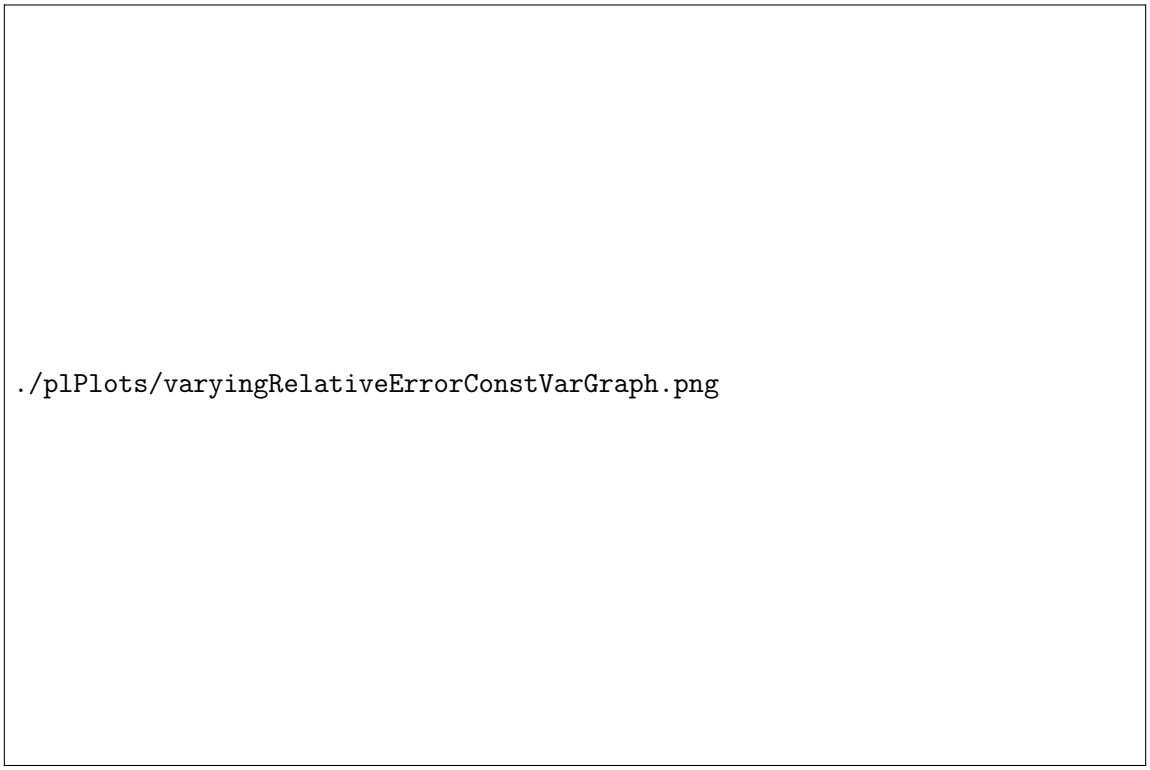

Figure 29: Host incidence data plotted using fitted parameters where the measurement errors is assumed to follow relative error and constant variance error respectively. Host incidence data using the true parameters and the noisy host incidence data is also plotted.

	β_h	β_v	γ
Relative Error	$[4.001 \cdot 10^{-5}, 0.000195]$	$[0.000448, 0.0283]$	$[0.0525, 10.336]$
Constant Variance Error	$[5.0748 \cdot 10^{-5}, 0.000122]$	$[0.000748, 0.00646]$	$[0.0710, 1.685]$

Figure 30: Simultaneous confidence intervals for β_h, β_v, and γ for relative error and constant variance error.

	R_0	Relative Error
True Parameters	7.855	0%
Relative Error	4.047	48.4%
Constant Variance Error	7.176	8.6%

Figure 31: R_0 values calculated from the fitted parameters where the measurement error is assumed to follow relative error and constant variance error respectively.

In Figure 29, both fitted parameters correctly describe the true host incidence data for $t \lessapprox 10$. The host incidence data from the fitted parameters using constant variance error was able to describe the true host incidence data for $t \gtrapprox 20$, but was unable to fully capture information about the true host incidence data at the peak. In contrast, the host incidence data from the fitted parameters using relative error was not able to describe the true host incidence data for $t \gtrapprox 10$. This is because the maximum likelihood estimate look at the difference between the predicted host incidence data and the noisy host incidence data relative to each other. This results in host incidence data that is unable to capture information about the peak and end behavior where a large amount of noise is present.

In Figure 31, the R_0 value calculated from the fitted parameter using relative error performs the worst with a R_0 value of 4.047 and a relative error from the true R_0 value of 48.4%. In contrast, the R_0 value calculated from the fitted parameter using constant variance error performs reasonably well with a R_0 value of 7.176 and a relative error of 8.6%. For comparison, the R_0 value calculated from fitted parameters using a maximum likelihood estimate where the measurement errors is assumed to follow a Poisson distribution is 7.265 with a relative error of 7.5%. This illustrates that ordinary least squares or constant variance error perform reasonably well even when the measurement error does not follow the normal distribution $\mathcal{N}(0, \sigma^2)$.

6 Conclusion

Compartmental models are a decisive tool for predicting the spread of an infectious disease early in a pandemic. However, without proper understanding of the identifiability of a model, one is unable to accurately predict the parameters of that model. We investigated the structural and practical identifiability of multiple variations of a simple vector-host SIS model.

In terms of Structural Identifiability, we found that a mix of the the differential algebra approach and the Taylor series approach often returned different results for identifiability of a few particular parameters. Specifically for variations of our simple vector-host model, we also found that DAISY, while being the most accurate method of computing structural identifiability, is the most computationally inefficient and often becomes less useful when working with more complex models. SAN was the second most computationally efficient method we used, with the IdentifiabilityAnalysis package in Wolfram Alpha being the most computationally efficient. However, this came at the cost of probabilistic accuracy, as shown by the difference in particular parameters. We found that in the software where we could fix initial conditions, all the models we ran were structurally identifiable, which matches our practical identifiability results.

From the Monte Carlo simulation method, we saw that with prevalence data from both populations, all parameters except for the host demographic parameters μ_h and Π_h were practically identifiable. But this was fragile to the choice of initial guess. We then fixed

all demographic parameters, leaving only β_h, β_v, and γ free. Here, all parameters were practically identifiable, even with an incorrect initial guess.

Next, we investigated factors independent of the model structure that impact the results of the Monte Carlo simulation method. We saw that having fewer observations increase the ARE values. However, only a small number of observations were needed to meet the threshold for identifiability. We also changed the final date of observations, and found that parameters became unidentifiable when estimating from before the sharp rise in cases. This shows the difficulties in predicting the future course of an epidemic.

Next, we investigated different types of observation data. We saw that when fitting from cumulative data, the practical identifiability depends greatly on whether the error is in incidence or cumulative data, as error in incidence data is more identifiable. Apart from this, no form of data consistently outperformed the other. We also saw that all parameters were practically identifiable with prevalence data from only vectors or only hosts, although the ARE values were lower with information from both populations.

We also looked at choices of error models and objective functions. We found that when adjusted for population size, ordinary least squares performs decently well on relative noise, whereas generalized least squares does not perform well on absolute noise.

Finally, we investigated the impacts of changes in objective functions, bounds, and initial guess. We found that with a correct initial guess, the objective functions and bounds had little impact on practical identifiability. However, when optimization algorithms may not find the correct minimum, the choice of bounds and optimization algorithm can have a large impact.

So when measuring practical identifiability, it is important to choose conditions that match the way parameters are estimated from real-world data.

Lastly, our analysis using profile likelihood reveals that β_h, β_v, and γ are practically identifiable with as few as 16 data points of host incidence data. However, there is a high uncertainty in estimating γ which require more incidence data to reduce the uncertainty. In estimating parameters and checking identifiability, information about the beginning and peak was the most helpful while information after the peak did not have the nearly the same effect. Also, we shown that more data of the same observation does not necessarily help with parameter estimation due to overfitting. However, data of different observations help much more in estimating parameters and constructing smaller confidence intervals. Lastly, we analyzed how parameter estimation, confidence intervals, and identifiability behave when the wrong distribution is chosen when finding the likelihood function. If the data is generated from the Poisson distribution, then using constant variance error leads to reasonable results while using relative error leads to poor results in estimating parameters and constructing confidence intervals.

7 Future Work

The impact on the bounds and initial guess given to optimization algorithms is a topic for future work. We made a first attempt at using global optimization algorithms for parameter estimation. But none of `ga`, `particleswarm`, `patternsearch`, `surrogateopt`, or `GlobalSearch` were able to identify the parameters of the vector-host model when using default settings.

The choice of noise distribution is another topic for future work. For the Monte Carlo simulation method, noise distributions besides relative and absolute noise can be looked at, such as Poisson noise. The impact of noise distributions that are not iid is another topic worthy of investigation.

The maximum likelihood approach for parameter estimation show that more data is not necessarily going to help fit the parameters due to overfitting. A point of future work is to perform a global sensitivity analysis of the parameters of the SIS Vector-Host model. The sensitivity of the parameters can be used to find the best data to fit the parameters as parameters that are sensitive might require more data to estimate with low uncertainty. Exploring sensitivity of the model is also related to Fisher Information Matrix which can be used for checking identifiability. Another point of future work is parameter estimation using Bayesian methods.

For finding profile likelihood, the code can be further optimized. Parallel processing can be used to compute multiple profile likelihood plots at once. Furthermore, different interpolation schemes such as spline interpolation can be used to reduce the number of points that need to be computed. This can provide a substantial speed up at a cost of a possibly less precise interpolation of χ^2_{PL}. Another technique to speed up the profile likelihood code is to use adaptive step size as opposed to a fixed step size when finding points to compute for χ^2_{PL}.

8 Acknowledgements

We would like to thank Vadim Ponomarenko, San Diego State University, and the National Science Foundation for hosting and supporting this research. And a special thank you to Dr. Tingting Tang and Anuradha Agarwal for their guidance and mentorship.

A Benchmark on Solving Differential Equations

Figure 32: Work-precision benchmark done on the SIS Vector-Host Model using the DiffEqDevTools package from DifferentialEquations ecosystem [Rackauckas and Nie, 2017]. Comparison between Julia, Matlab, and Scipy's Differential Equations Solvers. Calling from Julia: ≈ 2 ms overhead for Matlab and ≈ 3 times speed up for SciPy. Similar benchmarks can be found at `https://github.com/SciML/SciMLBenchmarks.jl`.

References

[Bellu et al., 2007] Bellu, G., Saccomani, M. P., Audoly, S., and D'Angiò, L. (2007). Daisy: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine*, 88(1):52–61.

[Bezanson et al., 2017] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98.

[Brauer et al., 2008] Brauer, F., van den Driessche, P., and Wu, J. (2008). *Mathematical Epidemiology*. Lecture Notes in Mathematics #1945. Springer, 1 edition.

[Broyden, 1970] Broyden, C. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA journal of applied mathematics*, 6(1):76–90.

[Byrd et al., 1996] Byrd, R. H., Gilbert, J. C., and Nocedal, J. (1996). *A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming*.

[Capaldi et al., 2012] Capaldi, A., Behrend, S., Berman, B., Smith, J., Wright, J., and Lloyd, A. L. (2012). Parameter estimation and uncertainty quantification for an epidemic model. *Mathematical Biosciences and Engineering*, 9(3):553–576.

[D'Errico, 2022] D'Errico, J. (2022). fminsearchbnd,fminsearchcon. `https://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon`. [Online; accessed August 4, 2022].

[Eisenberg, 2009] Eisenberg, M. (2009). Identifiability - 2019 tutorial workshop on parameter estimation for biological models.

[Feldt and Stukalov, 2018] Feldt, R. and Stukalov, A. (2018). Blackboxoptim.jl. `https://github.com/robertfeldt/BlackBoxOptim.jl`.

[Flanders and Kleinbaum, 1995] Flanders, W. D. and Kleinbaum, D. G. (1995). Basic Models for Disease Occurrence in Epidemiology. *International Journal of Epidemiology*, 24(1):1–7.

[Gill, 1981] Gill, P. E. (1981). *Practical optimization*. Academic Press, London ; New York.

[Glatter and Finkelman, 2021] Glatter, K. A. and Finkelman, P. (2021). History of the plague: An ancient pandemic for the age of covid-19.

[Hong et al., 2020] Hong, H., Ovchinnikov, A., Pogudin, G., and Yap, C. (2020). Global identifiability of differential models. *Communications on Pure and Applied Mathematics*, 73(9):1831–1879.

[Karlsson et al., 2012] Karlsson, J., Anguelova, M., and Jirstrand, M. (2012). An efficient method for structural identifiability analysis of large dynamic systems*. *IFAC Proceedings Volumes*, 45(16):941–946.

[Kolda et al., 2003] Kolda, T. G., Lewis, R. M., and Torczon, V. (2003). Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Rev.*, 45(3):385–482.

[Lagarias et al., 1998] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM journal on optimization*, 9(1):112–147.

[Larsen and Marx, 2006] Larsen, R. and Marx, M. (2006). *An Introduction to Mathematical Statistics and Its Applications*. Pearson Prentice Hall.

[Mejía et al., 2022] Mejía, J., Satman, M. H., and Monticone, P. (2022). jmejia8/metaheuristics.jl: v3.2.9.

[NIAID, 2016] NIAID (2016). Dengue fever. `https://www.niaid.nih.gov/diseases-conditions/dengue-fever`.

[Nocedal, 2006] Nocedal, J. (2006). *Numerical optimization*. Springer series in operations research. Springer, New York, 2nd ed.. edition.

[Price, 2013] Price, K. V. (2013). *Differential Evolution*, pages 187–214. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Rackauckas and Nie, 2017] Rackauckas, C. and Nie, Q. (2017). Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia. *Journal of Open Research Software*, 5(1).

[Raue et al., 2010] Raue, A., Becker, V., Klingmüller, U., and Timmer, J. (2010). Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4):045105.

[Raue et al., 2009] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.

[Tuncer and Le, 2018] Tuncer, N. and Le, T. T. (2018). Structural and practical identifiability analysis of outbreak models. *Math. Biosci.*, 299:1–18.

[WHO, 2020] WHO (2020). Vector-borne diseases. `https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases`.