Estimating time-varying transmission rates of the SIR model

Sang Woo Park

supervised by

Dr. Ben Bolker

A thesis presented for Math 4P06

Department of Mathematics & Statistics
McMaster University
April 9, 2019

Abstract

Seasonally-varying transmission rates drive irregular dynamics of childhood infections, such as measles. One way to infer transmission rates from a time series is to try to match the predicted dynamics of a model, whether deterministic or stochastic, with the observed dynamics. If a system is highly nonlinear and near chaotic, however, matching its time series can be difficult. Another way to approach this problem is to transform the time series and estimate transmission rates using a regression. In this study, we compare several different methods for estimating transmission rates against simulations and discuss how their assumptions affect their estimates of the transmission rates. We show that regression-based methods can correctly estimate the exact shape of the transmission rates but give biased estimates. Simulation-based methods are not able to infer the exact shape of the transmission rates precisely but give unbiased estimates of the transmission rates with good coverage properties. [**SWP:** *Come back*]

Acknowledgements

I would like to thank Dr. Ben Bolker and Dr. Jonathan Dushoff for their guidance throughout my undergraduate studies. Their enthusiasm for statistics and biology has been truly inspiring. Without them, I would not have known about the joy of doing research.

I would also like to thank the Mac-Theobio group for all the interesting scientific discussions we've had; Chyun-Fung Shi for making the lab a welcoming environment; Dr. David Earn, Dr. Stephen Ellner, Dr. Aaron King, and Dr. Bryan Grenfell for useful ideas and comments related to this project.

Finally, I would like to thank my family for their love and support.

Contents

1 Introduction

Over the last few decades, a plethora of computational tools have been developed to study irregular patterns observed in epidemic time series, such as those of measles (Fine and Clarkson, 1982; Grenfell et al., 1994; Finkenstadt and Grenfell, 2000; Cauchemez and Ferguson, 2008; Hooker et al., 2010; Xia et al., 2011; King et al., 2015; Fasiolo et al., 2016). Central to these studies are mechanistic models, which are sets of equations that represent our understanding of the underlying biological system (Bretó et al., 2009). Mechanistic modeling of disease dynamics allows us to study the qualitative dynamics of a system and infer underlying process that drove the observed patterns in the data. For example, we now know that the complex dynamics of measles outbreaks are driven by seasonal transmission patterns (Earn et al., 2000; Dalziel et al., 2016).

Major challenges in epidemiological modeling come from developing the actual model and parameterizing it. One must make an appropriate set of assumptions to balance generality, realism, and precision (Levins, 1966). These assumptions may depend on a system as well as an inferential tool. Fortunately, a simple epidemic model, such as the Susceptible-Infected-Recovered (SIR) model can describe the measles dynamics sufficiently well (Earn et al., 2000; Krylova and Earn, 2013; Hempel and Earn, 2015; Dalziel et al., 2016). Focusing on a simple, well-understood model allows us to make a fair comparison between different statistical methods in their ability to estimate underlying parameters; if any changes have to be made to the model to apply a statistical tool, they will be relatively easy to understand and control. Here, we focus on estimating the time-varying transmission rate of the SIR model.

One way to estimate the transmission rate is by matching the deterministic trajectory of the system with the observed data by minimizing their "distance", often measured by the sum of squares difference or negative log-likelihood (Riley et al., 2003; Chowell et al., 2004). This method of fitting a deterministic model (often referred to as trajectory matching) effectively assumes that *all* variation in data can be explained by observation error alone (Bolker, 2008). Likewise, it is possible to match gradients of the system (1) with the observed gradient, which can be estimated by fitting a smooth curve to a time series, such as piecewise polynomials, and taking its derivative (Ellner et al., 2002). These methods are computationally efficient and easy to implement.

Monte Carlo methods are computationally demanding but allow us to model *all* sources of variation (e.g., observation, demographic, and environmental stochasticity) explicitly. In particular, Sequential Monte Carlo methods (also known as particle filtering) has been successfully applied in mechanistic analyses of epidemic time series (Ionides et al., 2006; Bretó et al., 2009; He et al., 2009, 2011; Didelot et al., 2017); its ability to approximate the likelihood through a series of simulations provides a powerful technique for testing hypotheses and inferring underlying parameters (Ionides et al., 2006; Bretó et al., 2009; King et al., 2015). Approximate likelihood obtained by particle filtering can be maximized by iterated filtering (Ionides et al., 2011, 2015) or can be incorporated into a Bayesian scheme as a particle Monte Carlo Markov Chain (MCMC) (Andrieu

et al., 2010).

Unlike any of the previously discussed methods, which can be used to estimate parameters of a wide range of systems, the time series SIR (TSIR) method seeks to estimate time-varying transmission rates of fully immunizing childhood infections, specifically (Finkenstadt and Grenfell, 2000). The TSIR method first estimates the dynamics of the susceptible population based on the assumption that all individuals eventually become infected; using the estimated susceptible dynamics, transmission rates can be estimated using a regression. Even though the TSIR method is designed to tackle a specific kind of problems, it has been consistently used since its introduction. It is able to explain complex dynamics of measles (Finkenstadt and Grenfell, 2000; Bjørnstad et al., 2002; Grenfell et al., 2002; Finkenstädt et al., 2002; Metcalf et al., 2009; Mantilla-Beniers et al., 2009; Ferrari et al., 2010; Dalziel et al., 2016; Mahmud et al., 2017), cholera (Koelle and Pascual, 2004; Pascual et al., 2008), malaria (Pascual et al., 2007), rubella (Metcalf et al., 2010, 2011, 2013), chikungunya (Perkins et al., 2015), varicella (Jackson et al., 2014; Baker et al., 2018), and Hand, Foot, and Mouth Disease (Takahashi et al., 2016; Van Boeckel et al., 2016; Du et al., 2017; Takahashi et al., 2018). So what makes it work so well?

The purpose of this project is twofold: (1) to understand how different model assumptions affect our inference and (2) to compare each method's ability to estimate transmission rates. We use stochastic simulations to assess model assumptions and performances. Finally, we give practical guidance for developing epidemiological models and inferring their parameters.

2 Methods

2.1 SIR model

The Susceptible-Infected-Recovered (SIR) model is one of the most basic epidemic models (Kermack and McKendrick, 1927). It describes how disease spreads in a homogeneously mixing population. Here, we present the model as a set of ordinary differential equations:

$$\begin{aligned} \frac{dS}{dt} &= b(t) - \left(\beta(t)\frac{I}{N} + \mu\right) S \\ \frac{dI}{dt} &= \beta(t) S \frac{I}{N} - (\gamma + \mu) I \\ \frac{dR}{dt} &= \gamma I - \mu R \end{aligned} \tag{1}$$

where S is the number of susceptible individuals, I is the number of infected individuals, and R is the number of recovered (or removed) individuals. We assume that individuals enter the susceptible pool at a rate $b(t)$. Susceptible individuals become infected at a rate of $\beta(t)$ by contacting infected individuals. Infected individuals recover at a rate of γ; in other words, the infectious period

is exponentially distributed with mean $1/\gamma$ time units. Finally, all individuals experience natural mortality at a rate of μ.

Here, we mainly consider two sets of stochastic simulations: (1) a single outbreak with constant transmission rate (and no movement in and out of the susceptible population) and (2) recurrent outbreaks driven by sinusoidal transmission rates. To simulate a single outbreak, we use the following parameterization: $\beta(t) = 2, \gamma = 1, \mu = b(t) = 0, N = 100000$. Initial conditions are specified as follows: $I(0) = 10$ and $S(0) = N - I(0)$. To simulate recurrent outbreaks, we use the following parameterization: $\beta(t) = b_0(1 + b_1 \cos(t/26 \times 2\pi)), b_0 = 500/26, b_1 = 0.15, \gamma = 1, \mu = 1/(50 \times 26), N = 5000000, b(t) = \mu N$. Initial conditions are specified as follows: $I(0) = 1 \times 10^{-4} \times N$ and $S(0) = 0.05 \times N$. Parameters for recurrent epidemics are chosen to approximately match the parameters used in Earn et al. (1998), which yields a persistent biennial cycle. Note that γ is scaled to 1 to represent one generation; for measles, $1/\gamma$ will have a unit of biweeks (14 days). Stochastic simulation of the SIR model is performed using the Gillespie algorithm (Gillespie, 1976).

2.2 Epidemic time series

Epidemic time series can be broadly classified into three categories: incidence, mortality, and prevalence. Incidence is defined as the number of newly infected individuals generated over a reporting period. Then, true incidence of the SIR model (1) between time t and $t - t_{\text{rep}}$, where t_{rep} is the length of reporting time step, can be obtained by integrating total infection rate over the reporting period:

$$i_t = \int_{t-t_{\text{rep}}}^{t} \beta(s) S \frac{I}{N} ds. \tag{2}$$

Similarly, true mortality, which is defined as the number of individuals that died (or recovered for measles) during a reporting period, can be obtained by integrating total death rate over the reporting period:

$$m_t = \int_{t-t_{\text{rep}}}^{t} \gamma I ds. \tag{3}$$

Finally, prevalence is defined as the number of infected individuals that are present in the population; this corresponds to the state variable I. Incidence, mortality, and prevalence for a stochastic simulation are defined in the same manner.

Reported incidence, mortality, and prevalence cases are drawn from a beta-binomial distribution with reporting rate ρ, dispersion parameter θ_{rep}, and their corresponding true values (Morris, 1997). Unless noted otherwise, we use $\theta_{\text{rep}} = 10$ throughout simulations. Here, we assume that cases are reported instantaneously; a more realistic model can incorporate a delay between infection time and reporting time (e.g., infected individuals are likely to report after symptom onset but before recovery). We do not explore these ideas explicitly.

Instead, we focus on how different assumptions about an epidemic time series (whether it represents incidence, mortality, or prevalence) affect our inference. While this question was initially motivated by the TSIR framework, in which incidence is assumed to be equal to prevalence when reporting time is equal to mean generation time, misclassification of epidemic time series is not rare. For example, He et al. (2009) assume that incidence is measured upon recovery whereas Hooker et al. (2010); Althaus (2014) assume that incidence is measured upon transition from latent to infectious period. Understanding these differences is crucial in order to correctly estimate underlying parameters as well as their associated uncertainties. For all other purposes, we use the incidence time series throughout this study.

2.3 Fitting methods

Time-series SIR model

The time-series SIR (TSIR) method was developed by Finkenstadt and Grenfell (2000) to estimate seasonally varying transmission rates of measles. The TSIR method begins by discretizing the infection process of the SIR model, assuming that disease generation time is equal to reporting interval:

$$\begin{aligned} S_{t+1} &= B_t + S_t - i_{t+1} \\ i_{t+1} &= \beta_t S_t \frac{i_t^\alpha}{N_t} \end{aligned} \tag{4}$$

where α is often referred to as a conversion factor for discretization of a system or inhomogeneity (Liu et al., 1986; Bjørnstad et al., 2002; Glass et al., 2003). Taking log on both sides, estimation of the transmission rates β_t becomes a regression problem:

$$\log i_{t+1} = \log \beta_t + \log S_t + \alpha \log i_t - \log N_t + \epsilon_t. \tag{5}$$

A key assumption behind the TSIR model is that all, if not most, individuals become infected eventually (Finkenstadt and Grenfell, 2000). In other words, the total number of births should be equal to the total number of cases over a long period of time. Rewriting true incidence i_t as a product of reported cases $i_{t,\text{rep}}$ and reciprocal of reporting rate $1/\rho_t$ and iteratively solving (4), we get

$$\sum_{t=1}^{N} B_t = \sum_{t=1}^{N} \frac{i_{t,\text{rep}}}{\rho_{t+1}} + Z_{t+1} - Z_1, \tag{6}$$

where $Z_t = S_t - \bar{S}$. Under this assumption, fitting a regression model to cumulative birth as a function of cumulative cases allows us to estimate time-varying reporting rate ρ_t by taking the reciprocal of the slope and susceptible dynamics Z_t from the residuals (Finkenstadt and Grenfell, 2000). Conditioning on the known susceptible dynamics, we can apply the TSIR regression to epidemic time series to estimate the time-varying transmission rate.

Trajectory matching

Trajectory matching is a way of estimating underlying parameters of an ODE model by matching the observed time series and a deterministic solution of the model. Given observed incidence $i_{t,\text{rep}}$ (or equivalently, observed prevalence or mortality), we can define its expected dynamics using the solution of the SIR model:

$$\mu_t = \rho i_t = \rho \int_{t-t_{\text{rep}}}^{t} \beta(s) S \frac{I}{N} ds, \tag{7}$$

where ρ is the mean reporting rate. Directly computing this integral is practically infeasible. Instead, we can include an extra state variable C which represents cumulative number of cases to the SIR model: $dC/dt = \beta SI/N$. True incidence can be calculated by taking the difference of C between to consecutive reporting periods. Then, we get

$$\mu_t = \rho(C(t) - C(t - t_{\text{rep}})). \tag{8}$$

Finally, the observation likelihood can be defined as a deviation from this mean trajectory:

$$i_{t,\text{rep}} \sim \text{NegBin}(\mu_t, \phi_{\text{obs}}), \tag{9}$$

where ϕ_{obs} is the dispersion parameter of a negative binomial distribution.

Gradient matching

Like trajectory matching, gradient matching seeks to find a set of parameters such that the gradient of an ODE model matches the observed gradient (Ellner et al., 2002). Gradient matching provides a flexible way of modeling a system as it does not require the parametric form of the rate equations to be prespecified (e.g., we do not need to assume that per-capita rate of transmission is a linear function of I). Instead, parameters can be estimated by fitting a nonparametric regression model to the observed gradient as a function of relevant state variables. Therefore, all state variables must be observed in order to apply gradient matching.

In practice, it is difficult to observe all state variables of an epidemic model, even for simple models such as the SIR model. When either incidence or mortality are reported, *no* relevant state variables are observed. Nonetheless, it is still possible to estimate the transmission rates of the SIR model using gradient matching by making similar assumptions as the TSIR model. Here, we briefly outline the estimation process; note that the purpose of this section is to introduce the method, rather than to provide a rigorous justification of the statistical model.

First, we begin by assuming that the mean generation time is equal to the length of reporting time step and that incidence is proportional to prevalence, i.e., $i_t = aI_t$ for some $a > 0$. This assumption is slightly weaker than the conditions for the TSIR method. We then fit a smooth curve $f(t)$ to log reported

incidence $\log i_{t,\text{rep}}$:

$$\log i_{t,\text{rep}} \sim \text{Normal}\left(f(t), \sigma^2_{\text{smooth}}\right), \tag{10}$$

where natural cubic splines are used to model the function f. For the purpose of estimating time-varying transmission rates, placing basis knots every 6 generations appear to be sufficient.

The smooth curve $f(t)$ represents the dynamics of the log of prevalence ($\log I_t$) up to a constant. When we take the derivative of the curve $f(t_i)$, we get an estimate of the derivative of the log of prevalence:

$$E\left[\frac{df(t)}{dt}\right] = \beta(t)S - (\gamma + \mu), \tag{11}$$

where the unobserved dynamics of the susceptible S can be inferred from the susceptible reconstruction of the TSIR method. There are several ways we can estimate the transmission rate from this equation. First, we can model transmission rate as a function of susceptibles S by fitting a non-parametric regression by treating known values of γ and μ as offset terms:

$$\frac{df(t)}{dt} \sim \text{Normal}(g(S_t) - (\gamma + \mu), \sigma^2), \tag{12}$$

where g is a non-parametric function that represents the transmission rate. When transmission varies over time, we can either (1) model transmission rate as an interaction term between a factor variable that represents time of the year and a smoothing term (i.e., fit a separate nonparametric curve g for each factor level) or try to separate transmission rate from the susceptible dynamics using a log link function:

$$\frac{df(t)}{dt} + (\gamma + \mu) \sim \text{Normal}\left(\exp\left(\log \beta_t + \log S_t\right), \sigma^2\right), \tag{13}$$

where log number of susceptible individuals $\log S_t$ is treated as an offset term and the log transmission rate can be estimated is estimated as a function of time of year using a nonparametric regression. Here, we mainly focus on the last method.

Sequential Monte Carlo

Calculating the exact likelihood of a nonlinear stochastic system is practically impossible. Instead, Sequential Monte Carlo (also known as particle filtering) allows the likelihood to be approximated via simulations (Doucet et al., 2001; Arulampalam et al., 2002). Let $X_t = \{\mathbf{x}_{t,i} \mid i = 1, 2, \ldots, n\}$ be the set of states (particles) at time t and h be the process model. For the SIR model, each particle $\mathbf{x}_{t,i} \in X$ will consist of the number of susceptible individuals S_t and the number of infected individuals I_t. Then, we can predict the state of each particle at next time reporting step $t+1$ using the process model: $h(\mathbf{x}_{t,i})$. Note that it may take several *simulation* time steps to get to the next reporting time

step. For each prediction $h(\mathbf{x}_{t,i})$, we can calcuate the likelihood $w_{t,i}$ of the observation at time $t+1$ and define marginal likelihood by taking the average: $\bar{w}_t$. Then, drawing n weighted samples from the set $\{h(\mathbf{x}_{t,i}) \,|\, i = 1 = 2, \ldots, n\}$, where the weight of each $h(\mathbf{x}_{t,i})$ is proportional to w_i, yields the next set of particles X_{t+1}. This algorithm can be repeated until the final reporting time step. Finally, the complete likelihood of the data is given by the product of the marginal likelihood at each time step: $\prod \bar{w}_t$.

While Sequential Monte Carlo provides a straightforward way of approximating the likelihood for any process model that can be defined, it is still difficult fit a continuous-time stochastic model. For example, it can take as long as 10 hours to simulate a measles times series for 20 years with a gillspie algorithm. Computing a likelihood with large number of particles will take more than a few days; maximizing the likelihood will be practically impossible.

Instead of fitting a continuous-time model, we can assume that an epidemic process is a partially observed discrete-time Markov process and discretize the system by using a hazard approximation to the transition probabilities between states:

$$\begin{aligned}
B(t) &\sim \text{Poisson}(b(t)\Delta t) \\
F_S(t) &\sim \text{Binomial}(S_t, 1 - \exp(-(r_1 + r_2)\Delta t)) \\
F_I(t) &\sim \text{Binomial}(I_t, 1 - \exp(-(r_3 + r_4)\Delta t)) \\
F_{S,I}(t) &\sim \text{Binomial}\left(F_S(t), \frac{r_1}{r_1 + r_2}\right) \\
S(t + \Delta t) &= B(t) + S(t) - F_S(t) \\
I(t + \Delta t) &= FS, I(t) + I(t) - F_I(t) \\
C(t + \Delta t) &= C(t) + FS, I(t)
\end{aligned} \tag{14}$$

where $r_1 = \beta I/N$, $r_3 = \gamma$, and $r_2 = r_4 = \mu$. Δt represents simulation time step; as $\Delta t \to 0$, this model converges to a Gillespie simulation. Random variables $F_S(t)$ and $F_I(t)$ represent the total number of individuals that move out from susceptible S and infected I states. Likewise, $F_{S,I}(t)$ represent the total number of individual move from susceptible state S to infected state t. Finally, the state variable $C(t)$ keeps track of cumulative incidence. Again, we define observation likelihood using a negative-binomial distribution:

$$i_{t,\text{rep}} \sim \text{NegBin}(\rho(C(t) - C(t - t_{\text{rep}})), \phi_{\text{obs}}). \tag{15}$$

This model is implemented in the R package `POMP` (King et al., 2015).

Renewal equation

Many standard compartmental models, such as the SIR model, assume that the shape of generation-interval distribution is known exactly. The continuous-time SIR model assumes that generation intervals are exponentially distributed. The discrete-time SIR model represented as a partially observed Markov process assumes that generation intervals are geometrically distributed. Finally, the TSIR model assumes that generation intervals are fixed. We expect a realistic

generation-interval distribution of measles to be narrower than exponential but not fixed. While it is possible to use differently shaped generation-interval distributions by incrementing the number of compartments used for each state, the number of compartments has to be specified prior to model fitting.

Instead, renewal equation allows us to model incidence as a function of a previous incidence and generation-interval distributions explicitly:

$$i(t) = \mathcal{R}_0 S(t) \int_0^t g(t-s) i(s) ds, \tag{16}$$

where $i(t)$ represents incidence, $S(t)$ represents number of susceptible individuals, $g(t)$ represents generation-interval distribution and $\mathcal{R}_0$ represents the basic reproductive number. Because the renewal equation generalizes the compartmental models with Erlang distributed latent and infectious periods (i.e., models with multiple latent and infectious compartments), including the standard SIR model (Champredon et al., 2018), we expect this model to perform better in analyzing time series of outbreaks.

Previous studies used a discrete time version of the renewal equation model to analyze Ebola epidemics but assumed that the simulation time step was equal to the reporting time step (Li et al., 2018; Champredon et al., 2018). Here, we extend their models by relaxing this assumption and allowing for natural birth and death in the susceptible population:

$$\begin{aligned}
B(t) &\sim \text{Poisson}(b(t)\Delta t) \\
\phi(t) &= \mathcal{R}(t) \sum_{k=1}^{\ell} i(t-(k-1)\Delta t) g(k\Delta t) \\
F_S(t) &\sim \text{Binomial}\left(S(t), 1 - e^{-(\phi(t)+\mu\Delta t)}\right) \\
i(t+\Delta t) &\sim \text{Binomial}\left(F_S(t), \frac{\phi(t)}{\phi(t)+\mu\Delta t}\right) \\
S(t+\Delta t) &= B(t) + S(t) - F_S(t)
\end{aligned} \tag{17}$$

This model can be fitted using a Bayesian machine such as JAGS (Plummer et al., 2003), NIMBLE (de Valpine et al., 2017), and Stan (Carpenter et al., 2017) (see Li et al. (2018) for comparison across these platforms).

In order to make a fair comparison (without relying on a Bayesian prior), we take a further step to translate this model into a Markov process such that it can be used with particle filtering. We define a set of new state variables $j_1, j_2, \ldots, j_\ell$ such that $j_k(t)$ corresponds to $i(t-(\ell-k)\Delta t)$. Then, we can

rewrite the model as follows:

$$\begin{aligned} B(t) &\sim \text{Poisson}(b(t)\Delta t) \\ \phi(t) &= \mathcal{R}(t)\sum_{k=1}^{\ell} j_{\ell-k+1}(t)g(k\Delta t) \\ F_S(t) &\sim \text{Binomial}\left(S(t), 1-e^{-(\phi(t)+\mu\Delta t)}\right) \\ j_k(t+\Delta t) &= j_{k+1}(t) \qquad\qquad k = 1,2,\ldots,\ell-1 \\ j_\ell(t+\Delta t) &\sim \text{Binomial}\left(F_S(t), \frac{\phi(t)}{\phi(t)+\mu\Delta t}\right) \\ S(t+\Delta t) &= B(t)+S(t)-F_S(t) \end{aligned} \tag{18}$$

Finally, the observation likelihood can be written as follows:

$$i_{t,\text{rep}} \sim \text{NegBin}\left(\rho \sum_{k=1}^{\lceil t_{\text{rep}}/\Delta t\rceil} j_{\ell-k+1}(t), \phi_{\text{obs}}\right). \tag{19}$$

We model the generation-interval distribution using a truncated negative binomial distribution:

$$g(t) \propto t^{G_{\text{shape}}} \times \exp(-t/G_{\text{scale}}), \tag{20}$$

where $g(t)$ is normalized to sum to one. For estimation purposes, we use an alternate parameterization $(G_{\text{mean}}, G_{\text{var}})$ such that $G_{\text{scale}} = G_{\text{var}}/G_{\text{mean}}$ and $G_{\text{shape}} = G_{\text{mean}}/G_{\text{scale}}$. For sufficiently small Δt, G_{mean} and G_{var} corresponds to mean and variance of the generation-interval distribution. Here, we use a simulation time step Δt of one-twentieth of the mean generation interval (equivalently, one-twentieth of the reporting step t_{rep}) and $\ell = 120$; this allows the generation-interval distribution to span over 6 generations.

2.4 Splines

In order to estimate time-varying transmission rates as a smooth function of time, we model transmission rate $\beta(t)$ using a spline (Hooker et al., 2010):

$$\beta(t) = \sum_{i=1}^{k} \phi_i(t) b_i, \tag{21}$$

where each ϕ_i is a cyclic cubic B-spline basis and b_i is the corresponding coefficient. Here, we fix knot locations at every two biweeks (1/13 years); this parameterization results in almost a two-fold reduction in the number of parameters to be estimated compared to the regular TSIR procedure, which tries to estimate a separate transmission rate at every biweek. For regression-based methods (TSIR and gradient matching), we allow the number of knots to be estimated within the regression method by the Restricted Maximum Likelihood (REML) approach (Wood, 2012).

3 Results

3.1 Types of epidemic time series

When reporting step is equal to mean generation time, incidence and prevalence are often assumed to be equivalent (Fig. 1). In fact, we can show that incidence, prevalence, and mortality are nearly equivalent for simple SIR model (1) under Euler approximation when reporting step is equal to mean generation time (see Appendix A.1). Their deterministic dynamics are sufficiently similar that we expect these three reports to be nearly indistinguishable from each other when demographic stochasticity and observation error is introduced. However, this is not true for more complicated models. For a more realistic model with an exposed period (SEIR model), the incidence is comparable to prevalence when reporting time step is equal to mean *infectious period* rather than mean *generation time*. We are not able to see this distinction in the SIR model because it assumes that the mean infectious period is equal mean generation interval. For now, we focus on the SIR model.

Even if incidence, mortality, and prevalence exhibit similar dynamics when reporting step is equal to mean generation time, it is important to distinguish them when we are trying to estimate underlying parameters of the SIR model, especially when mean generation time $1/\gamma$ is not exactly known. To demonstrate this idea, we first simulate an epidemic time series representing prevalence over time by solving the deterministic SIR model numerically and calculating prevalence, $I(t)$, for 20 reporting periods, where reporting period is equal to the mean generation time. Then, we draw a beta-binomial random variable at each time step with a reporting rate of 70% and an overdispersion parameter of 10. Using this time series, we try to estimate four parameters – transmission rate β, recovery rate γ, reporting rate ρ, and an initial number of infected individuals $I(0)$ – by treating the same time series as if it were incidence, mortality, and prevalence report (Fig. 1).

We find that fitting prevalence and mortality curves to the same time series yields almost identical estimates of the transmission rate β and the recovery rate γ (as well as related uncertainty in the estimates) whereas fitting incidence and mortality curves to the same time series yields consistent estimates of the reporting rate ρ. As mortality always behaves like prevalence regardless of reporting period (reported mortality case at time t is approximately equal to $\rho\gamma I(t - t_{\text{rep}})t_{\text{rep}}$), it is possible to generate a mortality curve that matches its corresponding prevalence curve by adjusting the reporting rate of ρ. This explains why estimates of dynamical parameters, β and γ, are similar whether we fit prevalence or mortality. On the other hand, incidence report no longer matches prevalence report when reporting period is not equal to the mean generation time (i.e., when γ is being estimated); therefore, fitting incidence curve yields a different estimate of the dynamical parameters, β and γ, from fitting mortality or prevalence curves. Fitting incidence and mortality curve yields similar estimates of the reporting rate ρ because they both contain equivalent information about the final size of an epidemic: total *true* incidence and total

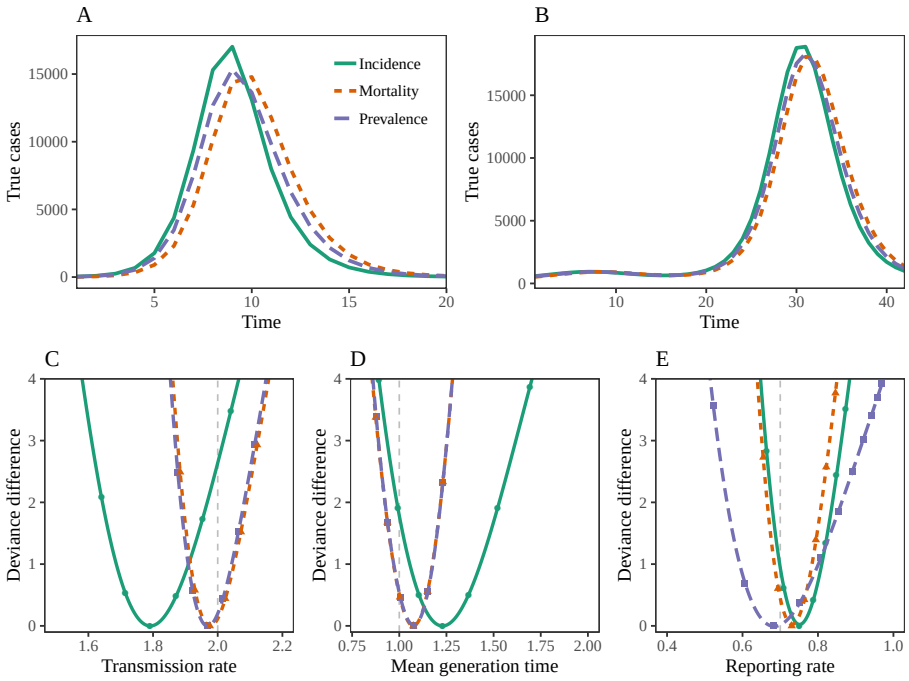

Figure 1: **Comparison of incidence, prevalence, and mortality from dynamical and estimation perspective.** A: deterministic dynamics of incidence, mortality, and prevalence of the SIR model for a single outbreak (no natural birth/death and fixed transmission rate) when mean generation time is equal to reporting step. The parameters and initial conditions are $\beta = 2$, $\gamma = 1$, $N = 1 \times 10^5$, $I(0) = 10$, and $S(0) = N - 10$. B: dynamics of incidence, mortality and prevalence of the SIR model with sinusoidal transmission rate ($\beta(t) = b_0(1 + b_1 \cos(2\pi t/26))$) when mean generation time ie equal to reporting step. The parameters and initial conditions are $b_0 = 500/26$, $b_1 = 0.15$, $\gamma = 1$, $\mu = 1/(50 \times 26)$, $N = 5 \times 10^6$, $I(0) = 0.0001N$, $S(0) = 0.05N$. C, D, E: deviance difference (profile likelihood minus the maximum likelihood) of each parameter when incidence, mortality, and prevalence curves are fit to same time series using trajectory matching. Grey dashed line represents the true value.

true mortality is equal to the final size of an epidemic.

Even when mean generation time is exactly known, we find that estimates of transmission rate β and reporting rate ρ depends on our assumptions about the time series. We find that fitting mortality and prevalence curves give consistent estimates of both β and ρ, whereas fitting incidence curve gives similar but still different estimates (see Appendix A.2). **[SWP:** *TODO*] Any analysis of an epidemic time series should carefully consider what the time series might actually represent.

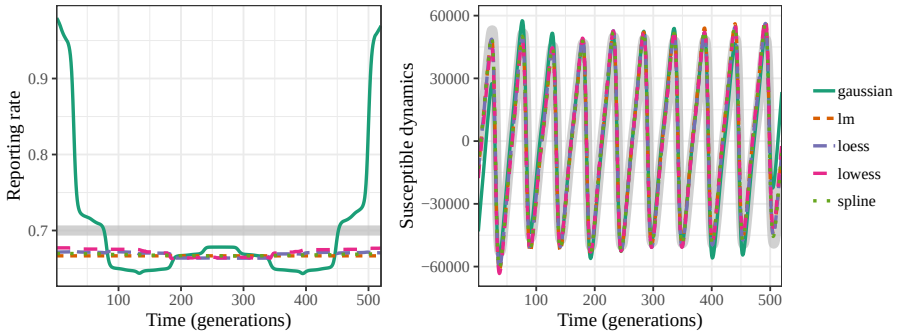

Figure 2: **Sensitivity of estimates of reporting rate and susceptible dynamics to regression methods.** Susceptible reconstruction is performed on a simulated time series (thick grey lines represent the true values). Although estimates of the reporting rates are highly sensitive to regression methods, susceptible dynamics can be robustly estimated.

3.2 Susceptible reconstruction

One of the main assumptions behind the TSIR method is that the dynamics susceptible population S_t is exactly known. Assuming that all individuals that enter the susceptible population become infected eventually, it is possible to recover the reporting rate, ρ_t, as well as the dynamics of the susceptible population as a deviation from its mean, $Z_t = S_t - \bar{S}$, by fitting a regression model between cumulative births and cumulative cases:

$$\sum_{t=1}^{N} B_t = \sum_{t=1}^{N} \frac{C_{t+1}}{\rho_{t+1}} + Z_{t+1} - Z_1, \tag{22}$$

where C_t is the number of observed cases: $C_t = \rho_t i_t$. Then, the reporting rate ρ_t can be estimated from the slope of the regression and the susceptible dynamics can be estimated from the residuals (Finkenstadt and Grenfell, 2000).

While the susceptible reconstruction has been successfully used with the TSIR method, several minor questions remain to be answered. First, how sensitive is our inference to regression methods? Second, does Z_t represent the deviation from the overall population mean $\bar{S}$ or from a long-term moving average (assuming that population changes over time)? Finkenstadt and Grenfell (2000) initially defined Z_t as $S_t - \bar{S}$, but this definition was corrected to $Z_t = S_t - \sigma N_t$, where σ is the mean proportion susceptible, by Dalziel et al. (2016) without justification. Finally, how do changes in population size or reporting rate over time affect our estimate of ρ_{t+1} or Z_t?

We first simulate an epidemic using the continuous deterministic SIR model with sinusoidal transmission rates for 10 years. After introducing observation error base on a beta-binomial distribution, we apply 5 different regression models (Gaussian regression, linear model, locally estimated scatterplot smoothing,

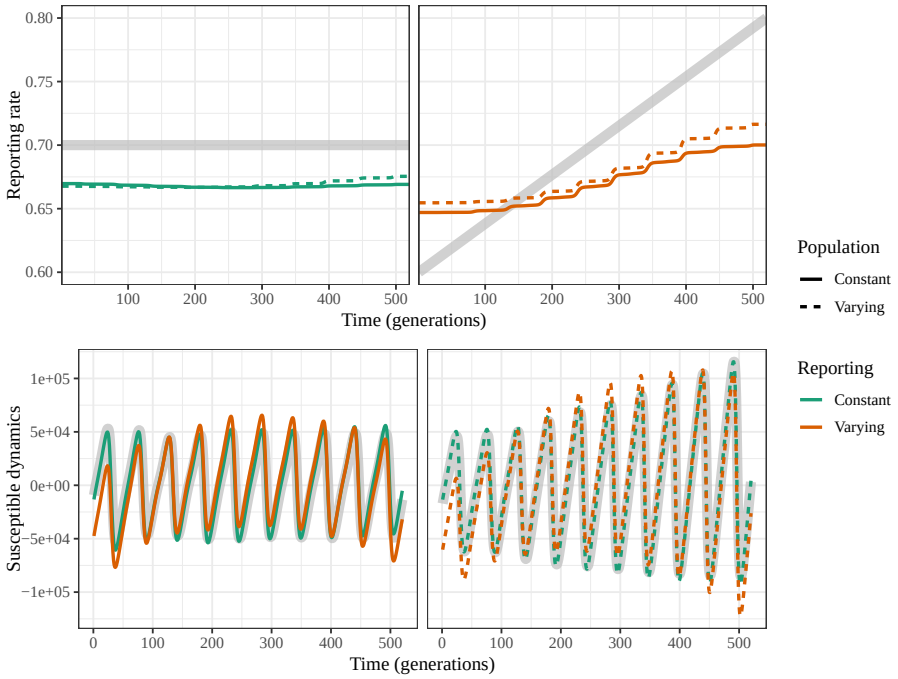

Figure 3: **Estimates of susceptible dynamics and reporting rate under different scenarios.** Susceptible reconstruction is performed on a simulated time series under different scenarios (thick grey lines represent the true values). Spline regression is used throughout these simulations.

locally weighted scatterplot smoothing, and spline model) available from the `tsiR` package to test their ability to infer the reporting rate and susceptible dynamics (Fig. **??**). We find that estimates of reporting rates are highly sensitive to regression methods. The Gaussian regression is particularly bad at estimating the reporting rates near the boundaries. Estimates of susceptible dynamics appear to be robust regardless of the regression method.

We then allow for both population size and reporting to stay constant or increase over time. Then, we compare estimates of susceptible dynamics and reporting rate under all combinations (total of 4) of these scenarios (Fig. 3). We find that estimating time-varying reporting rate is difficult; the increasing pattern in the reporting rate is only partially captured in the slope of the regression model. We also find that the susceptible dynamics Z_t estimated from a regression better matches $S_t - \bar{S}$ rather than $S_t - \sigma N$ (latter results are not shown). In other words, any long-term dynamics in the susceptible population is expected to be captured in Z_t.

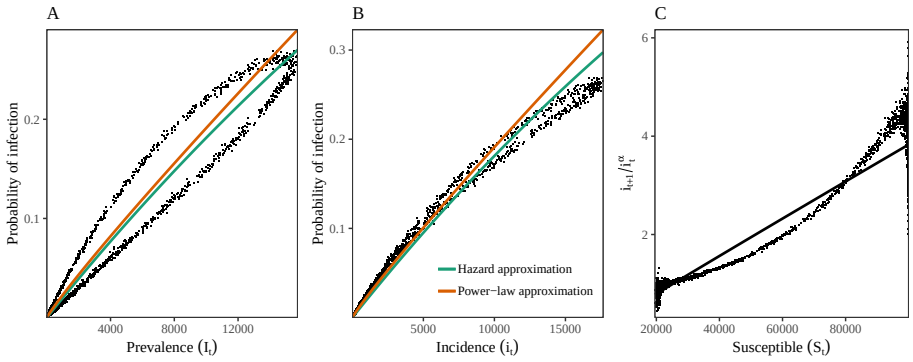

Figure 4: **Density dependence of the infection process.** A: the probability of infection over a generation as a function of prevalence. B: the probability of infection over a generation as a function of incidence. The hazard approximation is calculated using the true transmission rates. The power-law approximation is calculated by fitting the TSIR model to 100 simulations. C: ratio of next incidence i_{t+1} and a power function of current incidence i_t^{α} as a function of the current number of susceptibles S_t; this relationship is expected to be linear under the TSIR framework. Solid black line is the estimated relationship from the TSIR method when we introduce an exponent to S_t: $i_{t+1}/i_t^{0.92} = 4.9S_t^{0.98}$. Points are calculated from 100 Gillespie simulations for single outbreaks.

3.3 Density dependence

In a continuous time SIR model, density dependence in the infection process is implicitly captured in the nonlinear infection term $\beta SI/N$. As the number of infected individuals increases, competition among infected individuals to find a susceptible host also increases. This results in saturation in the probability of infection (Fig. 4A).

The simplest way to account for density dependence in a discrete time model is to approximate the probability of infection a hazard function, $1-\exp(-\phi\Delta t)$. This approximation can be obtained by integrating $dS/dt = -\phi S$ over Δt assuming a constant force of infection $\phi = \beta I/N$. While this approximation captures the *average* probability of infection reasonably well, the *realized* probability of infection deviates from this approximation (Fig. 4A). When an epidemic is growing, the force of infection increases over time, resulting in a higher realized probability of infection. Likewise, when an epidemic is dying out, the force of infection decreases over time, resulting in a lower probability of infection. It is still useful to model probability infection using the hazard function because it provides a probabilistic interpretation of the transition between states, which can be incorporated in stochastic models.

On the other hand, the TSIR model assumes that the probability of infection follows a power function of the current incidence: $\hat{\beta} i_t^{\alpha}/N$, where α is a correction factor for heterogeneity or discretization of a continuous model (Glass et al.,

2003). Glass et al. (2003) further derived an analytical expression for the optimal value of α as a function of birth rate and the basic reproductive number by comparing the second order Taylor expansion of the TSIR model to that of a continuous time model. In the absence of susceptible recruitment to the population, the optimal value of α based on their approximation is unity, which has been regarded as the condition for homogeneous mixing. This is false.

Instead, we propose that α should be interpreted as the strength of density dependence. Smaller values of α capture stronger density dependence and faster saturation of the probability of infection. Under this interpretation, it is clear that α should always be less than unity even in a homogeneously mixing population. Simulations of an SIR model confirms this prediction by demonstrating saturation in the probability of infection as incidence increases (Fig. 4B). Fitting the TSIR model to these simulations further shows that the optimal value for α is less than unity ($\alpha = 0.922$). However, there is a slight mismatch between the estimated probability of infection and the realized probability of infection because the TSIR method effectively tries to minimize the sum of squares on a log scale.

The high estimate of $\hat{\beta} = 3.9$, in comparison to the true basic reproductive number $\mathcal{R}_0 = 2$, suggests that the raw estimate of the transmission rate $\hat{\beta}$ from the TSIR regression should not be taken as an estimate of the reproductive number. Instead, we can translate this estimate to an estimate of the basic reproduction number of a continuous-time SIR model by matching the estimated probability infection from the TSIR model, $\hat{\beta}i^{\alpha}/N$, with the hazard function, $1-\exp(-\mathcal{R}_0 i/N)$, at mean incidence, $\bar{i}$, assuming that incidence and prevalence are equivalent:

$$\mathcal{R}_0 \approx -\frac{N}{\bar{i}} \log\left(1 - \hat{\beta}\bar{i}^{\alpha}/N\right). \tag{23}$$

After applying this correction, the TSIR method estimates $\mathcal{R}_0 = 2.1$, which is much closer to the true value.

We also compare how the effective transmission rate varies with the number of susceptibles (Fig. 4C). Under the TSIR framework, i_{t+1}/i_t^{α} should be a linear function of the number of susceptibles, where the slope is equal to the transmission rate divided by the population size: $\hat{\beta}/N$. However, for a continuous-time model, the ratio i_{t+1}/i_t^{α} decreases at a decelerating rate as the number of susceptible individual decreases due to density dependency. Even when we modify the TSIR model to include an exponent γ on the number of susceptible individuals,

$$\log i_{t+1} = \log\hat{\beta} + \alpha \log i_t + \gamma \log S_t - \log N + \epsilon_t, \tag{24}$$

we are not able to match the patterns of the ratio i_{t+1}/i_t^{α} (Fig. 4C; solid black line).

3.4 Simulation study: constant transmission rate

First, we use 100 Gillespie simulations with constant transmission rate (without natural birth or death) to assess the performance of each method to estimate

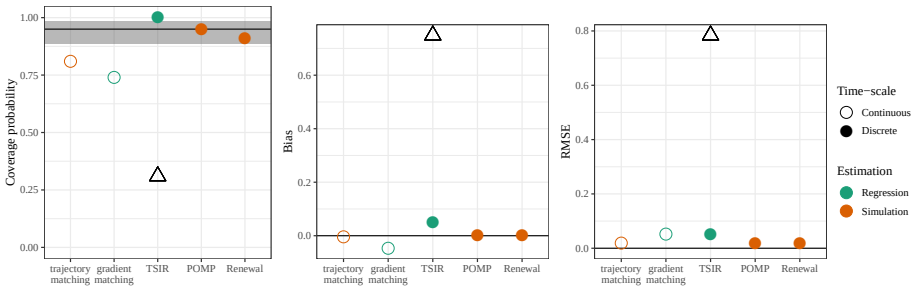

Figure 5: **Comparison of the estimates of a constant transmission rate from different method.** The triangles represent raw TSIR estimates before correcting for the discretization of a continuous-time model. The black lines show nominal coverage, zero bias, and zero RMSE, respectively. The grey bars represent the 95% binomial confidence intervals for 100 simulations.

the transmission rate of β. Trajectory matching and Sequential Monte Carlo estimate the transmission rate β, reporting rate ρ, over-dispersion parameter ϕ_{obs}, and the initial number of infected individuals $I(0)$ (initial number of the susceptible individual is assumed to be $N - I(0)$). The renewal equation estimates an additional parameter G_{var}, which is the approximate variance of the generation-interval distribution. On the other hand, gradient matching and TSIR estimate only the transmission rate β. True susceptible dynamics and reporting are assumed to be known for both methods because they cannot be estimated for a single outbreak with a regression.

Estimates of the transmission rates are summarized into coverage probability, bias, and root mean squared error (RMSE). Coverage probability is the proportion of confidence intervals that contain the true value over repeated experiments. For example, the 95% confidence interval should contain the true value 95% of the time. Bias and RMSE are computed on a log scale.

We find that the Monte Carlo methods (whether the model is represented as a compartmental model or a renewal equation) provide best estimates (Fig. 5). Estimates of the transmission rates based on Sequential Monte Carlo give good coverage and has the least bias and RMSE. However, in order for the renewal equation model to attain nominal coverage, the simulation time step Δt has to be as fine as one-twentieth of the mean generation length; simulating an epidemic at one-tenth of the mean generation length results in 60% coverage. While trajectory matching can fit a continuous-time model without having to rely on a discretization process, it gives a slightly lower coverage than the nominal probability.

Both regression-based methods (gradient matching and TSIR regression) perform similarly in terms of RMSE. Gradient matching and TSIR give slightly biased estimates in opposite directions. Surprisingly, the TSIR regression gives 100% coverage if we assume that we know the exact susceptible dynamics. How-

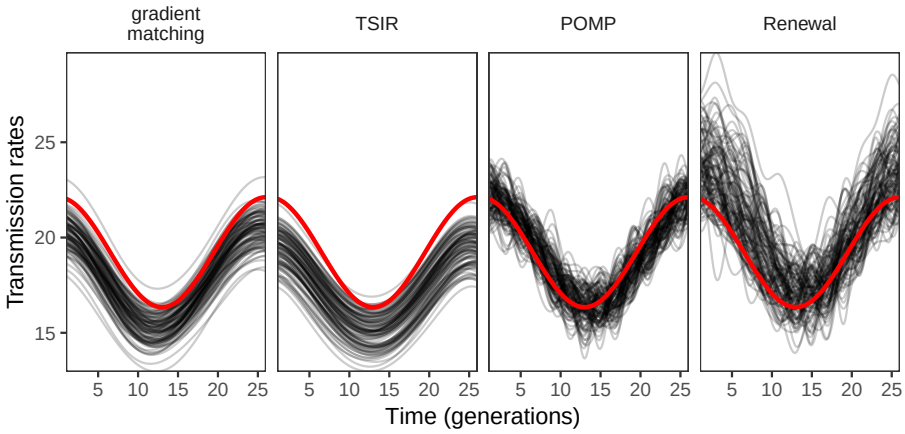

Figure 6: **Comparison of the estimates of time-varying transmission rates from different method.** The red line represents the true transmission rates. The black lines represent the estimated transmission rates from 100 simulations.

ever, the true coverage is likely to be lower because true dynamics of the susceptible population is never exactly known. Low coverage of the gradient matching method is likely driven by not accounting for uncertainty during the initial smoothing process; we do not explore this idea here. Finally, we note the summary statistics for the TSIR method are based on corrected estimates; using raw estimates of the transmission rate obtained from the TSIR regression, without correcting for discretization, gives estimates with strong bias and low coverage.

3.5 Simulation study: estimating time-varying transmission rates

Finally, we use 100 Gillespie simulations with sinusoidal transmission rates to assess the performance of each method to estimate the transmission rate. We excluded trajectory matching from this analysis because fitting a deterministic model to a time series that is driven by highly nonlinear and stochastic system is difficult and not likely to result in useful inference. We find that both regression-based methods are able to recover the true shape of the transmission rate reasonably well but give biased estimates. TSIR method yields a slightly more variable estimates; this is likely to be driven by inaccuracy in the estimates of the reporting rate.

Monte Carlo methods give more variable estimates. Estimates generally follow a sinusoidal shape – transmission rates are higher near the boundaries and lower in the middle – but are 'wigglier'. Fitting the renewal equation gives much more variable estimates of the transmission rates because the renewal equation accounts for extra uncertainty in the system by allowing for the variance of

the generation-interval distribution to vary. These results suggest that small changes in the transmission rate are difficult to identify in the presence of process error. In fact, we find similar patterns when we try to fit the TSIR model without using splines (see Appendix A.3). It is also possible that we may not have enough power to estimate the true shape of the transmission rates from 9 years of data (equivalent to 234 data points assuming biweekly reporting).

4 Discussion

Estimating biologically realistic parameters of a model is central to the analysis of epidemic time series. Here, we compare several statistical methods for estimating the transmission rates of the SIR model for recurrent epidemics. Simple methods that do not make an explicit distinction between different sources of errors (e.g., process and observation error) give reasonable estimates but fail to capture uncertainty associated with their estimates. More sophisticated methods that explicitly account for these errors provide good estimates. However, as we increase the flexibility of the model (e.g., by allowing for the shape of the generation-interval distribution to vary), it becomes more difficult to estimate the true shape of the transmission rates. The TSIR method works reasonably well under several strong assumptions.

4.1 Discretization of a continuous-time model

Fitting a continuous-time stochastic model to an epidemic time series is practically impossible. Many inferential frameworks, instead, rely on discrete-time approximations to estimate the parameters of the corresponding continuous-time model (Finkenstädt et al., 2002; King et al., 2015; Champredon et al., 2018; Li et al., 2018). However, the effects of these discrete-time approximations are rarely discussed. We show that the discretization of a continuous-time model can make a huge difference in the inference. In particular, the renewal equation has low coverage ($\approx 60\%$) when the simulation time step is one-tenth of the mean generation time.

The low coverage is likely to be driven by changes in the underlying generation-interval distribution. As the discretization time step changes, mean and variance of the distribution also changes even when the parameters of the distributions are held constant. For our example, changing the time step from one-twentieth of the mean generation time to one-tenth increases the mean generation time by 2.5%. This increase in the mean generation time is sufficient to introduce a small bias in our estimates and reduce coverage (Wallinga and Lipsitch, 2006). Using a coarser time scale is likely to increase bias and decrease coverage. This bias is not observed when we fit discrete-time compartmental SIR model via POMP because we adjust the rate parameter to match the mean generation time of the discrete distribution and the mean generation time of the continuous distribution (He et al., 2009).

These results suggest that discretization of a continuous-time model should

be done carefully. In particular, we prefer to separate the simulation time step from the reporting time step.

4.2 Generation-interval distributions

We expect a realistic generation-interval distribution of measles to be narrower than the exponential distribution but wider than the fixed distribution. We try to bridge these two extreme assumptions by using the renewal equation. However, the renewal equation does not perform better because we test our methods on Gillespie simulations of the SIR model, which assumes an exponentially decreasing generation-time distribution. While we might expect the renewal equation to perform better against real data, simulating the renewal equation with various shapes suggest that the shape of the generation-interval distribution has little effect on the overall dynamics (see Appendix A.4). [**SWP:** *TODO*] These results are consistent with several previous studies that reported that using non-exponential distributions made small differences to their conclusion (He et al., 2009). [**SWP:** *Confirm*]. Based on these results, it is not clear whether it is worth estimating the shape of the generation-interval distribution at the expense of power to estimate the transmission rates.

4.3 Process and observation error

Our study suggests that the assumptions about the amount of process and observation error affect the uncertainty of our inference. For weakly nonlinear systems with large observation error, not making this distinction may not have large effects on the inference (Ma et al., 2014). However, for highly nonlinear and stochastic systems, such as measles, assuming that there is only observation error leads to over-confidence (King et al., 2015; Taylor et al., 2016). On the other hand, assuming that there is only process error could lead to under-confidence ([**SWP:** *Fig?*]). Although ignoring observation error may not be as bad as ignoring process error (see Appendix A.5) [**SWP:** *TODO*], any formal analyses should consider different sources of errors explicitly.

5 Conclusions

We used the simple Susceptible-Infected-Recovered (SIR) model to compare some of the widely-used statistical methods for estimating transmission rates. In line with previous studies (King et al., 2015; Taylor et al., 2016), we argue again that fitting deterministic models should be avoided. While regression methods, such as gradient matching or TSIR method, can be useful for preliminary analyses, they rely on several assumptions which may be inappropriate. Any formal inference should be carried out using Monte Carlo methods that allow us to model all sources of variations separately.

Finally, we suggest using simulations to test mathematical and statistical models of interest before analyzing real epidemic time series. Often, the state-

of-art methods such as POMP (King et al., 2015), are taken for granted; using simulations allows us to assess statistical power of our model and identify which of the model assumptions are likely to have large effect on the inference. In order to estimate the transmission rates of measles, the shape of the generation-interval distribution [**SWP:** *Come back*]

References

Althaus, C. L. (2014). Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa. *PLoS currents 6*.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 269–342.

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing 50*(2), 174–188.

Baker, R. E., A. S. Mahmud, and C. J. E. Metcalf (2018). Dynamic response of airborne infections to climate change: predictions for varicella. *Climatic Change 148*, 547–560.

Bjørnstad, O. N., B. F. Finkenstädt, and B. T. Grenfell (2002). Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs 72*(2), 169–184.

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

Bretó, C., D. He, E. L. Ionides, A. A. King, et al. (2009). Time series analysis via mechanistic models. *The Annals of Applied Statistics 3*(1), 319–348.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of statistical software 76*(1).

Cauchemez, S. and N. M. Ferguson (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *Journal of the Royal Society Interface 5*(25), 885–897.

Champredon, D., J. Dushoff, and D. Earn (2018). Equivalence of the Erlang SEIR epidemic model and the renewal equation. *BioRxiv*, 319574.

Champredon, D., M. Li, B. M. Bolker, and J. Dushoff (2018). Two approaches to forecast Ebola synthetic epidemics. *Epidemics 22*, 36–42.

Chowell, G., N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore, and J. M. Hyman (2004). The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of theoretical biology 229*(1), 119–126.

Dalziel, B. D., O. N. Bjørnstad, W. G. van Panhuis, D. S. Burke, C. J. E. Metcalf, and B. T. Grenfell (2016). Persistent chaos of measles epidemics in the prevaccination United States caused by a small change in seasonal transmission patterns. *PLoS computational biology 12*(2), e1004655.

de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics 26*(2), 403–413.

Didelot, X., L. K. Whittles, and I. Hall (2017). Model-based analysis of an outbreak of bubonic plague in Cairo in 1801. *Journal of The Royal Society Interface 14*(131), 20170160.

Doucet, A., N. De Freitas, and N. Gordon (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pp. 3–14. Springer.

Du, Z., W. Zhang, D. Zhang, S. Yu, and Y. Hao (2017). Estimating the basic reproduction rate of HFMD using the time series SIR model in Guangdong, China. *PloS one 12*(7), e0179623.

Earn, D. J., P. Rohani, B. M. Bolker, and B. T. Grenfell (2000). A simple model for complex dynamical transitions in epidemics. *Science 287*(5453), 667–670.

Earn, D. J., P. Rohani, and B. T. Grenfell (1998). Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society of London. Series B: Biological Sciences 265*(1390), 7–10.

Ellner, S. P., Y. Seifu, and R. H. Smith (2002). Fitting population dynamic models to time-series data by gradient matching. *Ecology 83*(8), 2256–2270.

Fasiolo, M., N. Pya, S. N. Wood, et al. (2016). A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology. *Statistical Science 31*(1), 96–118.

Ferrari, M. J., A. Djibo, R. F. Grais, N. Bharti, B. T. Grenfell, and O. N. Bjornstad (2010). Rural–urban gradient in seasonal forcing of measles transmission in Niger. *Proceedings of the Royal Society B: Biological Sciences 277*(1695), 2775–2782.

Fine, P. E. and J. A. Clarkson (1982). Measles in England and Wales—I: an analysis of factors underlying seasonal patterns. *International journal of epidemiology 11*(1), 5–14.

Finkenstädt, B. F., O. N. Bjørnstad, and B. T. Grenfell (2002). A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. *Biostatistics 3*(4), 493–510.

Finkenstadt, B. F. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 49*(2), 187–205.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics 22*(4), 403–434.

Glass, K., Y. Xia, and B. Grenfell (2003). Interpreting time-series analyses for continuous-time biological models—measles as a case study. *Journal of theoretical biology 223*(1), 19–25.

Grenfell, B. T., O. N. BJøRNSTAD, and B. F. Finkenstädt (2002). Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. *Ecological Monographs 72*(2), 185–202.

Grenfell, B. T., A. Kleckzkowski, S. Ellner, and B. Bolker (1994). Measles as a case study in nonlinear forecasting and chaos. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences 348*(1688), 515–530.

He, D., J. Dushoff, T. Day, J. Ma, and D. J. Earn (2011). Mechanistic modelling of the three waves of the 1918 influenza pandemic. *Theoretical Ecology 4*(2), 283–288.

He, D., E. L. Ionides, and A. A. King (2009). Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*.

Hempel, K. and D. J. Earn (2015). A century of transitions in New York City's measles dynamics. *Journal of The Royal Society Interface 12*(106), 20150024.

Hooker, G., S. P. Ellner, L. D. V. Roditi, and D. J. Earn (2010). Parameterizing state–space models for infectious disease dynamics by generalized profiling: measles in Ontario. *Journal of The Royal Society Interface 8*(60), 961–974.

Ionides, E. L., A. Bhadra, Y. Atchadé, A. King, et al. (2011). Iterated filtering. *The Annals of Statistics 39*(3), 1776–1802.

Ionides, E. L., C. Bretó, and A. A. King (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences 103*(49), 18438–18443.

Ionides, E. L., D. Nguyen, Y. Atchadé, S. Stoev, and A. A. King (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences 112*(3), 719–724.

Jackson, C., P. Mangtani, P. Fine, and E. Vynnycky (2014). The effects of school holidays on transmission of varicella zoster virus, England and Wales, 1967–2008. *PloS one 9*(6), e99762.

Kermack, W. O. and A. G. McKendrick (1927). 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.

King, A. A., M. Domenech de Cellès, F. M. Magpantay, and P. Rohani (2015). Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society B: Biological Sciences 282*(1806), 20150347.

King, A. A., D. Nguyen, and E. L. Ionides (2015). Statistical inference for partially observed markov processes via the r package pomp. *arXiv preprint arXiv:1509.00503*.

Koelle, K. and M. Pascual (2004). Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *The American Naturalist 163*(6), 901–913.

Krylova, O. and D. J. Earn (2013). Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *Journal of The Royal Society Interface 10*(84), 20130098.

Levins, R. (1966). The strategy of model building in population biology. *American scientist 54*(4), 421–431.

Li, M., J. Dushoff, and B. M. Bolker (2018). Fitting mechanistic epidemic models to data: a comparison of simple Markov chain Monte Carlo approaches. *Statistical methods in medical research 27*(7), 1956–1967.

Liu, W., S. A. Levin, and Y. Iwasa (1986). Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. *Journal of mathematical biology 23*(2), 187–204.

Ma, J., J. Dushoff, B. M. Bolker, and D. J. Earn (2014). Estimating initial epidemic growth rates. *Bulletin of mathematical biology 76*(1), 245–260.

Mahmud, A., N. Alam, and C. Metcalf (2017). Drivers of measles mortality: the historic fatality burden of famine in Bangladesh. *Epidemiology & Infection 145*(16), 3361–3369.

Mantilla-Beniers, N., O. Bjørnstad, B. Grenfell, and P. Rohani (2009). Decreasing stochasticity through enhanced seasonality in measles epidemics. *Journal of The Royal Society Interface 7*(46), 727–739.

Metcalf, C., O. Bjørnstad, M. Ferrari, P. Klepac, N. Bharti, H. Lopez-Gatell, and B. Grenfell (2011). The epidemiology of rubella in Mexico: seasonality, stochasticity and regional variation. *Epidemiology & Infection 139*(7), 1029–1038.

Metcalf, C., C. Cohen, J. Lessler, J. McAnerney, G. Ntshoe, A. Puren, P. Klepac, A. Tatem, B. Grenfell, and O. Bjørnstad (2013). Implications of spatially heterogeneous vaccination coverage for the risk of congenital rubella syndrome in South Africa. *Journal of the Royal Society Interface 10*(78), 20120756.

Metcalf, C., C. Munayco, G. Chowell, B. Grenfell, and O. Bjørnstad (2010). Rubella metapopulation dynamics and importance of spatial coupling to the risk of congenital rubella syndrome in Peru. *Journal of the Royal Society Interface 8*(56), 369–376.

Metcalf, C. J. E., O. N. Bjørnstad, B. T. Grenfell, and V. Andreasen (2009). Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen. *Proceedings of the Royal Society B: Biological Sciences 276*(1676), 4111–4118.

Morris, W. F. (1997). Disentangling effects of induced plant defenses and food quantity on herbivores by fitting nonlinear models. *The American Naturalist 150*(3), 299–327.

Pascual, M., B. Cazelles, M. Bouma, L. Chaves, and K. Koelle (2007). Shifting patterns: malaria dynamics and rainfall variability in an African highland. *Proceedings of the Royal Society B: Biological Sciences 275*(1631), 123–132.

Pascual, M., L. Chaves, B. Cash, X. Rodó, and M. Yunus (2008). Predicting endemic cholera: the role of climate variability and disease dynamics. *Climate Research 36*(2), 131–140.

Perkins, T. A., C. J. E. Metcalf, B. T. Grenfell, and A. J. Tatem (2015). Estimating drivers of autochthonous transmission of chikungunya virus in its invasion of the Americas. *PLoS currents 7*.

Plummer, M. et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, Volume 124. Vienna, Austria.

Riley, S., C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L.-M. Ho, T.-H. Lam, T. Q. Thach, et al. (2003). Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science 300*(5627), 1961–1966.

Takahashi, S., Q. Liao, T. P. Van Boeckel, W. Xing, J. Sun, V. Y. Hsiao, C. J. E. Metcalf, Z. Chang, F. Liu, J. Zhang, et al. (2016). Hand, foot, and mouth disease in China: modeling epidemic dynamics of enterovirus serotypes and implications for vaccination. *PLoS medicine 13*(2), e1001958.

Takahashi, S., C. J. E. Metcalf, Y. Arima, T. Fujimoto, H. Shimizu, H. Rogier van Doorn, T. Le Van, Y.-F. Chan, J. J. Farrar, K. Oishi, et al. (2018). Epidemic dynamics, interactions and predictability of enteroviruses associated with hand, foot and mouth disease in Japan. *Journal of The Royal Society Interface 15*(146), 20180507.

Taylor, B. P., J. Dushoff, and J. S. Weitz (2016). Stochasticity and the limits to confidence when estimating R0 of Ebola and other emerging infectious diseases. *Journal of theoretical biology 408*, 145–154.

Van Boeckel, T. P., S. Takahashi, Q. Liao, W. Xing, S. Lai, V. Hsiao, F. Liu, Y. Zheng, Z. Chang, C. Yuan, et al. (2016). Hand, foot, and mouth disease in China: critical community size and spatial vaccination strategies. *Scientific reports 6*, 25248.

Wallinga, J. and M. Lipsitch (2006). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences 274*(1609), 599–604.

Wood, S. (2012). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation.

Xia, Y., H. Tong, et al. (2011). Feature matching in time series modeling. *Statistical Science 26*(1), 21–46.

A Appendix

A.1 Relating incidence, mortality, and prevalence

First, recall that mortality case can be written as

$$\int_{t-t_{\text{rep}}}^{t} \gamma I ds. \tag{25}$$

When reporting time step is equal to mean generation time, i.e., $t_{\text{rep}} = 1/\gamma$, it follows that:

$$\int_{t-t_{\text{rep}}}^{t} \gamma I ds \approx \gamma I(t - t_{\text{rep}}) t_{\text{rep}} = I(t - t_{\text{rep}}) \tag{26}$$

Likewise, we can derive a similar relation for incidence under the same condition:

$$\begin{aligned} I(t + t_{\text{rep}}) &= I(t) + \int_{t}^{t+t_{\text{rep}}} \frac{dI(s)}{ds} ds \\ &= I(t) + \int_{t}^{t+t_{\text{rep}}} \beta(s) S \frac{I}{N} ds - \int_{t}^{t+t_{\text{rep}}} \gamma I ds \\ &\approx \int_{t}^{t+t_{\text{rep}}} \beta(s) S \frac{I}{N} ds = i(t + t_{\text{rep}}) \end{aligned} \tag{27}$$

Here, we assumed that individuals leaving infected class I from natural mortality is negligible over the reporting period.

A.2 Fitting incidence, mortality, and prevalence when mean generation time is known

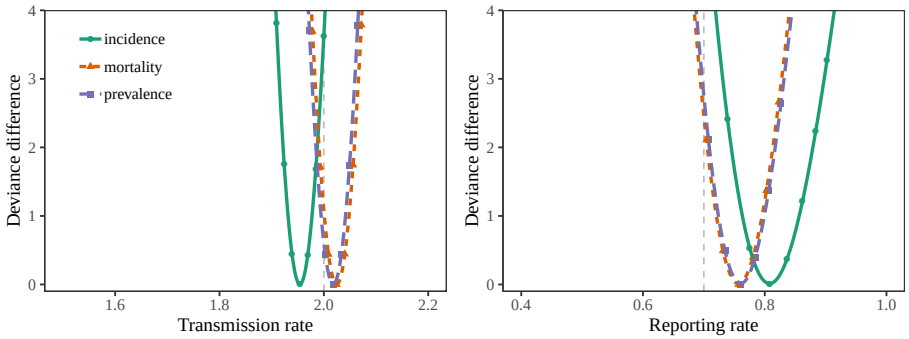

Figure A.2: **Comparison of fitting incidence, prevalence, and mortality curves when mean generation time is known.** Deviance difference (profile likelihood minus the maximum likelihood) of each parameter when incidence, mortality, and prevalence curves are fit to same time series using trajectory matching. We assume that the mean generation time is known. Grey dashed line represents the true value.

A.3 Fitting TSIR model without cyclic splines

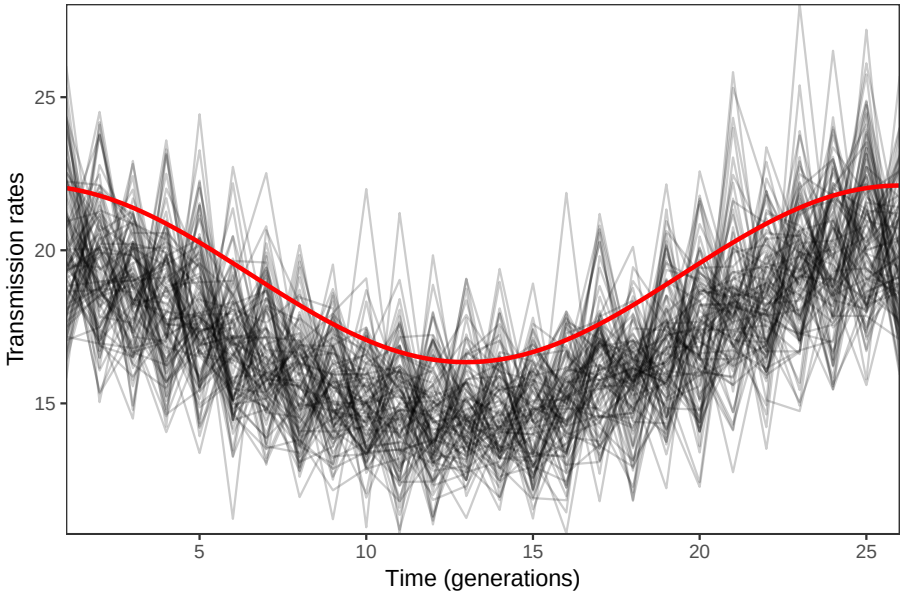

Figure A.3: **Estimates of time-varying transmission rates using TSIR without cyclic basis.** The red line represents the true transmission rates. The black lines represent the estimated transmission rates from 100 simulations.

A.4 The effects of the shape of generation-interval distributions on measles dynamics

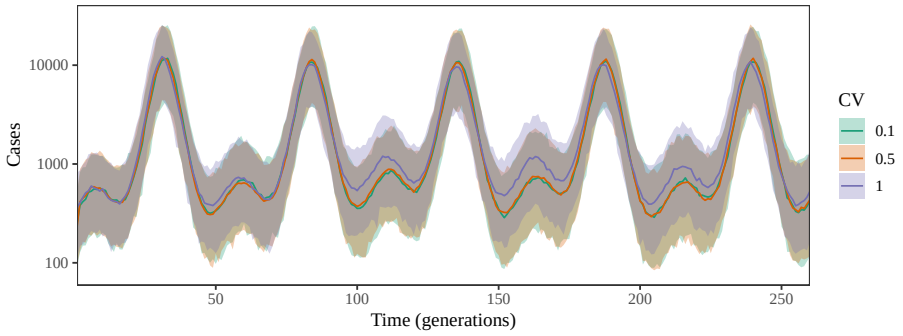

Figure A.4: **Effects of the shape of the generation-interval distribution on the dynamics of recurrent epidemics.** The stochastic renewal equation, represented as a Markov process, is simulated with measles-like parameters by changing coefficient variation (CV) of the generation-interval distribution. Ribbons show 95% quantiles of the simulated incidence report. Lines show the median of the simulated incidence report.

A.5 Process and observation error

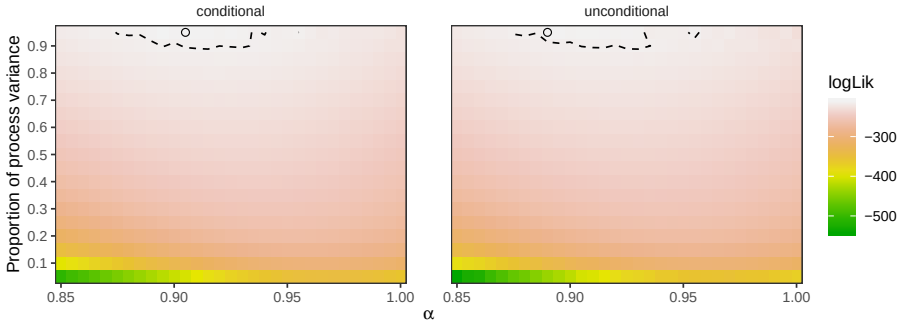

Figure A.5: **Likelihood (slice) surface for density dependence parameter α and proportion of process variance.** For each α, we fit a TSIR model to measles time series from Boston. We decompose the residual variance into observation variance and process variance. We then use particle filtering to estimate the likelihood of the TSIR model using the transmission rates inferred from the TSIR regression. Both the infection process and the measurement models are assumed to follow a lognormal distribution, i.e., $\log i_{t+1} \sim \text{Normal}(\log(\beta_t S_t i_t^\alpha / N_i), r\sigma^2)$ and $\log i_{t,\text{rep}} \sim \text{Normal}(\log(i_t/\rho), (1-r)\sigma^2)$, where σ^2 is the residual variance and r is the proportion of process variance. Conditional simulations assume that the susceptible dynamics is known (estimated via susceptible reconstruction). Unconditinal simulations treat susceptible dynamics as a latent variable: $S_{t+1} = B_t + S_t - i_{t+1}$. The circle represents the maximum likelihood estimate. Dashed lines represent the 95% confidence region.