Mathematical Approaches for Simulating Epidemic Progression: Addressing Limitations of the Linear Chain Trick in ODE Models

Submitted to

Dr. Jonathan Dushoff

McMaster University

Hamiltion, Ontario, Canada L8S 4K1

March 22, 2024

Reported by

Ningrui Xie

400323416

1   Abstract

Simple dynamical models based on Ordinary Differential Equations (ODE) are widely utilized to forecast the spread of infectious diseases and to analyze other biological systems. A limitation of this framework is its implicit assumption that residence time in any given state, such as the infectious state, follows an exponentially distributed pattern when a constant rate is employed. This assumption is often not realistic. A common solution is the application of the Linear Chain Trick (LCT), wherein a single stage is subdivided into multiple substages, each follows an identical exponential distribution, thereby yielding a more realistic Erlang distribution for the duration. However, the LCT poses challenges by restricting parametric flexibility, particularly resulting in less flexible dwell time distributions, and it often necessitates comparisons between models with varying structures (i.e. different numbers of substages). Here, we propose a Geometric Chain Trick (GCT) employing a fixed number of substages with transition rates forming a geometric sequence. Through mathematical analysis, simulation, and model fitting, we demonstrate that our approach retains the principal benefits of the LCT while providing increased flexibility and facilitating easier data fitting. Unlike the recently introduced Generalized Linear Chain Trick (GLCT), which incorporates higher-dimensional auxiliary stages to model delays in transitions between main compartments, our method offers a more direct means of achieving flexibility. Although this study focuses on the application of the LCT and GCT in the SIR (susceptible-infectious-recovered) model, the potential applications of our approach span a broad array of dynamical models, including those for infectious diseases and other systems (e.g. population dynamics, biochemical reaction networks), offering advantages in computational efficiency, ease of parameter estimation, and the flexibility of time distributions.

2   Introduction

Ordinary differential equations (ODEs) are extensively utilized for mathematical modeling in the biological sciences, offering a robust framework to accurately describe the temporal dynamics of systems. Their capacity to model continuous state changes makes them essential in

capturing complex biological phenomena such as population dynamics [1][2], epidemiological patterns of infectious diseases [3][4][5], and cellular processes like proliferation in oncogenesis [6][7]. The inherent flexibility and simplicity of ODEs make them not only easy to solve and understand but also easy to manipulate and accommodate to various situations.

In the field of epidemiology, ordinary differential equations (ODEs) serve as a crucial tool for characterizing disease progression and making predictions. During the COVID-19 pandemic, they were extensively utilized to guide policy-making, especially in disease control and public health interventions [8]. A commonly used model is the SIR compartmental model [3][9], which divides the population into three states: Susceptible (S), Infectious (I), and Recovered (R). In the model, a susceptible individual, upon contact with the infected population at a rate β ($\frac{1}{time}$), transitions immediately to the infectious state. An infected individual then moves to the recoverd state at a recovery rate of γ ($\frac{1}{time}$). However, one major limitation of this model it its implicit assumption that the duration of the infectious stage follows an exponential distribution. This can be observed in a simplified equation that models the dynamics of the infected population. It assumes the entire population was infected at the outset, and individuals can only transition out of the infectious stage at a rate of γ:

$$\frac{dI}{dt} = -\gamma I \tag{1}$$

Solving this equation explicitly yields

$$I(t) = I_0 e^{-\gamma t} \tag{2}$$

where I_0 represent the initial infected population. Formula (2) implies a stochastic state transition model where the time spend in the infectious I is exponentially distributed with a mean stage duration of $\frac{1}{\gamma}$. This dwell time distribution suggests that an individual has the highest probability of recovering from the disease immediately after infection, which is often not realistic. In reality, a delay between the successful infection of an individual and the recovery from the disease is inevitable. The work conducted by Byrne AW et al. [10] presents the simulation of parameter distribution inferred for the duration of the infectious

3

period, and the shape of this distribution closely resembles a gamma distribution. The study by Greenhalgh and Day (2017) [11], utilizing measles outbreak data, also demonstrates that the recovery rate from an infectious disease is not static but can vary based on a wide range of factors.

It is important to note that even when the mean is constant, the shape of the stage duration distributions (i.e., the timing of stage transitions) can significantly impact the dynamics and the outcomes of predictive models in applied settings [12][13][14][15]. Consequently, it becomes crucial to consider the type of distribution characterizing an individual in a specific state when constructing these models. A standard approach for developing models that incorporate different stage duration distributions is through the use of integro-differential equations (IDEs) or integral equations (IEs) [16][9][17]. IDEs facilitate the extension of the SIR model to include arbitrary distributions for the duration of infectiousness, as discussed in various sources [18][17][19]. This approach allows for the capture of the 'memory effect', where the current rate of change in each compartment of the model (Susceptible, Infectious, Recovered) is influenced by the entire history of the epidemic, rather than relying on a constant rate assumption.

However, while integro-differential equations increase the model's flexibility, they present challenges in formulation, mathematical analysis, and simulation, particularly in comparison to ODEs [12][16][20]. An alternative method, the Linear Chain Trick (LCT), overcomes some of these challenges, offering a simpler and more direct approach to modifying the shape of distributions [21][22]. This approach subdivides the single infected state into n substages, each characterized by an identical exponential distribution with a rate of $n\gamma$. This technique is based on the principle that the sum of a series of independent exponentially distributed random variables follows an Erlang distribution [12][23], which is a special case of the gamma distribution with an integer shape parameter. By employing this method, the distribution of the infectious stage duration transitions from exponential to Erlang. The central tendency of the Erlang distribution towards its mean value offers the model a more accurate representation of real-world scenarios.

When applying LCT into the SIR model, the number of substages n determines the shape of the stage duration distribution. As n increases, the variance decreases, resulting in a distribution that is more peaked and narrow around the mean value $\frac{1}{\gamma}$. As n approaches infinity, the distribution converges to a delta function (a deterministic value, implying that all individuals who become infectious at time t recover at exactly time $t + \frac{1}{\gamma}$), and the system evolves into a delay differential equation [12][17].

The Linear Chain Trick (LCT) is extensively employed in modeling biological processes, not only in quantifying disease progression [24][25], but also in analyzing virus [26][27] and population [28] dynamics. Studies have demonstrated that switching the stage duration distribution from an exponential to an Erlang distribution leads to predictions that align more closely with real data points [27]. However, despite the increased realism they bring to modeling disease processes, employing LCT in ODE models presents several limitations. One primary challenge is the inflexibility of the model's parameters. To integrate independent exponentially distributed random variables into the ODE system, n must be an integer. As mentioned above, since the shape of the stage duration distribution is determined by n, it would also be constrained to discrete changes. Another inherent limitation of using LCT is the cumbersome and time-consuming process of determining the appropriate number of substage to divide. Multiple models with varying substages must be constructed and subjected to repeated data-fitting processes to identify the optimized parameter value. Prior research has investigated approaches for estimating the number of states in the linear chain, employing techniques such as profile likelihood [29] and model reduction [30]. Yet, a comparative study of these methods indicates that the typical quality of experimental data is insufficient for reliably estimating the chain length using these approaches [31]. These limitations significantly restrict the adaptability and practical application performance of the model.

The recently proposed Generalized Linear Chain Trick (GLCT) enhances the Linear Chain Trick (LCT) by enabling models to incorporate a broader phase-type family of distributions [16][32][33], which includes exponential, Erlang, hypoexponential, and Coxian distributions, thus allowing for greater flexibility and the ability to capture more complexity. The GLCT

achieves this by creating a higher-dimensional Markov chain, enabling a more accurate representation of the complex nature of dwell time distributions. However, this process can be mathematically heavy and challenging to interpret. Additionally, this approach may lead to a increase in the number of variables or states within the model, potentially escalating computational demands. Therefore, finding a simpler and more efficient method to enhance the flexibility of the ODE model remains essential.

In this study, we focused on one particular direction: generalizing the Erlang distribution of stage durations currently present in the model. The goals were to have a more effective way for parameter estimation and to allow for a more diverse range of shapes, achieved by eliminating the constraints imposed by the shape parameter (κ). To address these limitations, we proposed the Geometric Chain Trick (GCT). Instead of having a variable number of substages (n) with a constant transition rate between them ($\gamma_i = n\gamma$), we fix n to a specific integer value and use a series of rates that follow a geometric pattern. This series is represented by the formula $\gamma_i = ar^{i-1}$, governed by parameters a and r. This approach enables a uniform model structure with the parameter of interest transitioning from discrete values (n and ng) to continuous values (a and r), thereby enhancing the model's adaptability (details in Section 4.x).

To illustrate this point, consider that the Gamma distribution can be parameterized in terms of its mean (D) and shape parameter (quadratic coefficient of variation, κ), which determine the overall shape and position of the distribution. As mentioned above, in subdividing a single stage into n substages, each exponentially distributed substage must have a rate of $n\gamma$ to maintain the overall rate of that stage. While the mean of the Erlang distribution in the ODE system can be chosen freely, κ is constrained to discrete values, shown as follows:

$$Mean: \quad \sum_{i=1}^{n_f} \frac{1}{n\gamma} = \frac{1}{\gamma}$$

$$Variance: \quad \sum_{i=1}^{n_f} \frac{1}{n^2\gamma^2} = \frac{1}{n\gamma^2}$$

$$\kappa: \quad \frac{\sigma^2}{M^2} = \frac{1}{n} \quad n \in \mathbb{N}$$

Our approach, using a geometric series of rates, addresses this issue by allowing κ to take any value within the range $\left[\frac{1}{n}, 1\right)$. A detailed explanation and the derivation process are provided in section 4.2. Another advantage of this method is that it enables a fixed model structure where only the parameters change, eliminating the need to determine optimal substage numbers through the construction of multiple models, a process currently required in the ODE model that employs the LCT.

Throughout this research, we primarily focused on the application of the LCT in the epidemic SIR model, which we refer to as the SI^nR model (or LCT model), where n represents the number of substages divided. The objective of our study is to develop and evaluate the method of using a fixed model structure and a geometrically distributed substage rate as a replacement for the constant rate, aiming to enhance flexibility and efficiency. We have named the model that utilizes this method the GCT model. We anticipate that this approach will not only be computationally efficient but also provide a better fit to real-world data. In the following sections, we first analyze the model parameters by calculating the formulae for two essential properties: the mean and shape parameter. We then invert the process: given a desired D and κ, we develop a computational system to determine the parameter values (a and r) that enable the model to produce a stage duration distribution with these properties. The obtained parameter values are then used to compare the GCT model with the LCT model through simulation. Subsequently, we conduct empirical validation of our model through data fitting. This fitting confirmed that the GCT model enables the use of a uniform model structure with enhanced flexibility and practicality, thus paving the way for the development of user-friendly software tools that empower modelers to efficiently select and employ the most suitable models for describing and predicting future epidemic scenarios.

3   Method

Several assumptions have been made for the unforced SIR model (Susceptible, Infectious, and Recovered) that our study is based on:

1. **Closed Population**: A constant population size with equal birth and death rates and

no disease-related mortalities.

2. **Homogeneous mixing**: Individuals have an equal chance of coming into contact with any other individual in the population.

3. **Ignoring Latency**: Individuals move immediately from a susceptible state to an infected state, ignoring any incubation period associated with the disease.

4. **Lifelong immunity**: Once an individual recovers, they do not become susceptible or infectious again.

5. **Fixed rate of transmission and recovery**: The rate of disease spread and the rate at which infected individuals recover and become immune are constant over time.

The dynamic of the population within the three compartment can be expressed as a system of nonlinear ordinary differential equations:

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} - \mu S \tag{3}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \mu I \tag{4}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{5}$$

where S, I, and R represent the susceptible, recovered, and infectious population. N is the size of the total population, and in this case $S + I + R = N$. The μ, β and γ are the rates of natural birth (death), transmission and recovery, respectively. β is the rate of new infections per unit time, so βSI is the number of new infections that occur per unit time (the incidence rate). Together, formulas $(3, 4, 5)$ will be referred to as the transmission model. Here, the mean recovery time for an infected individual is $\frac{1}{\gamma}$. By assuming that the entire population was infected at the onset and considering a zero natural birth (death) rate, we can focus solely on the dynamics of the recovered population, as indicated by formula (1).

The idea of dividing the infectious stage into n sub-stages exploits the fact that the sum of a sequence of independent, exponentially distributed random variables follows an Erlang

distribution (i.e. gamma distributed with integer shape parameter) [23]

$$f(x; n, n\gamma) = \frac{(n\gamma)^n}{(n-1)!} x^{n-1} e^{-n\gamma x} \quad x > 0, n \in \mathbb{N} \tag{6}$$

Utilizing the principle of the Linear Chain Trick (LCT), we divide the single infectious stage into n substages, each with a constant rate of $n\gamma$. Therefore, formula (1) will be replaced by the following set of equations, which we refer to as the cohort model:

$$\frac{dI_1}{dt} = -n\gamma I_1 \tag{7}$$

$$\frac{dI_2}{dt} = n\gamma I_1 - n\gamma I_2 \tag{8}$$

$$\vdots \tag{9}$$

$$\frac{dI_n}{dt} = n\gamma I_{n-1} - n\gamma I_n \tag{10}$$

$$\frac{dR}{dt} = n\gamma I_n \tag{11}$$

Note that this decomposition process will not change the mean stage duration, as the sum of the means of the n exponentially distributed random variables still equals $\frac{1}{\gamma}$

$$\sum_{i=1}^{n} \frac{1}{n\gamma} = \frac{1}{\gamma}$$

The Geometric Chain Trick (GCT), on the other hand, features a fixed number of substages along with a set of inter-substage transition rates characterized by a geometric series, rather

than a constant rate, as illustrated below:

$$\frac{dI_1}{dt} = -aI_1 \tag{12}$$

$$\frac{dI_2}{dt} = aI_1 - arI_2 \tag{13}$$

$$\frac{dI_3}{dt} = arI_2 - ar^2I_3 \tag{14}$$

$$\vdots \tag{15}$$

$$\frac{dI_n}{dt} = ar^{n-2}I_{n-1} - ar^{n-1}I_n \tag{16}$$

$$\frac{dR}{dt} = ar^{n-1}I_n \tag{17}$$

Having established the model, our initial step involved analyzing the parameters within the GCT model. We aim to determine whether employing a fixed model structure (i.e., a fixed number of substages) and manipulating the parameters a and r, which define the geometric transition rate, can replicate the effects of varying the number of substages. Additionally, we aim to evaluate whether the GCT structure offers greater flexibility in the choice of κ (CV^2). By calculating the mean (D), variance (σ), and shape parameter (κ), we are able to represent all these properties of the infectious stage duration distribution in terms of a and r. Similar to the LCT, utilizing the GCT facilitates independent alterations in the mean and κ. Given this, we can invert the process: starting with specific values of D and κ, we can identify the corresponding a and r that endow the model with such characteristics. However, due to the complexity of the formula, solving it explicitly is challenging. To overcome this, we have developed a system of functions in the **R** programming language, employing `uniroot()` to determine the values of a and r. These parameters are then integrated to the GCT model to compare simulation outcomes with those of the LCT model, examining alignment in both the infectious stage duration distribution and the incidence over time.

After confirming that the GCT model, which employs a fixed structure and varying parameters a and r, achieves comparable efficacy to the LCT model, we proceeded to empirically test the model's enhanced performance. Specifically, we focused on its improved flexibility in the kappa parameter through data fitting. Initially, we utilized simulated data as a re-

placement for actual data. This incidence data was generated through transmission model simulations, with deviations (noise) introduced via a mean reporting ratio and a negative binomial dispersion parameter. We performed model fitting using the **bbmle** package in R, estimating the model parameters through Maximum Likelihood Estimation (MLE). Our goal was to accurately retrieve the parameter values used to generate the incidence data, assessing the fitting quality through the mean squared error (MSE).

We began by applying the same model type (LCT or GCT) for both data generation and model fitting. Data were generated by simulating the LCT model with two substages ($n = 2$) and then fitted using the LCT model, evaluating both identical and modified structures. This methodology was similarly applied to the GCT model. To further test the GCT model, we increased the fitting challenge by generating incidence data using the LCT model and then fitting it with the GCT model, varying the number of substages. Subsequently, we modified both the fitting model's structure and the input data to observe the impact on fitting outcomes. Finally, we extracted the parameters estimated through fitting and compared them with the actual values (those used for simulation).

In the next phase of our study, we will approach the fitting of interval data from two perspectives: the rounding paradigm, which interprets 2 days as meaning any duration between 1.5 and 2.5 days, and the start/end paradigm, which considers potential starting and ending times. Initially, we will generate interval data using various distributions, such as gamma and lognormal, and subsequently apply both the LCT and GCT models for fitting. Although the GCT model can replicate certain properties (such as D and κ) observed with the gamma distribution, its inherent distribution does not match the gamma distribution precisely. Thus, it's necessary for us to develop a new likelihood function for fitting purposes. Furthermore, we intend to incorporate real-world data, including incidence and interval data, to evaluate the GCT model's performance comprehensively. By comparing the fitting results from both the LCT and GCT models, we aim to quantitatively determine the effectiveness of our proposed method in enhancing model accuracy and its reflection of actual epidemiological scenarios.

4   Results

4.1   Independent and Flexible Modification of Mean and Kappa in Stage Duration Distribution Using the GCT Model

As discussed in previous sections, the duration distribution of the stage can be parameterized by its mean (D) and shape parameter ($\kappa^2 = CV^2$). In the GCT model, the calculation of mean and variance is achieved by summing the respective properties of each substage, shown as follows:

$$Mean: \quad \sum_{i=1}^{n_f} \frac{1}{\gamma_i} = \frac{\frac{1}{a}(\frac{1}{r^n} - 1)}{\frac{1}{r} - 1} \tag{18}$$

$$Variance: \quad \sum_{i=1}^{n_f} \left(\frac{1}{\gamma_i}\right)^2 = \frac{\frac{1}{a^2}(\frac{1}{r^{2n}} - 1)}{\frac{1}{r^2} - 1} \tag{19}$$

The shape parameter κ, determined by $\frac{\sigma^2}{D}$, can thus be obtained:

$$\kappa = \frac{(1 + r^n)(1 - r)}{(1 + r)(1 - r^n)} \tag{20}$$

This simplification revealed that the value of κ is solely determined by the parameter r, thus enabling the alignment of kappa without affecting the mean, which can be adjusted through alterations to parameter a. Similarly, the mean can be scaled to a desired value without impacting κ. This grants the model independence in modifying properties.

Importantly, within the GCT model, even though the number of substages can be modified (n), the value of κ becomes independent of this number. It should also be noted that, according to formula (20), the range of κ can encompass any real number within the range $\left[\frac{1}{n}, 1\right)$ for $r > 0$, thereby significantly enhancing the model's flexibility. In contrast, when applying a similar calculation for κ in the LCT model, we find that $\kappa = \frac{1}{n}$ and the shape parameter is therefore constrained to a series of discrete values, as follows:

$$\kappa \in \left\{1, \frac{1}{2}, \frac{1}{3}, \dots\right\}$$

The transition from discrete to continuous values in kappa significantly enhances the model's adaptability to the shape of the stage duration distribution, enabling a more precise fit to the observed data and allowing for a nuanced interpretation of temporal patterns.

4.2 Matching GCT and LCT Model Performance through Parameter Adjustment within a Uniform Structure

Using formulas (18) and (20), we developed code to reverse-engineer the process: given specific values of D and kappa, we can identify the corresponding a and r that endow the model with such characteristics. These a and r values were subsequently integrated into the GCT model. Our initial goal was to determine if the stage duration distribution could match between the two models, assuming identical D and κ values. The number of substages in the LCT model was set to 2, 4, 6, and 8, while the substage number for the GCT model was fixed at 12. We selected a mean of 10 days. Applying the mean and kappa values $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}\right)$ to the system, we plotted the stage duration distributions generated from the LCT cohort model alongside those from the GCT with the corresponding a and r values. As depicted in Figure x.x, the stage duration distributions exhibit slightly different shapes, despite having identical mean and shape parameters. To further examine the match of properties, D and kappa were numerically calculated from the output distributions, revealing identical values across both models. This indicates that while the distributions' shapes may not completely match, they can possess the same properties.

By reintegrating the infectious stage duration distribution into the whole dynamic system, we present the dynamics of newly infected cases (incidence) as simulated by the transmission model. Figure x.x illustrates the incidence time series for the LCT model with 4 and 8 substages, represented by yellow and blue lines, respectively. The corresponding kappa values of the stage duration distribution for them are $\frac{1}{4}$ and $\frac{1}{8}$. Employing the GCT model with a fixed 12 substages and utilizing a and r values derived from the reverse-engineering system for the targeted kappa values, we successfully approximate the incidence patterns observed in the LCT model across different substage configurations, as shown by the black dashed line. This demonstrates the GCT model's capability to emulate the LCT model's
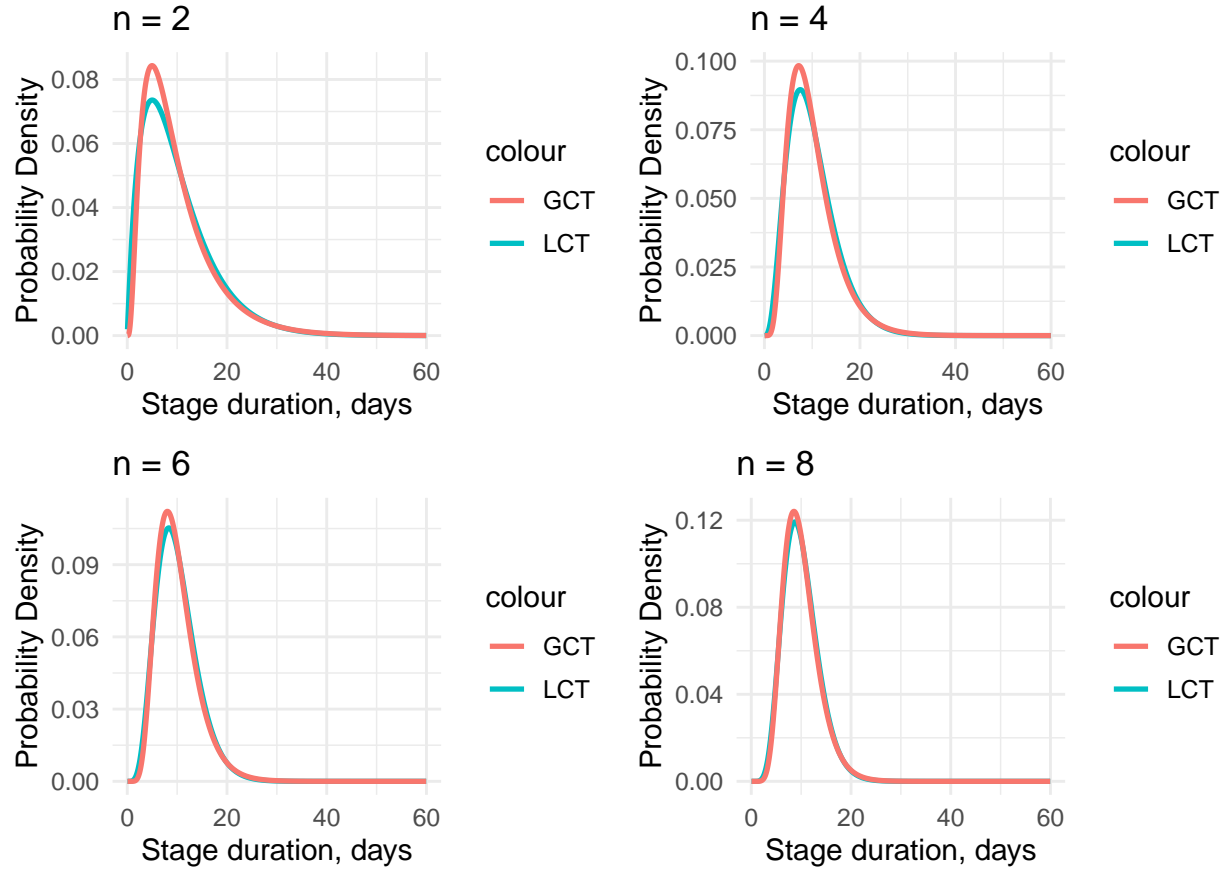
Figure 1: IMPORTANT

performance by adjusting parameters within a constant structural framework. Compared to modifying the model structure to achieve desired kappa values, the GCT model simplifies parameter estimation and model fitting processes.

4.3   Substage Configuration Minimally Impacts Model Fitting

Only temporarily. I'm going to change all of this...

After establishing the theoretical foundation, we proceeded with model fitting to assess the GCT model's performance. The incidence data were obtained from the model simulation, and the noises were added using a mean reporting ratio of 0.9 and a negative binomial dispersion parameter of 1000. Initially, we employed the same model type for both data generation and fitting. In Figure x.x.A, the data, generated by the LCT model with 2 substages, were fitted using an LCT model of the identical structure, as the plot title indicates. The quality
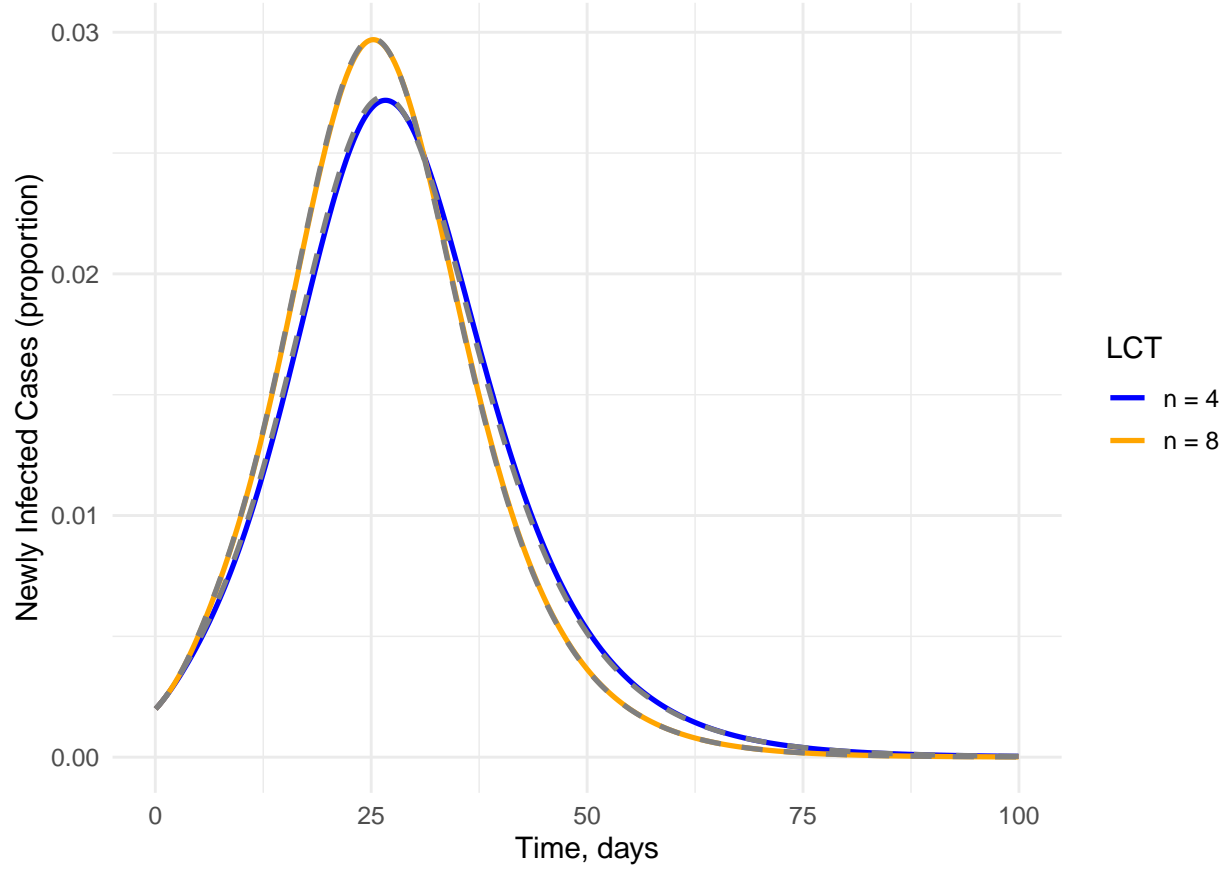
Figure 2: IMPORTANT

of the fitting was evaluated based on the mean squared error (MSE). With an MSE of 1.2, the fitting can be considered satisfactory. Figure x.x.B showcases the parameters estimated post-fitting. Conversely, employing the same incidence data but altering the fitting model's structure from an LCT with 2 substages to an LCT with 6 substages, as depicted in Figure x.x.C, resulted in less favorable outcomes, with the MSE escalating to 25. The estimated parameters, displayed in Figure x.x.D, showed significant discrepancies.

On the other hand, employing the GCT model with 12 substages for both generating and fitting incidence data resulted in a satisfactory fit, as indicated by an MSE of 0.01 (Figure x.x.A), and parameters post-fitting were close to the exact values (Figure x.x.B). However, altering the fitting model's structure from 12 substages to 6 substages, the model still exhibited a close match with the original in terms of both shape (Figure x.x.C) and parameters (Figure x.x.D), albeit with a slightly increased MSE (Figure x.x.C). This indicates that the
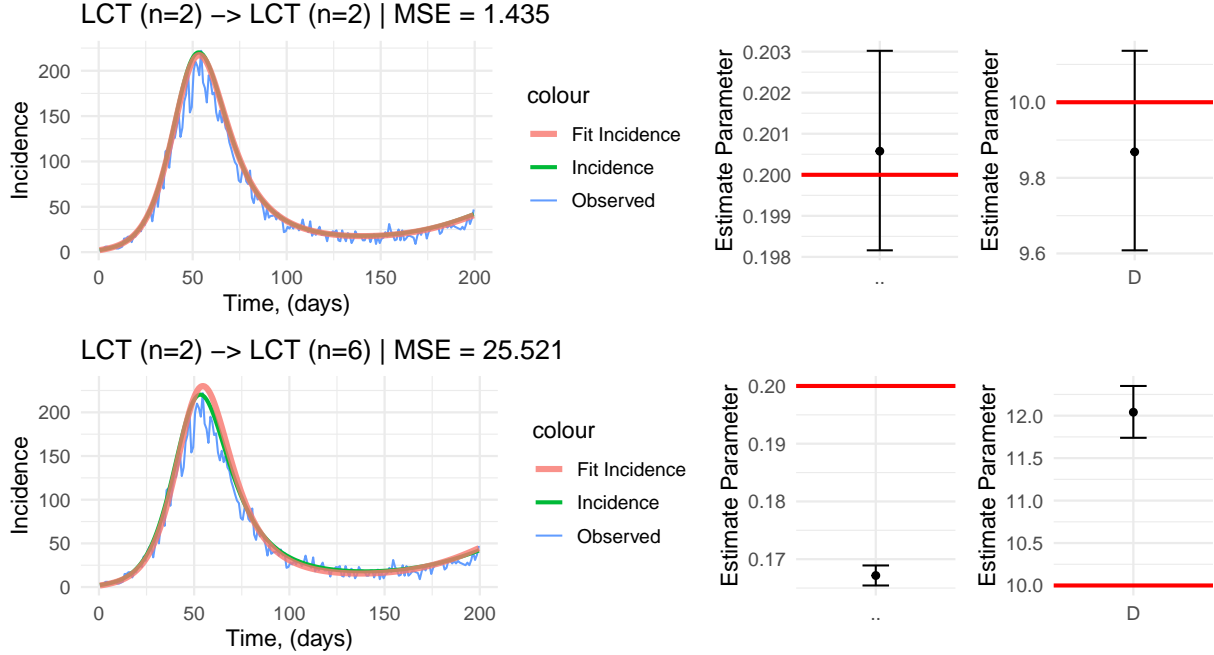
Figure 3: IMPORTANT

GCT model's fitting performance is minimally impacted by changes in the model structure.

4.4 Adaptability of the GCT Model to Data Input and Structural Changes in Fitting

Proceeding to a more challenging test, we utilized a different model type for data generation and model fitting. As demonstrated in Figure x.x.A, the data generated by the LCT model with 2 substages were fitted using the GCT model with 12 substages. Despite the significant difference in model structure, the fitting results were reasonably favorable, with an MSE of 2.4. Altering the fitting model from a GCT with 2 substages to a GCT with 6 substages had negligible impact on the fitting outcomes (Figure x.x.B). Similarly, changing the input data from an LCT with 2 substages to an LCT with 7 substages resulted in slightly improved fitting, though the difference was not substantial (Figure x.x.C). These findings suggest that regardless of changes to the fitting model structure or the data used, the GCT model is capable of adapting to these variations, providing satisfactory fitting results.
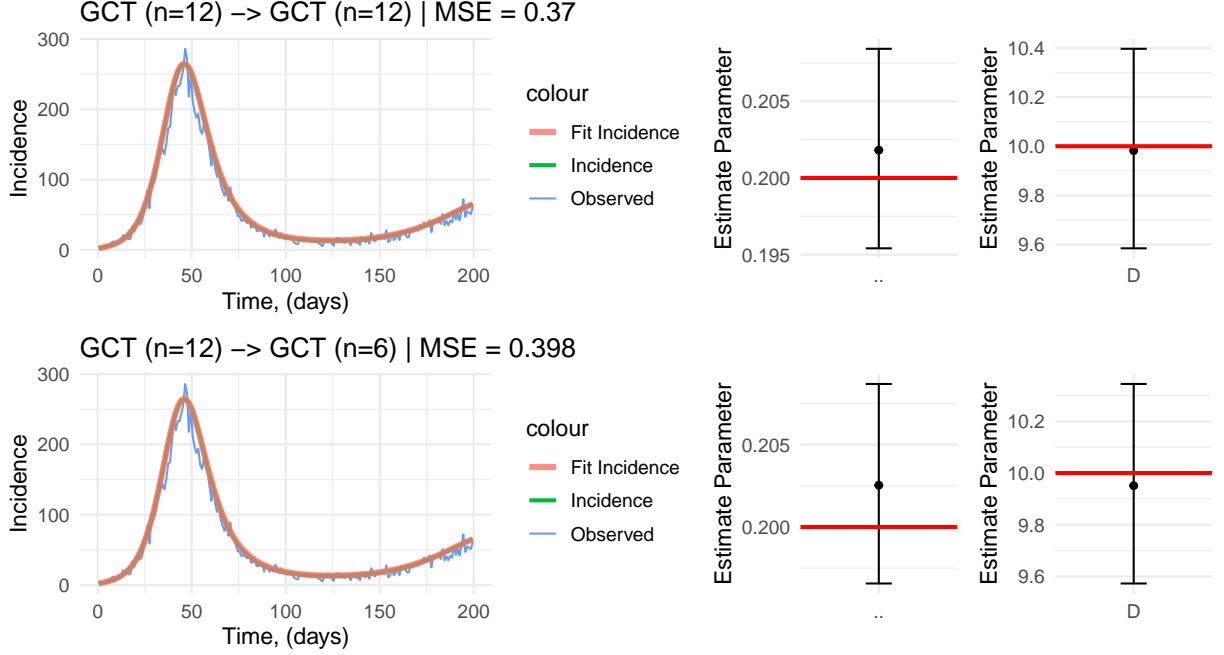
Figure 4: IMPORTANT

4.5   Parameter Estimation

In this section, we discuss the parameter estimation results obtained through model fitting.
The incidence data were generated using the GCT model with 5 substages, setting the D
to 10 and κ to $\frac{2}{9}$. The model was fitted using the GCT framework with 5 to 10 substages.
As depicted in Figure x.x, the precise values for each parameter are marked by a red line,
with post-fitting parameter values displayed alongside their 95% confidence intervals. It can
be observed that the parameter estimation is not perfect, with kappa and beta consistently
overestimated, and D underestimated. However, these discrepancies in parameter estimation
counterbalance each other, resulting in a fitting outcome characterized by a low mean squared
error. It is possible for the parameter to lack confidence intervals, which could be attributed
to structural inadequacies. As previously mentioned, the minimum possible value for kappa
in the model with n substages is $\frac{1}{n}$. The model encounters limitations when aligning with
a kappa value of less than $\frac{1}{n}$. Overall, it is important to acknowledge the imprecision of
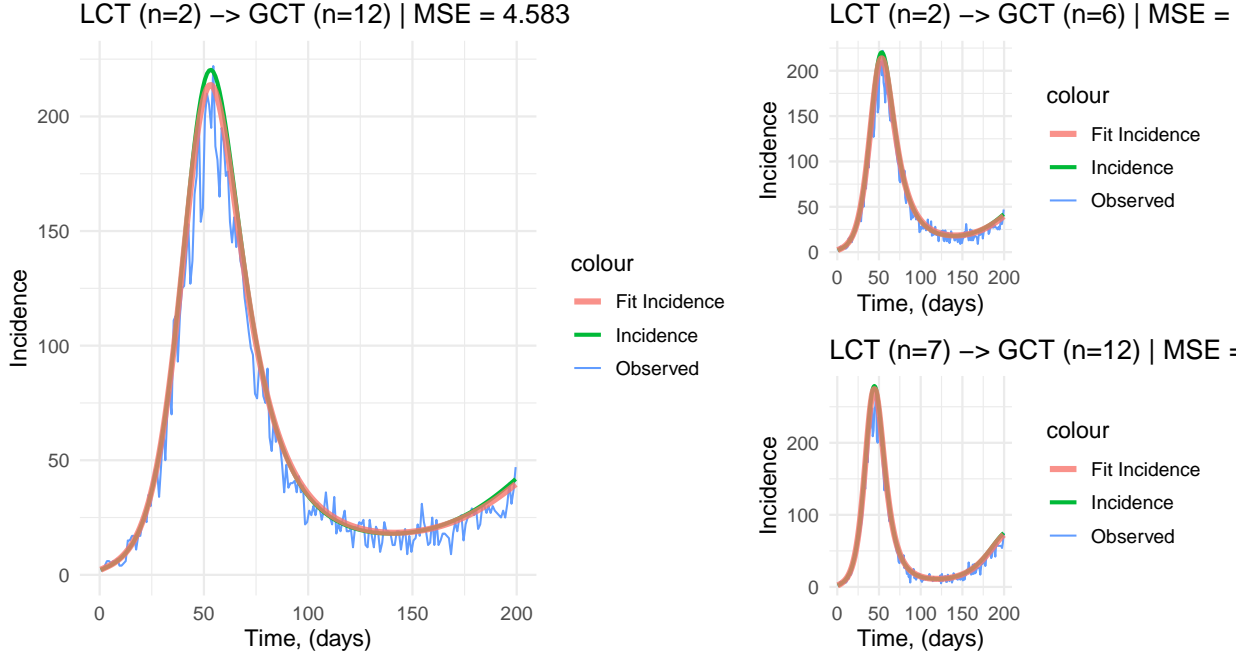parameter estimation, which will be a focus of our subsequent efforts.

17

Figure 5: IMPORTANT

5   Discussion

In this study, we introduce the Geometric Chain Trick (GCT), characterized by a fixed number of substages and a geometric series of transition rates. This approach aims to address the limitations identified in the Linear Chain Trick (LCT), including rigid parameter values and a cumbersome structure determination process. Our results indicate that the GCT model can utilize a uniform structure to achieve efficacy comparable to that of the LCT model with varying substage structures. The comparison is based on the consistency of simulation results, examining both the cohort model (stage duration distribution) and the transmission model (incidence over time). This enhancement simplifies the optimization of parameter values, requiring only a one-time process during data fitting. Furthermore, by theoretically analyzing and empirically validating the model through data fitting, we demonstrate that our model offers enhanced flexibility, specifically a broader range for the possible values of κ, in modeling the distribution of infectious stage duration. This flexibility allows for the selection of mean (M) and shape parameters (κ) for the distribution, with $M \in \mathbb{R}^+$ and $\kappa \in [\frac{1}{n}, 1)$.
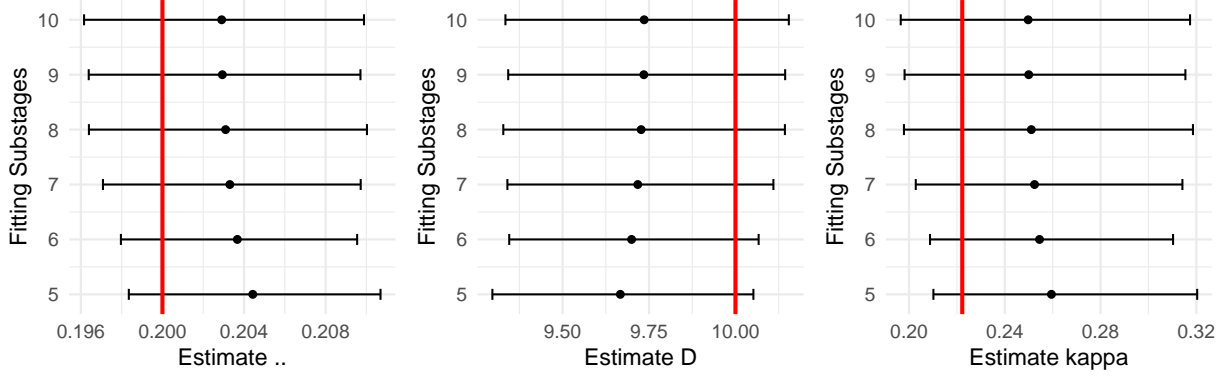
Figure 6: IMPORTANT

Compared to current studies that aim to provide a more flexible dwell time distribution within the ODE framework, such as the Generalized Linear Chain Trick (GLCT) proposed by Hurtado and Kirosingh [16], our method offers a simpler concept and structure for constructing, simulating, and analyzing models. This simplicity makes our approach more accessible to a wide range of scientists and fields of study. Additionally, our method allows for considering a more diverse substage transition pattern, for instance, a higher dimension transition matrix (e.g. Markov Chain). This idea is inspired by the GLCT [16][32], yet it is less complex. In our Geometric Chain Trick (GCT), transitions between substages follow a sequential pattern, where each substage receives a unique input and progresses to another distinct substage, with no cycles among them. Modifying this pattern of transitions between substages could possibly yield a wider variety of stage duration distributions.

There are potential challenges that need to be addressed in the current GCT model. As discussed in Section 4.x, the parameters estimated after data fitting do not precisely reflect the actual values, yet these values are often crucial for understanding disease dynamics. For example, the basic reproduction number, R_0, which indicates how many secondary infections can be expected from a single infected individual in a fully susceptible population, is essential for assessing the potential for disease outbreak and epidemic spread. The calculation of R_0 requires the use of the contact rate (β), recovery rate (γ), and birth/death rate (μ). Therefore, achieving accurate parameter values from model fitting is necessary, and more study is needed.

In addition to seeking solutions to the challenges discussed and exploring the diverse transition patterns mentioned above, future efforts will also involve fitting the GCT model to real-world incidence and interval data to assess its performance. We also plan to integrate this framework into the SEIR (Susceptible, Exposed, Infectious, Recovered) model to better account for delayed responses. Many infectious diseases exhibit a notable incubation period during which the pathogen replicates within the host without making the host infectious. Incorporating an exposed phase into the SEIR model enables a more accurate representation of disease transmission dynamics across populations.

In conclusion, this study suggests that our method simplifies the application of a uniform model architecture for data fitting and exhibits heightened adaptability for parameter value selection within an Ordinary Differential Equation (ODE) framework. Specifically, in the field of epidemiology, our research holds the potential to improve model performance and facilitate the creation of user-friendly tools for robust epidemic forecasting and simulation. The method demonstrates considerable flexibility in its structure, making it suitable to a wide array of applications across different biological systems.

References

[1] Alexander Salisbury. Mathematical models in population dynamics. 2011.

[2] TJ Sego, Josua O Aponte-Serrano, Juliano F Gianlupi, and James A Glazier. Generation of multicellular spatiotemporal models of population dynamics from ordinary differential equations, with applications in viral infection. *BMC biology*, 19(1):1–24, 2021.

[3] Anderson RM and May RM. *Infectious diseases of humans: dynamics and control.* Oxford university press, 1991.

[4] Diekmann O and Heesterbeek JA. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation.* John Wiley and Sons, 2000.

[5] Feng Z, Zheng Y, Hernandez-Ceron N, Zhao H, Glasser JW, and Hill AN. Mathematical

models of ebola—consequences of underlying assumptions. *Mathematical biosciences*, 227:89–107, 2016.

[6] CTH Baker, GA Bocharov, CAH Paul, and FA Rihan. Modelling and analysis of time-lags in some basic patterns of cell proliferation. *Journal of mathematical biology*, 37:341–371, 1998.

[7] Angela M Jarrett, Ernesto ABF Lima, David A Hormuth, Matthew T McKenna, Xinzeng Feng, David A Ekrut, Anna Claudia M Resende, Amy Brock, and Thomas E Yankeelov. Mathematical models of tumor cell proliferation: A review of the literature. *Expert review of anticancer therapy*, 18(12):1271–1286, 2018.

[8] Robin N Thompson. Epidemiological models are important tools for guiding covid-19 interventions. *BMC medicine*, 18:1–4, 2020.

[9] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[10] Andrew William Byrne, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, et al. Inferred duration of infectious period of sars-cov-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic covid-19 cases. *BMJ open*, 10(8):e039856, 2020.

[11] Scott Greenhalgh and Troy Day. Time-varying and state-dependent recovery rates in epidemiological models. *Infectious Disease Modelling*, 2(4):419–430, 2017.

[12] Olga Krylova and David JD Earn. Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *Journal of The Royal Society Interface*, 10(84):20130098, 2013.

[13] Matt J Keeling and Bryan T Grenfell. Understanding the persistence of measles: recon-

ciling theory, simulation and observation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1489):335–343, 2002.

[14] Helen J Wearing, Pejman Rohani, and Matt J Keeling. Appropriate models for the management of infectious diseases. *PLoS medicine*, 2(7):e174, 2005.

[15] Hanh TH Nguyen and Pejman Rohani. Noise, nonlinearity and seasonality: the epidemics of whooping cough revisited. *Journal of the Royal Society interface*, 5(21):403–413, 2008.

[16] Paul J Hurtado and Adam S Kirosingh. Generalizations of the 'linear chain trick': incorporating more flexible dwell time distributions into mean field ode models. *Journal of mathematical biology*, 79(5):1831–1883, 2019.

[17] Herbert W Hethcote and David W Tudor. Integral equation models for endemic infectious diseases. *Journal of mathematical biology*, 9:37–47, 1980.

[18] Zhilan Feng and Horst R Thieme. Endemic models with arbitrarily distributed periods of infection i: fundamental properties of the model. *SIAM Journal on Applied Mathematics*, 61(3):803–833, 2000.

[19] Junling Ma and David JD Earn. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of mathematical biology*, 68:679–702, 2006.

[20] Ted A Burton. *Volterra integral and differential equations*. Elsevier, 2005.

[21] N MACDONALD. Time lags in biological models. 1978.

[22] Hal L Smith. *An introduction to delay differential equations with applications to the life sciences*, volume 57. springer New York, 2011.

[23] Charles Therrien and Murali Tummala. *Probability and random processes for electrical and computer engineers*. CRC press, 2018.

[24] Alun L Lloyd. Destabilization of epidemic models with the inclusion of realistic dis-

tributions of infectious periods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1470):985–993, 2001.

[25] Alun L Lloyd. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theoretical population biology*, 60(1):59–71, 2001.

[26] Alun L Lloyd. The dependence of viral parameter estimates on the assumed viral life cycle: limitations of studies of viral load data. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1469):847–854, 2001.

[27] Yusuke Kakizoe, Shinji Nakaoka, Catherine AA Beauchemin, Satoru Morita, Hiromi Mori, Tatsuhiko Igarashi, Kazuyuki Aihara, Tomoyuki Miura, and Shingo Iwami. A method to determine the duration of the eclipse phase for in vitro infection with a highly pathogenic shiv strain. *Scientific reports*, 5(1):10371, 2015.

[28] *Integrodifferential equations and delay models in population dynamics*, volume 20. Springer Science & Business Media, 2013.

[29] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[30] Tim Maiwald, Helge Hass, Bernhard Steiert, Joep Vanlier, Raphael Engesser, Andreas Raue, Friederike Kipkeew, Hans H Bock, Daniel Kaschek, Clemens Kreutz, et al. Driving the model to its limit: profile likelihood based model reduction. *PloS one*, 11(9):e0162366, 2016.

[31] Adrian L Hauber, Raphael Engesser, Joep Vanlier, and Jens Timmer. Estimating chain length for time delays in dynamical systems using profile likelihood. *Bioinformatics*, 36(6):1848–1854, 2020.

[32] Paul J Hurtado and Cameron Richards. Building mean field ode models using the

generalized linear chain trick & markov chain theory. *Journal of Biological Dynamics*, 15(sup1):S248–S272, 2021.

[33] Mogens Bladt, Bo Friis Nielsen, Mogens Bladt, and Bo Friis Nielsen. Phase-type distributions. *Matrix-Exponential Distributions in Applied Probability*, pages 125–197, 2017.