Bulletin of Mathematical Biology

Fitting epidemic models to data – a tutorial in memory of Fred Brauer

--Manuscript Draft--

Manuscript Number:	BMAB-D-24-00128R1
Full Title:	Fitting epidemic models to data – a tutorial in memory of Fred Brauer
Article Type:	Education
Keywords:	ordinary differential equations parameter estimation
Corresponding Author:	David J. D. Earn, Ph.D. McMaster University CANADA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	McMaster University
Corresponding Author's Secondary Institution:	
First Author:	David J. D. Earn, Ph.D.
First Author Secondary Information:	
Order of Authors:	David J. D. Earn, Ph.D.
	Sang Woo Park, B.Sc.
	Benjamin M. Bolker, Ph.D.
Order of Authors Secondary Information:	
Funding Information:	NSERC — Prof. David J. D. Earn Prof. Benjamin M. Bolker

Abstract:

Fred Brauer was an eminent mathematician who studied dynamical systems, especially differential equations. He made many contributions to mathematical epidemiology, a field that is strongly connected to data, but he always chose to avoid data analysis. Nevertheless, he recognized that fitting models to data is usually necessary when attempting to apply infectious disease transmission models to real public health problems. He was curious to know how one goes about fitting dynamical models to data, and why it can be hard. Initially in response to Fred's questions, we developed a user-friendly \code{R} package, \code{fitode}, that facilitates fitting ordinary differential equations to observed time series. Here, we use this package to provide a brief tutorial introduction to fitting compartmental epidemic models to a single observed time series. We assume that, like Fred, the reader is familiar with dynamical systems from a mathematical perspective, but has limited experience with statistical methodology or optimization techniques.

Figure 1, top panel

Click here to access/download;Figure;Bombay-figure-1.pdf

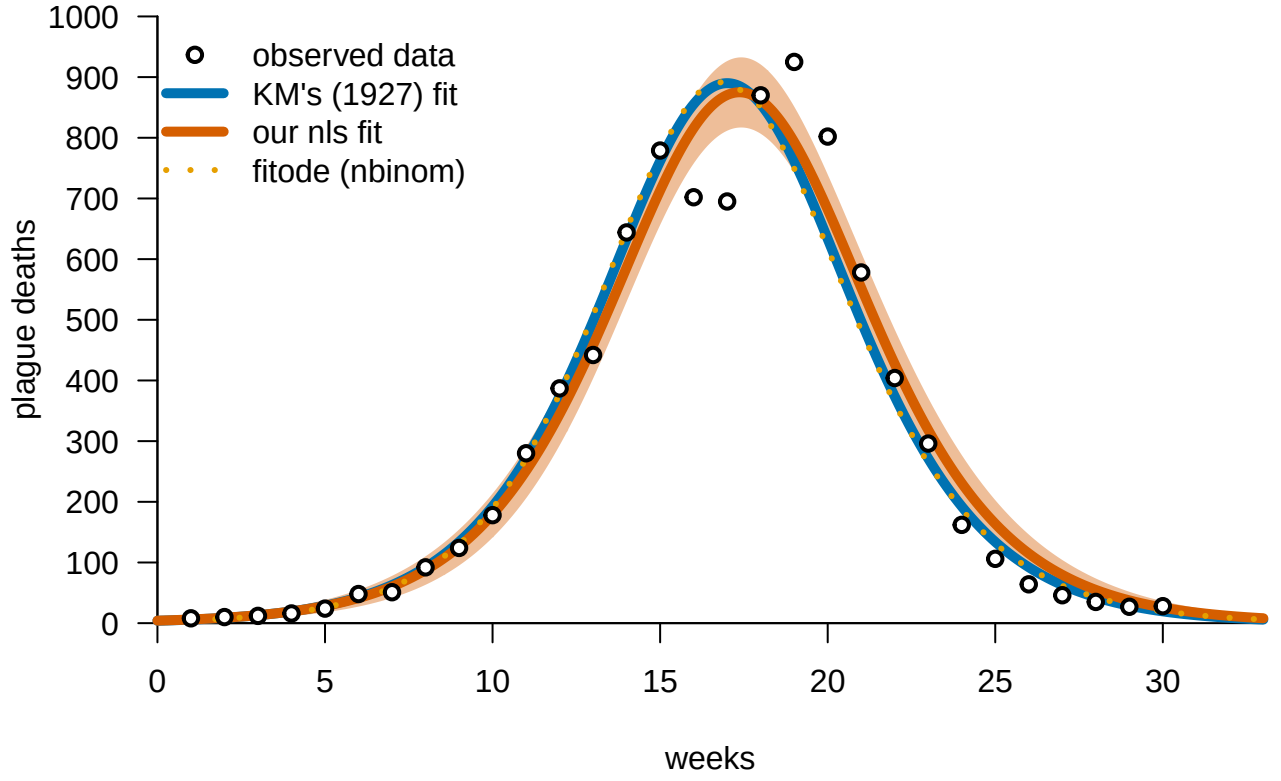


Figure 1, bottom panel

Click here to access/download;Figure;Bombay-figure-2.pdf

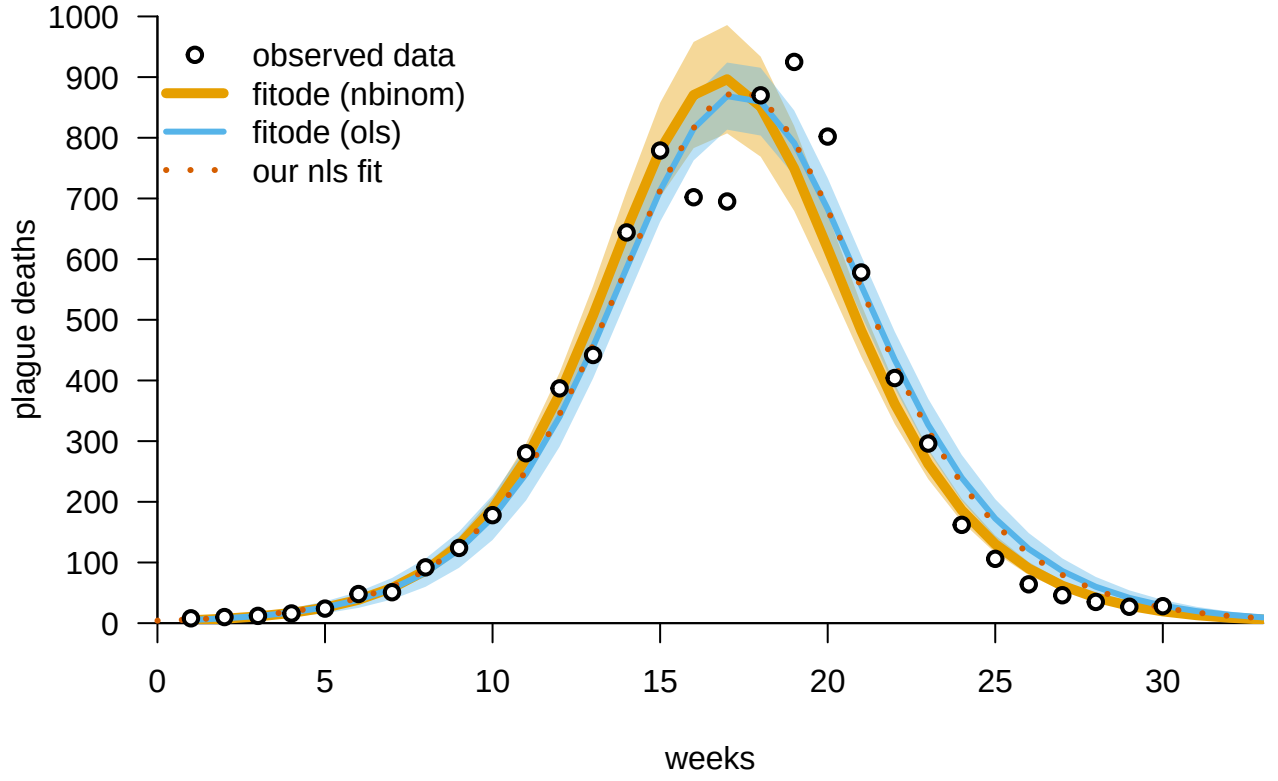


Figure 2

Click here to access/download;Figure;phila-figure-1.pdf

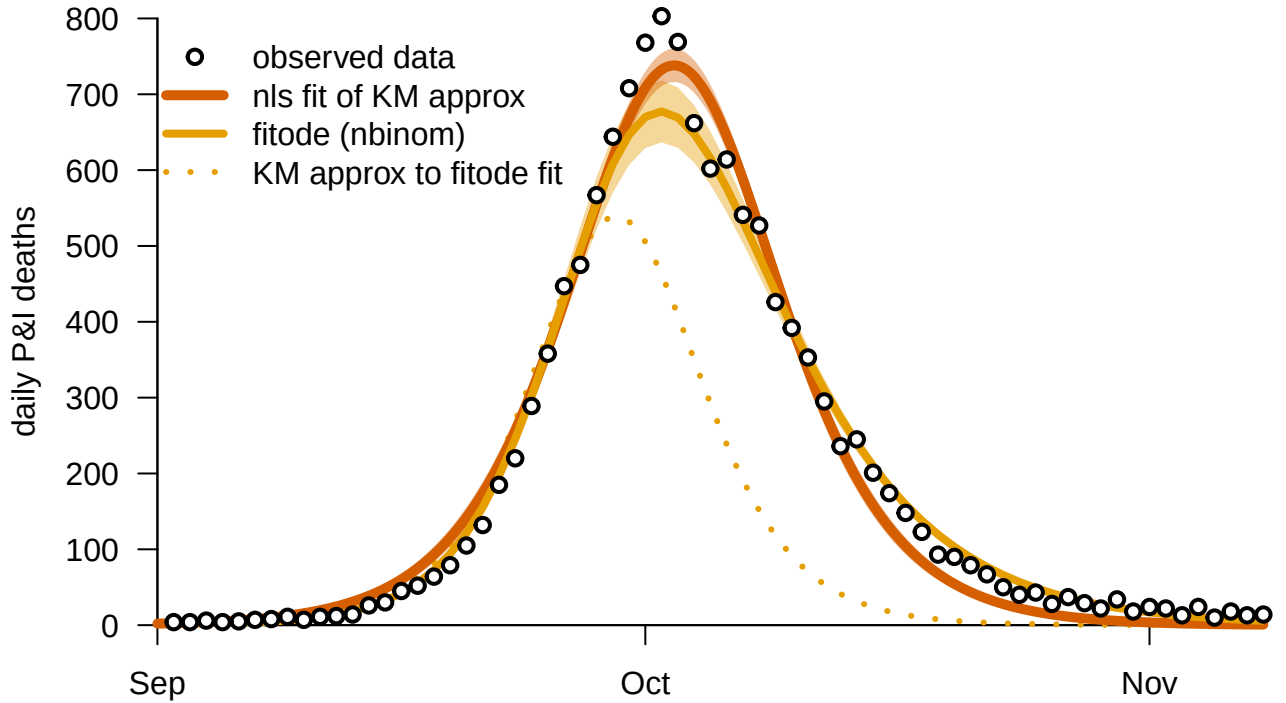


Figure 3, top panel

Click here to access/download;Figure;plot_stochastic_SIR-1.pdf

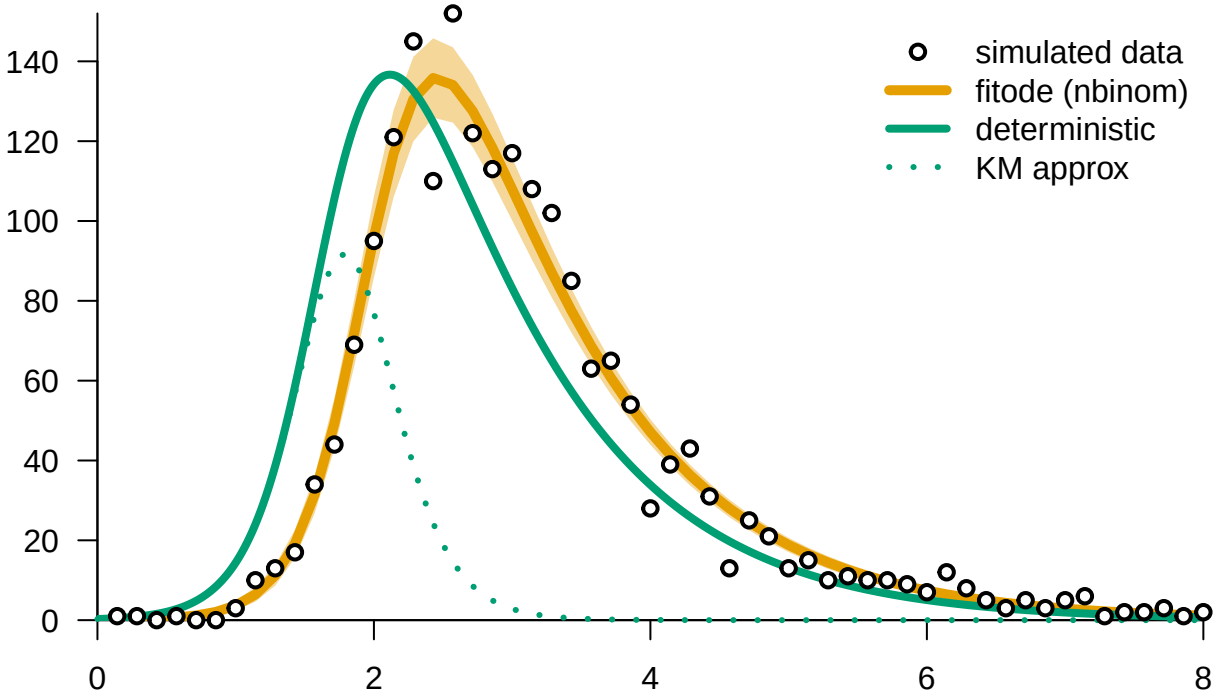


Figure 3, bottom panel

Click here to access/download;Figure;plot_stochastic_SIR-2.pdf

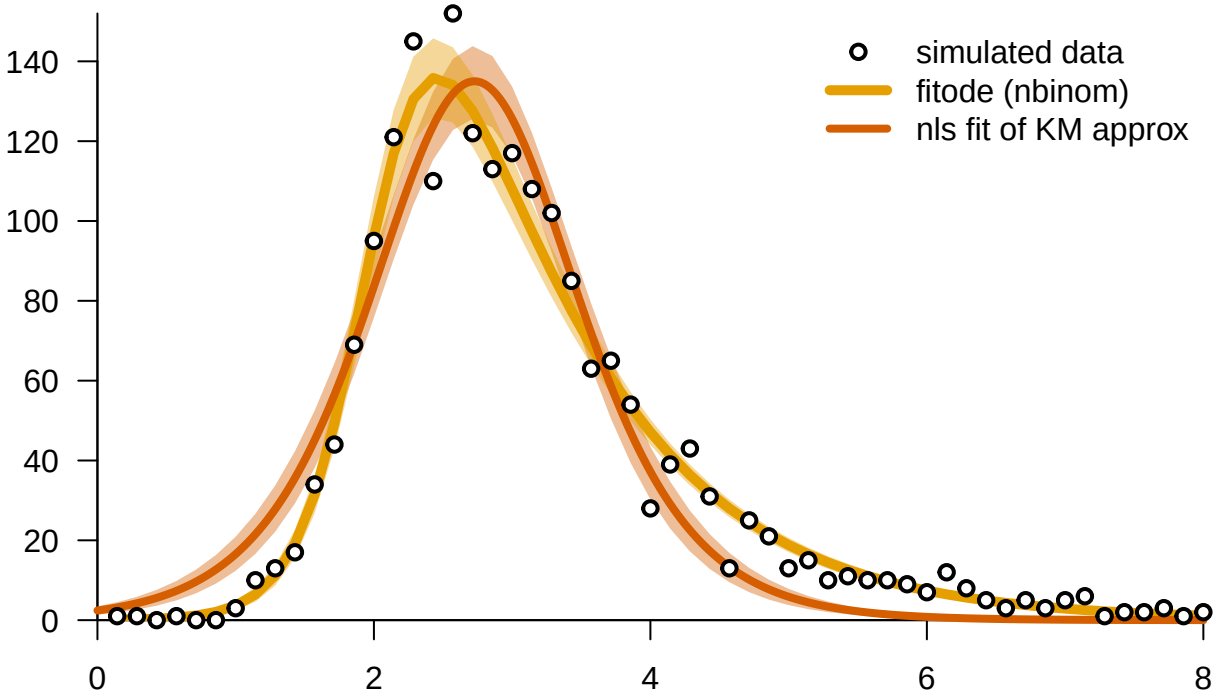
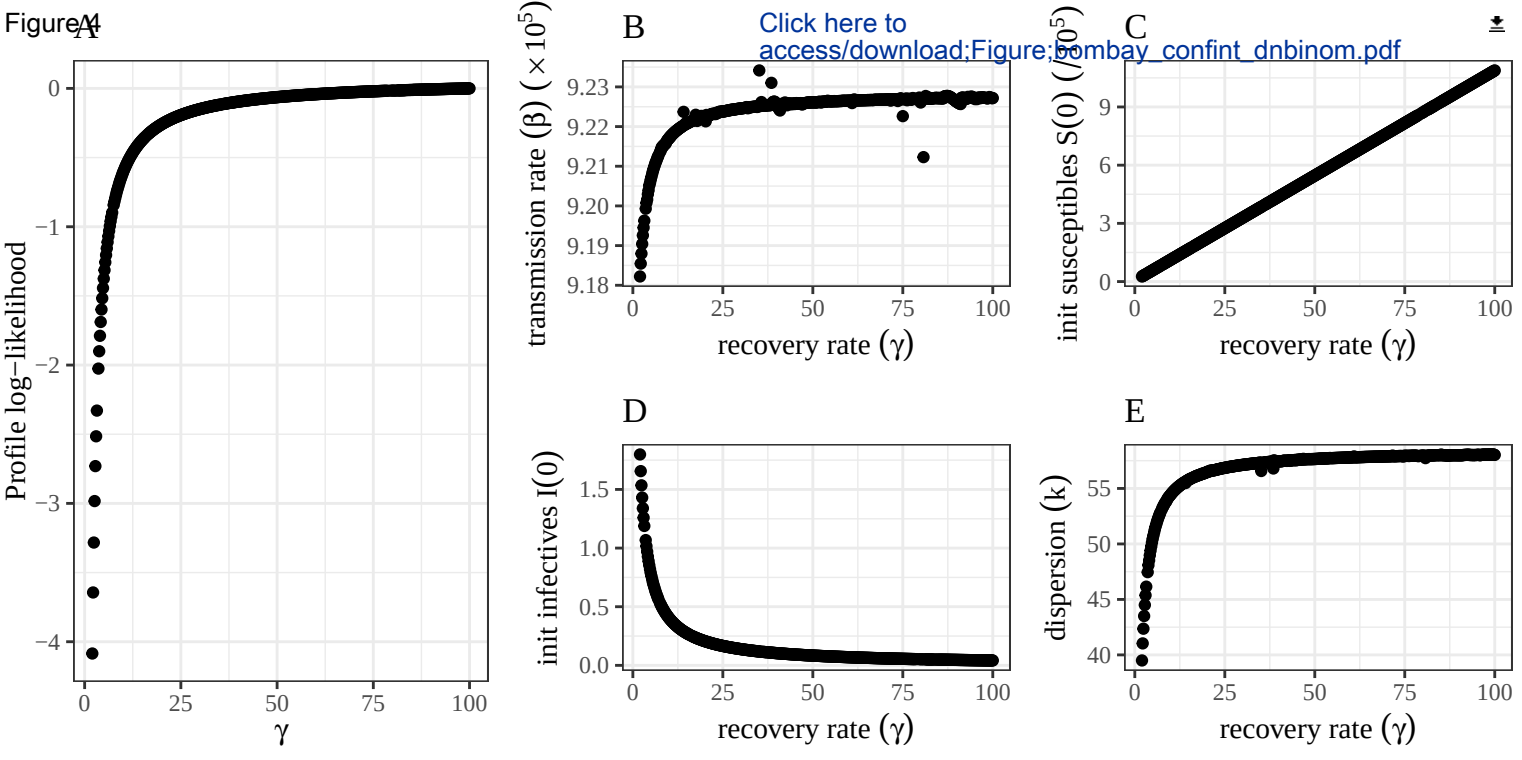


Figure 4



ARTICLE TYPE: EDUCATION

Fitting epidemic models to data – a tutorial in memory of Fred Brauer

David J. D. Earn[a], Sang Woo Park[b] and Benjamin M. Bolker[a,c],

[a]Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada, L8S 4K1;

[b]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544;

[c]Department of Biology, McMaster University, Hamilton, Ontario, Canada, L8S 4K1

Article compiled:
June 4, 2024

ABSTRACT

Fred Brauer was an eminent mathematician who studied dynamical systems, especially differential equations. He made many contributions to mathematical epidemiology, a field that is strongly connected to data, but he always chose to avoid data analysis. Nevertheless, he recognized that fitting models to data is usually necessary when attempting to apply infectious disease transmission models to real public health problems. He was curious to know how one goes about fitting dynamical models to data, and why it can be hard. Initially in response to Fred's questions, we developed a user-friendly `R` package, `fitode`, that facilitates fitting ordinary differential equations to observed time series. Here, we use this package to provide a brief tutorial introduction to fitting compartmental epidemic models to a single observed time series. We assume that, like Fred, the reader is familiar with dynamical systems from a mathematical perspective, but has limited experience with statistical methodology or optimization techniques.

KEYWORDS

epidemic models; infectious diseases; ordinary differential equations; parameter estimation; maximum likelihood; `fitode`

Submitted to *Bulletin of Mathematical Biology*

D.J.D. Earn. earn@math.mcmaster.ca. ORCID: 0000-0002-7562-1341
S.W. Park. swp2@princeton.edu. ORCID 0000-0003-2202-3361,
B.M. Bolker. bolkerb@mcmaster.ca. ORCID: 0000-0002-2127-0443

1. Introduction

In their landmark 1927 paper, Kermack and McKendrick [1927, p. 713] (KM) introduced the now-standard susceptible-infected-removed (SIR) epidemic model,

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI\,, \tag{1a}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI - \gamma I\,, \tag{1b}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I\,, \tag{1c}$$

where S, I and R represent the numbers of individuals who are susceptible, infected or removed[1], β is the transmission rate, and γ is the removal (or recovery) rate. In that original paper, KM [p. 714] also fit their model to plague mortality data from an epidemic in Bombay (now Mumbai) that occurred about 20 years before their paper was written.

In the century that has elapsed since publication of KM's initial paper, the field of mathematical epidemiology has expanded and matured, and has been the subject of many books [Bartlett, 1960, Bailey, 1975, Anderson and May, 1991, Andersson and Britton, 2000, Diekmann and Heesterbeek, 2000, Brauer and Castillo-Chavez, 2001, Brauer et al., 2019] and review articles [Hethcote, 2000, Earn et al., 2002, Earn, 2008, 2009]. Researchers have primarily focused on ***compartmental models*** like the SIR model, cast either as differential equations following the tradition of KM, or as ***stochastic processes*** in the tradition of McKendrick [1926] and Bartlett [1960]. In recent years, as the power of computers has grown, some researchers have turned to ***agent-based models***, which represent each individual as a separate unit that can have unique properties [Eubank et al., 2004].

Throughout the history of the subject, and regardless of the modelling frameworks they have exploited, mathematical epidemiologists have frequently attempted to fit—or at least to compare—their models to observed infectious disease data. Such fits have often been naïve, with limited consideration of their quality. Over the years, however, there has been a trend towards greater sophistication and statistical rigour in parameter estimation for infectious disease models; books that explain these methods have begun to appear in recent decades [Bolker, 2008, Bjørnstad, 2018]. Careful consideration of uncertainty is especially important when epidemic models are used for the development and analysis of policy options for infectious disease management [Elderd et al., 2006], a challenge that began to absorb the attention of many mathematical epidemiologists as soon as the emergence of SARS-CoV-2 ignited the COVID-19 pandemic [Brooks-Pollock et al., 2021, Hillmer et al., 2021, Nixon et al., 2022, Howerton et al., 2023].

While visiting the University of British Columbia in 2014–2015, one of us (DE) had many conversations with Fred Brauer about epidemic models and how they can be used in practical applications. While he had no desire to analyze data himself, Fred was acutely aware that fitting to data is essential if one wishes to apply epidemic models to real public health problems, and he did want to understand what was involved in doing so.

Fred's curiosity inspired us to develop user-friendly software for fitting ordinary dif-

[1] In the words of KM [p. 701], "removed from the number of those who are sick, by recovery or by death".

ferential equation (ODE) models to observed time series, with the goal of illustrating the process and challenges of model fitting to Fred and others like him, i.e., individuals who are comfortable with mathematical analysis of ODEs but have little or no experience with statistics and parameter estimation. Unfortunately, we have lost the opportunity to present our work to Fred, but it seems fitting (!) to highlight Fred's role in the history of this work, and to dedicate this tutorial to his memory.[2]

2. Kermack and McKendrick's fit

We begin by revisiting KM's application of their SIR model (1) to the epidemic of plague in Bombay in 1905–1906. The observed data (dots in Fig. 1) were weekly numbers of deaths from plague.

Referring to their version of Fig. 1, KM [p. 714] argued that "As at least 80 to 90 per cent. of the cases reported terminate fatally, the ordinate may be taken as approximately representing $[\mathrm{d}R/\mathrm{d}t]$ as a function of t." Since (non-human) computers did not yet exist [Campbell-Kelly, 2009], and an exact analytical form for this function could not be found, they proceeded to assume [KM, p. 713] that $\frac{\beta}{\gamma}R(t) \ll 1$, which yields the approximate analytical form,

$$\frac{\mathrm{d}R}{\mathrm{d}t} \approx a\,\mathrm{sech}^2(\omega\,t-\phi)\,. \tag{2}$$

Noting that the ***basic reproduction number*** is[3]

$$\mathcal{R}_0 = \frac{N\beta}{\gamma}\,, \tag{3}$$

where N is the total population size, the assumption that yields KM's approximation (2) can be written

$$\frac{R(t)-R(0)}{N} \ll \frac{1}{\mathcal{R}_0}\,, \tag{4}$$

(KM assumed $R(0)=0$); thus, Eq. (2) is a good approximation as long as the proportion of the population that has been infected and removed since the initial time is much less than $1/\mathcal{R}_0$.

Given Eq. (3), the ***effective reproduction number*** at time $t=0$ is

$$\mathcal{R}_\mathrm{e} = \frac{S(0)\beta}{\gamma}\,. \tag{5}$$

[2] We had originally intended to submit this paper to a collection in honour of Fred's memory [Kribs and van den Driessche, 2023].

[3] $\mathcal{R}_0$ is the expected number of secondary cases resulting from a primary case in a wholly susceptible population [Anderson and May, 1991].

In terms of $\mathcal{R}_e$, γ, $S(0)$ and $I(0)$, the parameters in Eq. (2) can be written[4]

$$\omega = \frac{\gamma}{2}\sqrt{(\mathcal{R}_e - 1)^2 + \frac{2I(0)}{S(0)}\mathcal{R}_e^2}\,, \tag{6a}$$

$$\phi = \operatorname{arctanh}\left(\frac{\mathcal{R}_e - 1}{2\,\omega/\gamma}\right)\,, \tag{6b}$$

$$\text{and} \qquad a = \frac{2\,\omega^2 S(0)}{\gamma\,\mathcal{R}_e^2}\,. \tag{6c}$$

The values of these parameters that KM estimated for the Bombay plague epidemic are listed in the KM column of Table 1. Using these values, KM plotted their "calculated" curve, which we have reproduced in blue in Fig. 1.

3. How to fit the model to the data

The blue curve in Fig. 1 does appear to provide a reasonable fit to the data, but KM gave no indication of how their parameter estimates were obtained. Whatever their process, they must have engaged in some sort of ***trajectory matching***, i.e., adjusting parameter values until the model—Eq. (2) in their case—is, by some measure, close to the observed data points. The most obvious metric for this purpose is the Euclidean distance between the model curve and the data. Thus, a natural ***objective function*** to minimize is

$$\sum_{\ell=1}^{n_t} \big(x(t_\ell;\boldsymbol{\theta}) - x_{\text{obs}}(t_\ell)\big)^2\,, \tag{7}$$

where the observed data are the points $\{(t_\ell, x_{\text{obs}}(t_\ell)) : \ell = 1,\ldots,n_t\}$, $\boldsymbol{\theta}$ is the vector of parameters, and $x(t;\boldsymbol{\theta})$ is the model; for KM's problem, the parameter vector is $\boldsymbol{\theta} = (a,\omega,\phi)$ and the model is given by Eq. (2). (Note that we write $x_{\text{obs}}(\cdot)$ when referring to observations of the variable x and $x(\cdot\,;\cdot)$ when referring to the model.) Choosing this objective function is equivalent to assuming that the $x_{\text{obs}}(t_\ell)$ values are direct (but noisy) observations of the state variable $x(t)$. When the connection between the dynamical system and our observations is more complicated, we need to define an explicit ***observation process***; see Sect. 4. Minimizing (7) with respect to $\boldsymbol{\theta}$ would have required some heroic arithmetic with a pencil and paper in 1927, but it is a simple task with the aid of a modern computer.

In the following segment of R code, we fit equation (2) to the Bombay plague data (which are included in the `fitode` package that we describe below, as a data frame with columns `week` and `mort`). We exploit R's nonlinear least squares function (`nls`), which attempts to minimize the distance (7) to the data, starting from an initial guess (`start`).

```
sech <- function(x) {1/cosh(x)}
```

[4] There is a typographical error in equation (31) of KM: their factor $\sqrt{-q}$ should be $(-q)$ in their equivalent of the parameter we call a. Bacaër [2012, §3] corrected this error without comment.

```
KM_approx <- function(t, a, omega, phi) {a * sech(omega*t - phi)^2}
KM.parameters <- c(a = 890, omega = 0.2, phi = 3.4)
nlsfit <- nls(mort ~ KM_approx(week, a, omega, phi),
              data = fitode::bombay,
              start = KM.parameters)
nls.parameters <- coef(nlsfit)
print(nls.parameters)
```

```
##           a       omega         phi
## 874.7545749   0.1935916   3.3720557
```

Above, we chose as our starting value the fitted parameter values of KM. Our least squares parameter values differ from KM's by a few percent (see Table 1). The least squares fitted function is shown in orange in Fig. 1.

Starting from someone else's fit is not a great way to test the method, but fortunately the least squares fit for this problem is not very sensitive to the starting value. To pick reasonable starting values, it often helps to think about the meaning of parameters. For example, in the case of Eq. (2), it is useful to note that a is the maximum of the function, and if write $\omega t - \phi$ as $\omega(t - t_\mathrm{p})$ then

$$t_\mathrm{p} = \frac{\phi}{\omega} \tag{8}$$

is the ***peak time*** (at which the maximum occurs); both a and t_p can be approximated by looking at the plotted data. Assuming $I(0)/S(0) \ll 1$, ω is half the initial exponential growth rate[5], so it can be approximated easily by plotting the data on a log scale, estimating the initial slope, and dividing by 2. Very rough guesses for a, t_p and ω are sufficient to converge on the same fit:

```
a.guess <- 1000    # crude "by eye" estimate of peak value,
tpeak.guess <- 15  # peak time,
omega.guess <- 1   # and half the initial growth rate
phi.guess <- omega.guess * tpeak.guess
nlsfit <- nls(mort ~ KM_approx(week, a, omega, phi),
              data = fitode::bombay,
              start = c(a = a.guess, omega = omega.guess,
                        phi = phi.guess))
print(nls.parameters <- coef(nlsfit))
```

```
##           a       omega         phi
## 874.7550490   0.1935918   3.3720589
```

However, if you experiment with starting values, you will find that if you pick sufficiently *bad* starting values, then `nls` will fail. For example, starting from $a = 2000$, $t_\mathrm{p} = 5$, and $\omega = 0.1$ yields a "singular gradient" error. More interestingly, starting from $a = 500$, $t_\mathrm{p} = 5$, and $\omega = 0.1$ yields $a = 869$, $\omega = -0.19$, $\phi = -3.48$, which is far from our fitted values and illustrates an important fact: there is not necessarily a unique best fit set of parameters! In this case, the alternative solution exists because $\mathrm{sech}^2(x)$ is symmetric about the y axis, but in general, there can be multiple local

[5]From Eq. (1b), the initial exponential growth rate is $\beta S(0) - \gamma = \gamma(\mathcal{R}_\mathrm{e} - 1)$.

minima that cause nonlinear optimizers to converge to points that may or may not represent equally good fits to the data. The potential existence of multiple local optima makes fitting to data hard; you need to be cautious, and use common sense, in interpreting the solutions found by your software (always plot the solutions!). Raue et al. [2013] give suggestions for how to diagnose and handle multiple optima.

If you know that your parameters should be in a certain range, then you can exclude values outside that range. For example, to ensure that all the parameters are non-negative (and exclude the alternative fit above), you would add the `nls` option

```
lower = c(a = 0, omega = 0, phi = 0)
```

which would prevent convergence to negative ω and ϕ. Alternatively, you could write

$$a = e^A, \quad \omega = e^\Omega, \quad \phi = e^\Phi \,, \tag{9}$$

and fit A, Ω, and Φ, which would guarantee positive a, ω, and ϕ without having to constrain the values of the fitted parameters. While this last suggestion may just seem like a cute trick, there is more to it than that. Many more optimization algorithms are available for unconstrained fitting; numerical parameter values of very small magnitude can also lead to numerical instability, so it is advantageous to link parameters that must lie in a given range to unconstrained parameters that can be fit more easily [Bolker, 2008, pp. 328–329]. In Eq. (9), the ***link function*** that converts the parameters to the unconstrained scale is $\log(x)$. Another common link function is $\mathrm{logit}(x) = \log(x/(1-x))$ (the log-odds function, or the inverse of the logistic function), which converts the unit interval $(0, 1)$ to $(-\infty, \infty)$, and is convenient when parameters represent proportions or probabilities. (Requiring positivity is so common that `fitode` uses a log link for all parameters by default.)

If we accept our fit as satisfactory, what can we infer about the dynamics of plague that KM were attempting to capture with the SIR model (1)? We need to convert the parameters of KM's approximation (6) back to the original parameters that are directly related to the mechanism of disease spread formalized by the model (i.e., β and γ, and initial conditions $S(0)$ and $I(0)$).

The nonlinear algebraic relationships specified by Eq. (6) can be inverted[6] analytically[7] [Bacaër, 2012, §3], to obtain

$$\mathcal{R}_\mathrm{e} = 1 + \frac{2\,\omega I(0)\sinh(\phi)\cosh(\phi)}{a}\,, \tag{10a}$$

$$\gamma = \frac{2\,\omega\tanh\phi}{\mathcal{R}_\mathrm{e} - 1}\,, \tag{10b}$$

$$S(0) = \frac{2\,\mathcal{R}_\mathrm{e}^2\,I(0)\sinh^2\phi}{(\mathcal{R}_\mathrm{e}-1)^2}\,. \tag{10c}$$

Since there are four original parameters (β, γ, $S(0)$, $I(0)$) and only three parameters in KM's approximation (2) (a, ω, ϕ), one of the four original parameters needs to be specified separately; in Eq. (10) above we have taken this to be the initial prevalence

[6] Our expressions are slightly different from those of Bacaër [2012, eq. (3)] because we have corrected a minor error. At the start of §3 of Bacaër [2012], in the expression for Q, the term $2Ry_0/x_0$ should be $2R^2y_0/x_0$; this missing square is propagated through to the inversion formulae.

[7] In (common) situations in which nonlinear algebraic equations cannot be solved analytically, they can still be solved numerically, for example with the `nleqslv` package in R.

$I(0)$. From Eq. (10), we can compute the transmission rate,

$$\beta = \frac{\mathcal{R}_\mathrm{e}\gamma}{S(0)}\,, \tag{11}$$

and the mean intrinsic generation interval [Champredon and Dushoff, 2015],

$$T_\mathrm{g} = \frac{1}{\gamma}\,, \tag{12}$$

which is the same as the mean infectious period in this simple model [Pybus et al., 2001, Roberts and Heesterbeek, 2007, Wallinga and Lipsitch, 2007, Krylova and Earn, 2013, Champredon et al., 2018]. Table 1 lists the values of the parameters as estimated by KM and by us using `nls`.

Correctly handling weekly mortality. We have glossed over the fact that we have fitted observed weekly mortality to the *instantaneous* rate, $\mathrm{d}R/\mathrm{d}t$ (2), which is not observed. We did this because it is what KM did, and we wanted to be able to compare formal nonlinear least squares fits to KM's results[8]. Weekly mortality reported at time t_ℓ should really be modelled as the aggregation of $\mathrm{d}R/\mathrm{d}t$ over the preceding week, i.e., it would be better to define

$$x(t_\ell;\boldsymbol{\theta}) = \int_{t_{\ell-1}}^{t_\ell} \frac{\mathrm{d}R}{\mathrm{d}t}\,\mathrm{d}t \tag{13a}$$

$$= \int_{t_{\ell-1}}^{t_\ell} a\,\mathrm{sech}^2(\omega\,t-\phi)\,\mathrm{d}t \tag{13b}$$

$$= \frac{a}{\omega}\Big(\tanh\left(\omega\,t_\ell-\phi\right)-\tanh\left(\omega\,t_{\ell-1}-\phi\right)\Big)\,. \tag{13c}$$

Indeed, whether we are fitting to mortality or incidence or another instantaneous rate, we should be integrating over the observation interval, which is precisely what we do below when fitting to the ODEs directly. In addition, we really ought to consider the fact that not all infections end in death—we have followed KM in assuming that the ***infection fatality proportion*** is 100%. Similarly, when analyzing incidence data, the ***reporting proportion*** ought to be taken into account.

4. Uncertainty

To this point, we have addressed only an optimization problem. We solved it using the method of nonlinear least squares, which yields estimates of the values of the parameters of the model (2). But our best estimates are just that: *estimates*, not known values of the parameters.

To quantify uncertainty in our estimates, we need a statistical framework. The typical output of such a framework is a ***confidence interval*** (CI) within which our best estimate lies. For example, the final column of Table 1 lists 95% CIs on our `nls` parameter estimates, and the light orange shaded region in the top panel of Fig. 1 is a 95% ***confidence band***, which shows CIs for each point of the fitted model curve.

[8] In his reanalysis of KM's results, Bacaër [2012] also retained this conceptual error.

To understand how to estimate CIs, we will start by thinking about our observation model, the probability of observing the data $\{x_{\text{obs}}(t_\ell)\}$ given the model trajectory $x(t;\boldsymbol{\theta})$. We imagine that the model—for now, KM's approximation (2)—is a perfect representation of reality, and we consider the deviations from the model curve in Fig. 1 to be observation errors. A simple observation model assumes that the observation error for each data point is independent and identically distributed (iid), and drawn from a Normal distribution with zero mean and standard deviation σ equal to the standard deviation of the residuals (the differences between the model curve and the observed data). Then the joint probability density $\mathbb{p}$ of the data given the model is

$$\mathbb{p}(\text{data} \mid \text{model}) = \mathbb{p}(\{x_{\text{obs}}(t_\ell)\} \mid \boldsymbol{\theta}) \tag{14a}$$

$$= \prod_{\ell=1}^{n} \left[\lim_{\Delta x_\ell \to 0} \frac{\mathbb{P}\big(x(t_\ell;\boldsymbol{\theta}) \le x_{\text{obs}}(t_\ell) < x(t_\ell;\boldsymbol{\theta}) + \Delta x_\ell\big)}{\Delta x_\ell} \right] \tag{14b}$$

$$= \prod_{\ell=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\big(x(t_\ell;\boldsymbol{\theta}) - x_{\text{obs}}(t_\ell)\big)^2}{2\sigma^2} \right) \right] . \tag{14c}$$

Note that we write $\mathbb{P}$ for the probability measure and $\mathbb{p}$ for the probability density above. We use a probability density function here because the Normal is a continuous distribution; we would use a probability mass function for a discrete response distribution such as the Poisson. In practice, we don't have to worry about this distinction when we are estimating the parameters of an epidemic model (the elements Δx_ℓ will always appear as constant multipliers or divisors and don't affect any of our conclusions). Consequently, in the interests of brevity, below we interpret $\mathbb{p}$ as either probability mass or probability density, depending on whether the associated distribution is discrete or continuous, and refer simply to "probability".

Using these assumptions we can adopt a ***maximum likelihood*** framework, where we consider parameter values that maximize the probability of observing the data (14) to be the best [Bolker, 2008]. We define the ***likelihood*** $\mathcal{L}$ of a set of parameter values $\boldsymbol{\theta}$ as

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{p}(\{x_{\text{obs}}(t_\ell)\} \mid \boldsymbol{\theta}) . \tag{15}$$

Maximizing $\mathcal{L}$ with respect to $\boldsymbol{\theta}$ or, equivalently, minimizing the negative log-likelihood, yields an estimate,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \tag{16a}$$

$$= \arg\min_{\boldsymbol{\theta}} \big(-\log \mathcal{L}(\boldsymbol{\theta})\big) \tag{16b}$$

$$= \arg\min_{\boldsymbol{\theta}} \Big(\sum_{\ell=1}^{n_t} \big(x(t_\ell;\boldsymbol{\theta}) - x_{\text{obs}}(t_\ell)\big)^2 \;+\; \text{constant} \Big) \tag{16c}$$

$$= \arg\min_{\boldsymbol{\theta}} \sum_{\ell=1}^{n_t} \big(x(t_\ell;\boldsymbol{\theta}) - x_{\text{obs}}(t_\ell)\big)^2 , \tag{16d}$$

which—lo and behold—agrees exactly with (7), the ***ordinary least squares*** (OLS) solution! The standard way of expressing this is to say that the OLS solution $\hat{\boldsymbol{\theta}}$ is the ***maximum likelihood estimate*** (MLE) of $\boldsymbol{\theta}$, under the assumption of independent,

identically distributed (i.e., mean-zero, constant-variance) Normal observation errors in the time series.

Having introduced the idea of maximum likelihood, we can do better by making a more realistic assumption about the error distribution. We will then end up with a different likelihood function to maximize, and obtain a different $\hat{\boldsymbol{\theta}}$, but the basic idea is the same.

So what is a better assumption about the observation error distribution, and how can we use the likelihood function to estimate uncertainty in $\hat{\boldsymbol{\theta}}$ and on the fitted trajectory?

Our data are actually non-negative, discrete counts of deaths (or cases in other epidemiological contexts), so a continuous, real-valued Normal distribution is somewhat unrealistic. More importantly, we expect (and can see in the plots of our fitted curves) that the magnitude of error in the observations will vary over the course of the epidemic; the error might be ± 2 at the beginning of the epidemic when mortality is low and ± 50 at the peak.

We could address both of these problems by using a Poisson distribution of observations with mean equal to the fitted model trajectory [Eq. (1c) or Eq. (2)]. This approach handles discrete observations and allows the variance to change as a function of the mean. However, the Poisson distribution assumes ***equidispersion***—the variance is equal to the mean—while typical observation errors are ***overdispersed***, meaning that the variance is greater than the mean. Ignoring overdispersion will underestimate the uncertainty in the parameters and lead to overly narrow confidence intervals on parameters and predictions [Li et al., 2018]. The negative binomial distribution is the most common way to generalize the Poisson to allow for overdispersion [Lindén and Mäntyniemi, 2011], although other distributions such as the generalized Poisson are occasionally used [Brooks et al., 2019, Kim et al., 2022].

The probability mass function for the ***negative binomial distribution*** (for counts $x = 0, 1, 2, \ldots$) is

$$\texttt{NB}(x; \mu, k) = \frac{\Gamma(k+x)}{\Gamma(k)x!}\left(\frac{k}{k+\mu}\right)^k\left(\frac{\mu}{k+\mu}\right)^x. \tag{17}$$

The predicted variance of a particular observation $x_{\text{obs}}(t_\ell)$ is given by $\mu_\ell(1+\mu_\ell/k)$, where $\mu_\ell(\boldsymbol{\theta}) = x(t_\ell;\boldsymbol{\theta})$ is the model evaluated at the ℓ^{th} observed data point [*cf.* (7) and (13)]. The maximum likelihood estimate is, therefore,

$$\begin{aligned}\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{\ell=1}^{n_t} \Big(&-\log\Gamma\big(x_{\text{obs}}(t_\ell)+k\big) + \log\Gamma(k) + \log(x_{\text{obs}}(t_\ell)!) \\ &- k\log\left(\frac{k}{k+\mu_\ell(\boldsymbol{\theta})}\right) - x_{\text{obs}}(t_\ell)\log\left(\frac{\mu_\ell(\boldsymbol{\theta})}{k+\mu_\ell(\boldsymbol{\theta})}\right)\Big).\end{aligned} \tag{18}$$

Here, the overdispersion parameter k also needs to be estimated alongside $\hat{\boldsymbol{\theta}}$ to maximize the likelihood. This is different from the likelihood associated with Normal errors, where σ^2 can be either computed as the variance of the residuals across the full time series or estimated jointly with model parameters.

Regardless of the form of the likelihood function, we can use it to obtain CIs on the MLE $\hat{\boldsymbol{\theta}}$. A relatively simple approach is to use the the curvature of $-\log\mathcal{L}(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$ to infer parameter values of a multivariate Normal distribution for $\boldsymbol{\theta}$. At $\hat{\boldsymbol{\theta}}$, the

shape of $-\log\mathcal{L}$ is described by its ***Hessian matrix*** (the matrix of second order partial derivatives of $-\log\mathcal{L}$, also known as the ***Fisher information matrix***), and the inverse of the Hessian is the ***variance-covariance matrix*** $\mathrm{Cov}(\boldsymbol{\theta})$ that specifies the desired multivariate Normal with mean $\hat{\boldsymbol{\theta}}$. This relationship between $\mathrm{Cov}(\boldsymbol{\theta})$ and the Hessian of $-\log\mathcal{L}$ is, admittedly, not obvious! See Bolker [2008, §6.5] for a heuristic explanation or Wasserman [2010, §§9.7, 9.10] for a rigorous (if terse) explanation.

The diagonal elements of $\mathrm{Cov}(\boldsymbol{\theta})$ are the (estimated) variances of the parameter estimates, so we can take their (positive) square roots to get the standard error (SE) and compute approximate 95% confidence intervals by adding $\pm 1.96\,\mathrm{SE}$ to $\hat{\boldsymbol{\theta}}$ (± 1.96 represents a range containing 95% of the probability of a standard Normal distribution). To obtain CIs on *functions of the fitted parameters* (e.g., $\mathcal{R}_0$ or γ if our model is KM's approximation (2)), we build on the idea that if the error in a parameter a is Δa, then the associated error in a (differentiable) function $g(a)$ is $\Delta g \approx g'(a)\Delta a$. Given a (smooth) nonlinear function $g(\boldsymbol{\theta})$ of the parameters, the ***Delta Method*** [Dorfman, 1938, Ver Hoef, 2012] expands $\mathrm{Var}(g(\boldsymbol{\theta}))$ to first order about $\hat{\boldsymbol{\theta}}$, which gives us the variance-covariance matrix of $g(\boldsymbol{\theta})$ [Bolker, 2008, §7.5.2] [Wasserman, 2010, §9.9]. In particular, the variance of $g(\boldsymbol{\theta})$ is

$$\begin{aligned}
\mathrm{Var}(g(\boldsymbol{\theta})) &\approx \mathrm{Var}\big[g(\hat{\boldsymbol{\theta}}) + (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\big] && \text{(19a)}\\
&= \mathrm{Var}\big[(\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\big] && \text{(19b)}\\
&= \mathbb{E}\big[\big((\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\big)^2\big] && \text{(19c)}\\
&= \mathbb{E}\big[(\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})^{\mathsf{T}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathsf{T}} (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})\big] && \text{(19d)}\\
&= (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})^{\mathsf{T}}\, \mathbb{E}\big[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathsf{T}}\big]\, (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}}) && \text{(19e)}\\
&= (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})^{\mathsf{T}}\, \mathrm{Cov}(\boldsymbol{\theta})\, (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}}) && \text{(19f)}
\end{aligned}$$

We can again get the 95% CIs by taking square roots and computing $g(\hat{\boldsymbol{\theta}}) \pm 1.96\,\mathrm{SE}$.

Given a fit of KM's approximation (2) to the time series data, which yields $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{\omega}, \hat{\phi})$, we can apply the Delta method (19) to the nonlinear relationships (10) to obtain CIs on $g(\hat{\boldsymbol{\theta}}) = (\widehat{\mathcal{R}_{\mathrm{e}}}, \widehat{\gamma}, \widehat{S(0)})$. This is precisely how we obtained the CIs on the derived parameters listed in Table 1. Perhaps less obviously, we can also use the Delta method to obtain CIs on the fitted trajectory at each observation time t_ℓ (and hence obtain a confidence band) by considering $g(\boldsymbol{\theta}) = x(t_\ell; \boldsymbol{\theta})$. This is how we obtained the confidence band for the nonlinear least squares fit (light orange) shown in Fig. 1.

Better confidence intervals can be obtained using the ***profile likelihood***, which is calculated by fixing a set of model parameters to specific values and fitting the remaining parameters to maximize the likelihood [Bolker, 2008, §7.5.1]. By calculating the profile likelihood across a range of parameter values, we obtain the profile likelihood surface, from which confidence intervals can be estimated using the likelihood ratio test [Bolker, 2008, §6.4.1.1]. While profile likelihoods generally give more accurate estimates of confidence intervals, calculating the profile likelihood can be challenging, if not practically impossible, for derived parameters or epidemic trajectories [Bolker, 2008, §7.5.1.2]. Consequently, we rely on the Delta Method here.

5. Fitting the ODE

Until now, we have focused on fitting KM's approximation (2) rather than actual solutions of the SIR model (1). If we had an exact analytical solution of the SIR ODE (1) then we could proceed as above, replacing the approximate analytical expression (2) with the exact formula. Since we do not have an exact solution, we instead rely on numerical solutions of the ODE. Fitting numerical solutions of ODEs to data introduces significant coding/computational challenges, but conceptually the problem is the same as if we did have an analytical formula. We can still use the Delta method (19) to estimate uncertainty, but calculating the gradient $(\nabla_{\boldsymbol{\theta}} g)(\boldsymbol{\theta})$ is not straightforward if g is a numerical solution of an ODE; we must simultaneously solve a set of ***sensitivity equations*** [Raue et al., 2013, Eq. (6)] alongside the main differential equations. Sensitivity equations define the time derivatives of the gradients of trajectories with respect to the parameters. They can easily be derived using the chain rule; if we write a generic, autonomous ODE for $\boldsymbol{x}(t;\boldsymbol{\theta})$ as

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\,, \qquad \boldsymbol{x}(0,\boldsymbol{\theta}) = \boldsymbol{x}_0(\boldsymbol{\theta}), \tag{20}$$

then the sensitivity equations are

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\nabla_{\boldsymbol{\theta}}\boldsymbol{x}(t;\boldsymbol{\theta})\Big) = \nabla_{\boldsymbol{\theta}}\Big(\frac{\mathrm{d}\boldsymbol{x}(t;\boldsymbol{\theta})}{\mathrm{d}t}\Big) = \nabla_{\boldsymbol{\theta}}\Big(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\Big) \tag{21a}$$

$$= \nabla_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\,\nabla_{\boldsymbol{\theta}}\boldsymbol{x}(t;\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\,. \tag{21b}$$

If $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are $n_{\boldsymbol{x}}$ and $n_{\boldsymbol{\theta}}$dimensional, respectively, then the $n_{\boldsymbol{x}} n_{\boldsymbol{\theta}}$ ***sensitivities*** $\mathsf{S}_{ij}(t)$ are given by the $n_{\boldsymbol{x}} \times n_{\boldsymbol{\theta}}$ ***sensitivity matrix***,

$$\mathbf{S}(t) = \nabla_{\boldsymbol{\theta}}\boldsymbol{x}(t;\boldsymbol{\theta})\,. \tag{22}$$

Eq. (21) defines a set of $n_{\boldsymbol{x}} n_{\boldsymbol{\theta}}$ differential equations for S_{ij},

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \big[\nabla_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\big]\,\mathbf{S} + \big[\nabla_{\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\big]\,, \tag{23a}$$

which can be solved jointly with the original ODEs (20) for the state variables ($\boldsymbol{x}$) by specifying initial conditions

$$\mathbf{S}(0) = \nabla_{\boldsymbol{\theta}}\big(\boldsymbol{x}_0(\boldsymbol{\theta})\big)\,. \tag{23b}$$

We can then use a further chain-rule step to compute the (total) derivative of the log-likelihood of the observations with respect to the parameters. To get this right, it helps to make explicit the dependence on the trajectory ($\boldsymbol{x}$) versus dependence on the parameters ($\boldsymbol{\theta}$, by which we will now mean all parameters, including parameters of the trajectory model and of the observation process model). For a general function $\boldsymbol{\Phi}(\boldsymbol{x},\boldsymbol{\theta})$, the total derivative with respect to $\boldsymbol{\theta}$ is

$$\frac{\mathrm{d}\boldsymbol{\Phi}}{\mathrm{d}\boldsymbol{\theta}} = \nabla_{\boldsymbol{x}}\boldsymbol{\Phi}\;\nabla_{\boldsymbol{\theta}}\boldsymbol{x} + \nabla_{\boldsymbol{\theta}}\boldsymbol{\Phi}\,. \tag{24}$$

To apply this to the log-likelihood, it is helpful to make dependence on the trajectory $\boldsymbol{x}$ explicit. Consistent with our notation above [e.g., Eq. (7)], we write $\boldsymbol{x}_{\text{obs}}(t_\ell)$ for the observations at times $t_\ell \in \{t_1, t_2, \ldots, t_{n_t}\}$, making it easier to distinguish them from the fitted model trajectory evaluated at these times, $\boldsymbol{x}(t_\ell; \boldsymbol{\theta})$. Then

$$\frac{\mathrm{d} \log \mathcal{L}(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \Big(\log \mathbb{p}\big(\{\boldsymbol{x}_{\text{obs}}(t_\ell) : \ell = 1, \ldots, n_t\} \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) \Big) \tag{25a}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \Big(\log \prod_{\ell=1}^{n_t} \mathbb{p}\big(\boldsymbol{x}_{\text{obs}}(t_\ell) \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) \Big) \tag{25b}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \sum_{\ell=1}^{n_t} \Big(\log \mathbb{p}\big(\boldsymbol{x}_{\text{obs}}(t_\ell) \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) \Big) \tag{25c}$$

$$= \sum_{\ell=1}^{n_t} \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \Big(\log \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \Big) \qquad \big[\text{abbreviating } \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \mathbb{p}\big(\boldsymbol{x}_{\text{obs}}(t_\ell) \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big)\big] \tag{25d}$$

$$= \sum_{\ell=1}^{n_t} \frac{1}{\mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})} \Big(\nabla_{\boldsymbol{x}} \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})\, \nabla_{\boldsymbol{\theta}} \boldsymbol{x} + \nabla_{\boldsymbol{\theta}} \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \Big) \Big|_{\boldsymbol{x} = \boldsymbol{x}(t_\ell; \boldsymbol{\theta})} \tag{25e}$$

$$= \sum_{\ell=1}^{n_t} \frac{1}{\mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})} \Big(\nabla_{\boldsymbol{x}} \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})\, \mathbf{S}(t_\ell) + \nabla_{\boldsymbol{\theta}} \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \Big) \Big|_{\boldsymbol{x} = \boldsymbol{x}(t_\ell; \boldsymbol{\theta})}, \tag{25f}$$

where we typically assume the probability distribution

$$\mathbb{p}\big(\boldsymbol{x}_{\text{obs}}(t_\ell) \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) = \prod_{i=1}^{n_x} \mathtt{NB}(x_{\text{obs},i}(t_\ell);\, x_i(t_\ell, \boldsymbol{\theta}),\, \boldsymbol{\theta})\,. \tag{26}$$

We have slightly abused notation here, compared with Eq. (17); we have written $\boldsymbol{\theta}$ rather than k as the final argument of the negative binomial distribution, since there might be a different k for each observed variable x_i, and we collect all parameters into the single vector $\boldsymbol{\theta}$. (The examples we discuss in this paper involve only a single observed time series, so $n_x = 1$.)

Integrating the sensitivity equations (23) in parallel with the ODEs (20) is a computationally efficient and numerically stable way to calculate the overall gradients of the log-likelihood with respect to the parameters, which makes nonlinear estimation more robust and efficient. We can also use these gradients to calculate CIs using the Delta method. Raue et al. [2013] give a detailed comparison between using the sensitivity equations and computing gradients by finite-difference approximations. (Bjørnstad [2018, Chapter 9] also gives an introduction to trajectory matching.)

The `fitode` package[9] does all of this computational work under the hood, and makes it as easy for a user to fit an ODE to data as it was for us to use `nls` above to fit a curve based on an analytical formula. We begin illustrating the use of the package by fitting the SIR model (1) to the Bombay plague epidemic.

We first load the package

```
library(fitode)
```

and define a model object:

[9] `fitode` is available on CRAN, and can be installed via `install.packages("fitode")`.

```
SIR_model <- odemodel(
    name="SIR model",
    model=list(
        S ~ - beta * S * I,
        I ~ beta * S * I - gamma * I,
        R ~ gamma * I
    ),
    observation = list(
        mort ~ dnbinom(mu = R, size = k)
    ),
    diffnames="R",
    initial=list(
        S ~ S0,
        I ~ I0,
        R ~ 0
    ),
    par=c("beta", "gamma", "S0", "I0", "k")
)
```

In the model definition above:

`model` specifies the vector field given by the ODE (1).

`observation` specifies the observation model: the observed data (`mort`) are assumed to arise from sampling from the negative binomial distribution [`dnbinom`, Eq. (17)] with overdispersion parameter k. Ordinary least squares (with normally distributed observation errors) can be implemented by changing the `observation` argument to `mort ~ ols(mean = R)`. The mean of the distribution is given by the incidence derived from the fitted model trajectory [Eq. (13a)],

$$\mu(t_\ell) = \int_{t_{\ell-1}}^{t_\ell} \frac{\mathrm{d}R}{\mathrm{d}t}\,\mathrm{d}t = R(t_\ell) - R(t_{\ell-1})\,, \tag{27}$$

Fitting to such differences, useful whenever the observations represent accumulated values of processes (such as infections, recoveries, or deaths) between observation times, is implemented by using the `diffnames` argument to specify the state variable for which consecutive differences are to be used (so, if the focal variable is R then `fitode` fits to $R(t_\ell) - R(t_{\ell-1})$ rather than $R(t_\ell)$).

`initial` conditions are expressed as numbers of individuals.

`par` refers to the parameters to be fitted: β, γ, initial conditions $S(0)$ and $I(0)$, and the overdispersion parameter k.

Since we are taking the difference $\mu(t_\ell) = R(t_\ell) - R(t_{\ell-1})$ to calculate the mortality trajectory,[10] we have to add an extra row representing t_0 to the data set in order to compute $\mu(t_1) = R(t_1) - R(t_0)$:

```
bombay2 <- rbind(
```

[10] Modelers often fit trajectories to cumulative curves. However, doing so is ill-advised because points in a cumulative time series are not independent, making it difficult to define CIs [King et al., 2015].

```
    c(times=bombay$week[1] -
          diff(bombay$week)[1], mort=NA),
    bombay
)
```

Taking our previous parameter estimates from `nls` as starting values (and choosing a starting value for k), we can fit the model by calling the `fitode` function:

```
SIR_start <- c(beta=beta.nls, gamma=gamma.nls,
                I0=I0.KM, S0=S0.nls, k=50)
SIR_fit <- fitode( model = SIR_model, data = bombay2,
                   fixed = list(gamma=gamma.nls),
                   start = SIR_start, tcol = "week" )
```

In the fitting function above:

`model` specifies the ODE model to be fitted.
`data` specifies the data.
`fixed` specifies parameter values to be fixed (and therefore not estimated); above, we chose to assume that the recovery rate γ is known (due to parameter unidentifiability[11]).
`start` specifies the starting parameter set for the optimization[12].
`tcol` specifies the name of the time column of the data frame.

The resulting fits are plotted in Fig. 1 and summarized in Table 2. The estimated parameter values (the ***coefficients*** of the model) can be obtained via `coef(SIR_fit)`. The coefficients together with associated confidence intervals are obtained via `confint(SIR_fit)`, which can also provide confidence intervals for derived parameters using the Delta method. Note that `fitode` gives discrete predictions (rather than smooth curves) because we are calculating mortality at discrete (weekly) time intervals using Eq. (27).

6. Cautionary remarks concerning fits to Bombay plague

We have highlighted the Bombay plague data because of their prominent role in KM's paper [Kermack and McKendrick, 1927] and, consequently, for the history of mathematical epidemiology. However, while they provide an interesting example with which to illustrate the process of fitting an epidemiological model to data, modelling plague dynamics with the simple SIR model is, at best, difficult to justify: Bacaër [2012] argues that the trajectory of the Bombay plague epidemic was primarily governed by seasonality rather than SIR dynamics. Indeed, KM themselves recognized that their model involves a sequence of optimistic assumptions, which they admitted were not

[11] In short, unidentifiability of γ means that we can obtain nearly identical fits across a wide range of γ. While it is possible to fit the model without fixing γ, the resulting estimates are sensitive to starting conditions and numerically unstable, preventing a reliable calculation of the Hessian matrix and therefore precluding estimation of confidence intervals. These issues could be addressed alternatively by fixing a different parameter instead and estimating γ. We typically choose to fix γ because the mean duration of infection ($1/\gamma$) can often be estimated from independent data sources; here, to make comparisons of fits easier to interpret, we have fixed γ to the value we estimated via `nls` fits of the KM approximation (2).

[12] In general, worse models (providing a poorer or less identifiable fit to the data) and worse data (fewer data points and more noise) will increase the sensitivity of fits to the starting values.

"strictly" satisfied:

> "We are, in fact, assuming that plague in [humans] is a reflection of plague in rats, and that with respect to the rat (1) the uninfected population was uniformly susceptible; (2) that all susceptible rats in the island had an equal chance of being infected; (3) that the infectivity, recovery, and death rates were of constant value throughout the course of sickness of each rat; (4) that all cases ended fatally or became immune; and (5) that the flea population was so large that the condition approximated to one of contact infection. None of these assumptions are strictly fulfilled and consequently the numerical equation can only be a very rough approximation. A close fit is not to be expected, and deductions as to the actual values of the various constants should not be drawn."
>
> — KM [p. 715]

Given the mental gymnastics required to motivate applying the SIR model to plague transmission, it is surprising that KM did not choose to examine a more obviously suitable disease. The surprise is especially extreme given that the most salient infectious disease epidemic in the 1920s would have been the 1918 influenza pandemic, which did involve direct human-to-human transmission, and for which much more detailed data were available at the time [Rogers, 1920, Frost, 1920, Eichel, 1923].

7. Influenza in Philadelphia, October 1918

Deaths caused ultimately by influenza are often attributed to pneumonia [Earn et al., 2002], so influenza mortality studies typically combine pneumonia and influenza (P&I). Among published tables summarizing P&I mortality during the 1918 pandemic, a particularly valuable example concerns the main wave in the city of Philadelphia [Rogers, 1920]. These data are exceptional because they are restricted to a single, large city, and because they provide *daily* counts that capture the detailed temporal pattern (large dots in Fig. 2).

As for Bombay plague, we can fit KM's approximation (2) to the Philadelphia influenza epidemic using nonlinear least squares, which yields the orange curve in Fig. 2. While this `nls` fit does not look unreasonable at a glance, the fitted parameter values (Table 3) are absurd, including a basic reproduction number $\mathcal{R}_0 \approx 2500$ and a mean generation interval $T_g \approx 1.5$ years.

Matching numerically computed trajectories of the exact SIR model using `fitode` gives a fit—the solid gold curve in Fig. 2—that is visually similar to the (orange) fit of KM's approximation, but provides much more realistic parameter estimates (Table 4); in particular, $\mathcal{R}_0 \approx 6.4$ and $T_g \approx 4.3$ days.

If we convert the `fitode` estimates of the SIR parameters to the parameters of KM's approximation, we obtain the dotted gold curve in Fig. 2, which grossly underestimates the magnitude of the epidemic (the epidemic peak occurs much too soon). The KM approximation (2) is good initially, but becomes poorer and poorer over time as the underlying assumption on which it is based (4) becomes less and less valid.

8. Fitting the deterministic SIR model to stochastic simulations

The most compelling tests of estimation methods involve fitting models to data that have been generated from a known model, so we know the true underlying values of the parameters we are trying to estimate.

The most basic test is essentially a consistency check: in the context of the SIR

model, we choose initial conditions $(S(0), I(0))$ and parameter values $(\mathcal{R}_0, T_\mathrm{g})$, compute the associated trajectory by solving Eq. (1) numerically, and then use `fitode` to estimate the parameters. At least if we choose starting values reasonably close to the correct underlying values, `fitode` should converge to those values.

The next level of testing is to take our numerically computed solution and artificially "observe" it with error, i.e., using a noise distribution that we specify. For example, we could take observation errors to be negative binomially distributed with overdispersion parameter k, and then use `fitode` to estimate k together with the other parameters $(S(0), I(0), \mathcal{R}_0, T_\mathrm{g})$.

A still more stringent test is to simulate data from a model that is more complex and realistic than the idealized model that we want to fit, and then see if we can nevertheless recover parameters that correspond to those of our idealized model (e.g., $\mathcal{R}_0$ and T_g for the SIR model). We will take a step in this direction in this section by fitting the deterministic SIR model (1) to data generated by a fully stochastic version of the model.

The standard stochastic SIR model [Andersson and Britton, 2000] can be defined by interpreting the individual terms in Eq. (1) as event rates for stochastic processes in a population of N individuals (in the limit $N \to \infty$ the stochastic model approaches the ODEs (1); see Ethier and Kurtz [1986]). Realizations of this discrete-state model can be generated exactly using the Gillespie algorithm [Gillespie, 1976], or approximately (as we do here) using the "τ-leaping" approach [Gillespie, 2001], which is implemented in the `adaptivetau` R package [Johnson, 2023]. The demographic stochasticity that these algorithms simulate is essential to capture real effects that occur when the number of infected individuals is small (especially the possibility that an epidemic can burn out [Parsons et al., 2024]).

In Fig. 3, the simulated data points show a single realization of the stochastic SIR model with initial state $(S(0), I(0), R(0)) = (1998, 2, 0)$, basic reproduction number $\mathcal{R}_0 = 5$, and mean generation interval $T_\mathrm{g} = \gamma^{-1} = 1$ week. In the top panel, $\mathrm{d}R/\mathrm{d}t$ [Eq. (1c)] with the correct initial conditions and parameter values is shown with solid green, and the KM approximation (2) based on those parameter values is shown with dotted green. The `fitode` fit [based on $\int (\mathrm{d}R/\mathrm{d}t)\,\mathrm{d}t$] and confidence band are shown in gold. The time shift between the deterministic solution and the stochastic realization arises because the stochastic model captures the demographic noise (which causes a randomly distributed delay until the tipping point is reached, i.e., until the epidemic takes off in a roughly deterministic fashion).

As expected, with the correct parameter values, KM's approximation (2) fails once the requirement (4) that $R(t)/N \ll 1/\mathcal{R}_0$ is violated. We can, of course, find values of the parameters (a, ω, ϕ) such that the function $a\,\mathrm{sech}^2(\omega\, t - \phi)$ [Eq. (2)] more closely matches the shape of the full simulated epidemic. Using nonlinear least squares (`nls`) as in previous sections, we obtain visually reasonable agreement (Fig. 3, bottom panel, orange curve; Table 5). This `nls` fit cannot be improved further because the function we are fitting (2) is symmetric about its peak, whereas the rise is steeper than the fall in the simulated epidemic. It is also worth emphasizing that the parameter values that yield the orange curve in Fig. 3 are far from the true parameters that were used in the simulation (Table 5).

The excellent fit of the deterministic trajectory that `fitode` finds (gold in Fig. 3) is achieved by estimating an initial prevalence that is only a third of the true initial prevalence, thereby mimicking the stochastic delay with the deterministic model; all other parameter estimates are nearly identical to the true parameter values used to generate the stochastic trajectory (Table 6).

9. Discussion

We have presented a basic theoretical and practical introduction to standard methods for fitting dynamical models to time series, in the context of infectious disease epidemiology. We explained how to use nonlinear least squares (`nls`) to fit a given function to a time series, and illustrated the approach using the Kermack and McKendrick (KM) analytical approximation (2) to the solution of the standard SIR model (1). We also explained how to fit solutions of ordinary differential equations (ODEs) to a time series—using our R package `fitode`—and obtain parameter estimates and confidence intervals, regardless of whether analytical solutions of the ODEs are available.

`fitode` is flexible enough to handle most compartmental epidemiological and ecological models [Brauer and Castillo-Chavez, 2001, Brauer and Kribs, 2016, Brauer et al., 2019], including non-autonomous models, such as seasonally forced epidemic models [London and Yorke, 1973, Earn et al., 2000, He and Earn, 2007, 2016, Papst and Earn, 2019]. We hope the package will be useful for many readers, not only as a pedagogical tool but also to fit models to novel data. Potential applications abound (we have ourselves used `fitode`'s predecessor, `fitsir`, to study music popularity [Rosati et al., 2021]).

We focused here on three illustrative examples of epidemic time series. The first was the reported weekly mortality from plague in Bombay in 1906 (Fig. 1), which was examined by KM in their original paper [Kermack and McKendrick, 1927]. Although historically important, it is certainly debatable whether we can trust any inferences we might draw from fitting the simple SIR model (1) to these plague data. As we quoted at the end of Sect. 5, to justify the application of their SIR model to these data, KM highlighted five implicit assumptions, any or all of which might be violated. Furthermore, Bacaër found that over the longer term seasonal epidemics of plague occurred in Bombay every year from 1897 to 1911 [Bacaër, 2012, Fig. 2], suggesting that the 1906 epidemic was just one in a long sequence of epidemics that were "driven by seasonality" [Bacaër, 2012, p. 403]. Of course, other mechanisms (e.g., heterogeneity in contact patterns) might play a role as well.

To obtain a deeper understanding of the Bombay plague epidemic, we could formulate a variety of models, fit them to the data using `fitode` or other software, and use a statistical framework for model selection [Burnham and Anderson, 2002] to rank the relative importance of the various mechanisms included in the sequence of models (see, e.g., He et al. [2013] for an example of using this approach to understand the occurrence of three distinct waves in the 1918 influenza pandemic). Alternatively, we could formulate one model that included *all* of the processes and attempt to measure their relative importance by comparing the magnitudes of parameters [Bolker, 2023]. We have not attempted such a study here, since our goal was simply to explain and illustrate the fitting methodology. However, it is worth highlighting that our analysis using the SIR model did reveal a computational challenge that—in the absence of additional information about the Bombay plague outbreak—would likely limit how much can be learned from a model selection exercise: the mean generation interval (T_g) appears to be ***unidentifiable***, i.e., difficult to estimate reliably from the reported weekly plague deaths alone (see Fig. 4).

Our second example was the main wave of the 1918 influenza pandemic in the city of Philadelphia, for which daily mortality from pneumonia and influenza (P&I) was reported (Fig. 2). Again we fitted a numerical solution of the SIR model (1) using `fitode`, and KM's analytical approximation (2), but found—unlike the situation for Bombay plague—that only the `fitode` fit yielded plausible parameter estimates (see

Tables 3 and 4).

Finally, we conducted a kind of test that truly makes most sense to perform *before* fitting to a real, empirically observed time series: we fit models to a simulation that we ran, so we knew the parameter values used to generate the simulated "observations". The simulation was a realization of the stochastic SIR model, to which, again, we fit both the deterministic SIR model (1) using `fitode` and KM's analytical approximation (2) using `nls`. At a glance, both provide visually roughly reasonable fits (Fig. 3, bottom panel) but KM's approximation cannot represent the asymmetry about the peak in the epidemic curve and yields absurd parameter values, whereas `fitode` estimates an epidemic curve with the correct shape and the correct values of the underlying disease-related parameters (Tables 5 and 6). (We did find a discrepancy in the estimates of initial conditions; this was driven by the failure of the stochastic outbreak simulation to take off immediately. A lower initial prevalence is the only mechanism by which the deterministic model can capture the delayed onset of the epidemic. In practice, modelers fitting to epidemic time series by trajectory matching usually pick an "epidemic window" that corresponds to the part of the epidemic that can be reasonably captured by a deterministic model [Earn et al., 2020].)

KM's approximation (2) estimates the simulation parameters badly because the assumption on which it is based (4) is strongly violated in the simulation (Fig. 3). Consequently, the parameters of the KM approximation cannot be interpreted biologically or mechanistically. More generally, a purely phenomenological model with the same number of parameters can sometimes fit a stochastic simulation just as closely or even closer than the deterministic limit of the model that generated the data [Rosati et al., 2021]; a good fit is not, on its own, sufficient to conclude that a model matches the underlying processes of a dynamical system.

While `fitode` provides a relatively easy way to specify ODEs and estimate their parameters from data, any programming language will work to implement the steps we have outlined above, including both free general-purpose languages such as Python [Batista and da Silva, 2022, Gupta, 2023] or commercial, domain-specific tools such as MATLAB [Chowell, 2017] or Berkeley Madonna [Zha et al., 2020]. As long as a language provides tools for integrating arbitrary sets of ODEs (e.g., MATLAB's `ode45`) and optimizing nonlinear functions (e.g., MATLAB's `fminunc` or `lsqnonlin`), it can be used to estimate parameters of ODEs. However, `fitode`'s simple interface, automatic derivation of sensitivity equations, flexible specification of observation models, and provision of confidence intervals make it both convenient and powerful.

Beyond the basics that we have discussed here, `fitode` contains a number of useful advanced features. In particular, `fitode` can

fit to multiple data streams: `fitode` is not limited to fitting a trajectory to a single state variable, such as incidence or prevalence of infected individuals. For example, during the later stages of the COVID-19 pandemic modelers often had access to time series of case reports, hospitalization reports, and wastewater sampling for the same geographic region. If we build a model that includes state variables for hospitalized individuals and for virus concentrations in wastewater, `fitode` can fit the model's parameters using all of the available data (as in Nourbakhsh et al. [2022]).

compute confidence intervals via importance sampling: While the Delta method can compute confidence intervals for derived quantities such as predicted trajectories, it rests on strong and sometimes unreliable assumptions. A more accurate but computationally expensive approach starts by sampling

parameter sets randomly from a multivariate normal distribution with a mean and covariance matrix drawn from the maximum likelihood fit. For each set of parameters in the ensemble, `fitode` computes the likelihood and a predicted trajectory (or some quantity such as the total size of the epidemic); an average value and confidence intervals are derived from weighted moments (means) or quantiles (medians or extremes such as 10th and 90th percentiles).

specify priors and apply Bayesian inference: Unlike maximum likelihood approaches, which seek to estimate the best-fitting parameter set, Bayesian methods aim to estimate a distribution of parameters (also known as the posterior distribution) that are consistent with our previous knowledge about the system (encapsulated in ***prior distributions***) as well as the observed data. These approaches are generally better at handling parameter uncertainties [Elderd et al., 2006] but are usually much more computationally expensive.

`fitode` allows the user to specify prior distributions on parameters; these priors can either reflect previous knowledge of a disease system, or can be used to ***regularize*** a fitting procedure by downweighting extreme values of parameters [Lemoine, 2019], which can help mitigate problems with identifiability (see below).

Bayesian modelers typically use ***Markov chain Monte Carlo*** algorithms to explore the parameter space and approximate the target distribution. `fitode` implements a simple ***Metropolis-Hasting*** sampler [Bolker, 2008, §7.3.1]. (The Stan platform provides a much more powerful Bayesian sampling algorithm using sensitivity equations, built on top of a fully general system for specifying ODEs; however, this tool requires significantly more computational and statistical background to use effectively [Grinsztajn et al., 2021].)

Even with these extensions, modelers may face many challenges when fitting ODEs to data with the `fitode` package, as with fitting any nonlinear model to data. For example, it is often difficult to ensure that the model has converged properly or reached its true maximum. More generally, when they first start attempting to fit models to data, naïve and optimistic epidemic modelers often run into problems of ***structural identifiability*** (the impossibility of estimating particular sets of parameters from data, regardless of how much data is available [Tuncer and Le, 2018, Chowell et al., 2023]) and ***practical identifiability*** (the impossibility of reliably estimating parameters from a particular small, noisy data set [Gallo et al., 2022, Chowell et al., 2023]). In addition to the rigorous methods described by Chowell et al. [2023], using a ***multistart method*** (performing optimization from multiple starting conditions [Raue et al., 2013]), or plotting likelihood surfaces, can help diagnose these problems. Using different optimization methods or reparameterizing the model can also help [Raue et al., 2013, Bolker et al., 2013]. We encourage users of `fitode` who encounter these or other fitting challenges to open issues via the `fitode` GitHub repository (`https://github.com/parksw3/fitode`).

As its name suggests, `fitode` is limited to fitting ODEs to time series. Consequently, by design, `fitode` ignores ***process error***, i.e., random variability that affects both current and future steps of the trajectory—as opposed to ***observation error***, which arises from imperfect measurements or reporting and is usually assumed to be independent of the trajectory itself. A key component of process error is the demographic stochasticity that is inherent to the discrete-state stochastic SIR model discussed above (and to any real host-pathogen system). Parameters of models can also be subject to process error; for example, the transmission rate might depend on

random fluctuations in weather. Properly accounting for process error can be critical for accurately quantifying uncertainties in parameter estimates and model forecasts [King et al., 2015, Taylor et al., 2016, Li et al., 2018]; however, the required statistical and computational procedures are significantly more challenging than the approaches considered here. Popular `R` packages that can fit models with process error include `pomp` [King et al., 2016] and `mcstate` [FitzJohn et al., 2024].

10. Closing remarks: from Fred Brauer to `fitode`

The idea of digging into data seemed like punishment to Fred Brauer, but while he never—to our knowledge—did any data analysis himself, he did develop a sincere appreciation for the value of data in epidemiological research. Fred's curiosity—about how dynamical models can be fit to data, and why it is hard—convinced us that it would be worth writing a paper (and accompanying software) that could draw more dynamicists working on epidemic models into the world of data.

We have provided two answers to Fred's question of "how" to fit models to data (via `nls` or `fitode`), and through examples we have hinted at some of the reasons "why" such fitting can be very difficult. A true understanding of "why it is hard" is something that builds over time with experience, but the key points are that finding optima of a complex multi-dimensional function is hard enough on its own [Raue et al., 2013], and estimating statistically meaningful uncertainty in those optima is extremely challenging [Elderd et al., 2006, Li et al., 2018].

Fred would never have used `fitode`, but would have delighted in seeing it demonstrated and in discussing the theoretical background on model fitting that we have presented in this paper. We hope that others like him, as well as students and researchers who actually do want to dig into data, will benefit from this exposition.

References

R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1991.

H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis*, volume 151 of *Lecture notes in statistics*. Springer-Verlag, New York, 2000.

N. Bacaër. The model of Kermack and McKendrick for the plague epidemic in Bombay and the type reproduction number with seasonality. *Journal of Mathematical Biology*, 64(3):403–422, Feb. 2012. ISSN 0303-6812, 1432-1416. . http://link.springer.com/article/10.1007/s00285-011-0417-5.

N. T. J. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Hafner Press, New York, second edition, 1975.

M. S. Bartlett. *Stochastic population models in ecology and epidemiology*, volume 4 of *Methuen's Monographs on Applied Probability and Statistics*. Spottiswoode, Ballantyne & Co. Ltd., London, 1960.

A. A. Batista and S. H. da Silva. An epidemiological compartmental model with automated parameter estimation and forecasting of the spread of COVID-19 with analysis of data from Germany and Brazil. *Frontiers in Applied Mathematics and Statistics*, 8, Apr. 2022. ISSN 2297-4687. . URL `https://www.frontiersin.org/articles/10.3389/fams.2022.645614`. Publisher: Frontiers.

O. N. Bjørnstad. *Epidemics: Models and Data Using R*. Springer, New York, NY, 1st ed. 2018 edition edition, Nov. 2018. ISBN 978-3-319-97486-6.

B. Bolker. Multimodel approaches are not the best way to understand multifactorial systems. July 2023. URL https://ecoevorxiv.org/repository/view/5722/. Publisher: EcoEvoRxiv.

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

B. M. Bolker, B. Gardner, M. Maunder, C. W. Berg, M. Brooks, L. Comita, E. Crone, S. Cubaynes, T. Davies, P. de Valpine, J. Ford, O. Gimenez, M. Kéry, E. J. Kim, C. Lennert-Cody, A. Magnusson, S. Martell, J. Nash, A. Nielsen, J. Regetz, H. Skaug, and E. Zipkin. Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and Evolution*, 4(6):501–512, June 2013. ISSN 2041210X. . URL http://doi.wiley.com/10.1111/2041-210X.12044.

F. Brauer and C. Castillo-Chavez. *Mathematical models in population biology and epidemiology*, volume 40 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2001.

F. Brauer and C. Kribs. *Dynamical systems for biological modeling: An introduction*. CRC press, 2016.

F. Brauer, C. Castillo-Chavez, and Z. Feng. *Mathematical models in epidemiology*, volume 32. Springer, 2019.

M. E. Brooks, K. Kristensen, M. R. Darrigo, P. Rubim, M. Uriarte, E. Bruna, and B. M. Bolker. Statistical modeling of patterns in annual reproductive rates. *Ecology*, 100(7):e02706, 2019. ISSN 1939-9170. .

E. Brooks-Pollock, L. Danon, T. Jombart, and L. Pellis. Modelling that shaped the early COVID-19 pandemic response in the UK. *Philosophical Transactions of the Royal Society B*, 376(1829):20210001, 2021.

K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: A practical information-theoretic approach*. Springer, New York, 2nd edition, 2002.

M. Campbell-Kelly. Origin of computing. *Scientific American*, 301(3):62–69, 2009.

D. Champredon and J. Dushoff. Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821):20152026, 2015.

D. Champredon, J. Dushoff, and D. J. D. Earn. Equivalence of the Erlang SEIR epidemic model and the renewal equation. *SIAM Journal on Applied Mathematics*, 78(6):3258–3278, 2018. . URL https://epubs.siam.org/doi/10.1137/18M1186411.

G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, Aug. 2017. ISSN 2468-0427. .

G. Chowell, S. Dahal, Y. R. Liyanage, A. Tariq, and N. Tuncer. Structural identifiability analysis of epidemic models based on differential equations: A tutorial-based primer. *Journal of Mathematical Biology*, 87(6):79, Nov. 2023. ISSN 1432-1416. .

O. Diekmann and J. A. P. Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Wiley Series in Mathematical and Computational Biology. John Wiley & Sons, LTD, New York, 2000.

R. A. Dorfman. A note on the δ-method for finding variance formulae. *The Biometric Bulletin*, 1:129–137, 1938.

D. J. D. Earn. A Light Introduction to Modelling Recurrent Epidemics. In F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, volume 1945 of *Lecture Notes in Mathematics*, pages 3–17. Springer, 2008. . URL https://link.springer.com/chapter/10.1007%2F978-3-540-78911-6_1.

D. J. D. Earn. Mathematical epidemiology of infectious diseases. In M. A. Lewis, M. A. J. Chaplain, J. P. Keener, and P. K. Maini, editors, *Mathematical Biology*, volume 14 of *IAS/Park City Mathematics Series*, pages 151–186. American Mathematical Society, 2009. . URL http://www.ams.org/books/pcms/014/.

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

D. J. D. Earn, J. Dushoff, and S. A. Levin. Ecology and evolution of the flu. *Trends in Ecology*

and Evolution, 17(7):334–340, 2002. .

D. J. D. Earn, J. Ma, H. Poinar, J. Dushoff, and B. M. Bolker. Acceleration of plague outbreaks in the second pandemic. *PNAS – Proceedings of the National Academy of Sciences of the U.S.A.*, 117(44):27703–27711, 2020. ISSN 0027-8424. . URL `https://doi.org/10.1073/pnas.2004904117`.

O. R. Eichel. *A Special Report on the Mortality From Influenza in New York State During the Epidemic of 1918–19*. New York State Department of Health, Albany, NY, 1923.

B. D. Elderd, V. M. Dukic, and G. Dwyer. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proceedings of the National Academy of Sciences*, 103(42):15693–15697, 2006.

S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley and Sons, New York, 1986.

S. Eubank, H. Guclu, V. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429 (6988):180–184, 2004.

R. FitzJohn, M. Baguelin, E. Knock, L. Whittles, J. Lees, and R. Sonabend. *mcstate: Monte Carlo Methods for State Space Models*, 2024. URL `https://github.com/mrc-ide/mcstate`. R package version 0.9.20.

W. H. Frost. Statistics of influenza morbidity: with special reference to certain factors in case incidence and case fatality. *Public Health Reports*, 35:584–597, 1920.

L. Gallo, M. Frasca, V. Latora, and G. Russo. Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models. *Science Advances*, 8(3):eabg5234, Jan. 2022. ISSN 2375-2548. .

D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.

D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

E. Goldstein, J. Dushoff, J. Ma, J. Plotkin, D. J. D. Earn, and M. Lipsitch. Reconstructing influenza incidence by deconvolution of daily mortality time series. *PNAS – Proceedings of the National Academy of Sciences of the U.S.A.*, 106(51):21825–21829, 2009. .

L. Grinsztajn, E. Semenova, C. C. Margossian, and J. Riou. Bayesian workflow for disease transmission modeling in Stan. *Statistics in Medicine*, 40(27):6209–6234, 2021. ISSN 1097-0258. . URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9164`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.9164.

N. Gupta. On the Calibration of Compartmental Epidemiological Models. Master's thesis, New York University Tandon School of Engineering, United States – New York, 2023. URL `https://www.proquest.com/docview/2820207706/abstract/C69EF306BFF041E0PQ/1`. ISBN: 9798379583248.

D. He and D. J. D. Earn. Epidemiological effects of seasonal oscillations in birth rates. *Theoretical Population Biology*, 72:274–291, 2007. .

D. He and D. J. D. Earn. The cohort effect in childhood disease dynamics. *Journal of the Royal Society of London, Interface*, 13:20160156, 2016. .

D. He, J. Dushoff, T. Day, J. Ma, and D. J. D. Earn. Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proceedings of the Royal Society of London, Series B*, 280(1766):20131345, 2013. .

H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

M. P. Hillmer, P. Feng, J. R. McLaughlin, V. K. Murty, B. Sander, A. Greenberg, and A. D. Brown. Ontario's COVID-19 Modelling Consensus Table: mobilizing scientific expertise to support pandemic response. *Canadian Journal of Public Health*, 112(5):799–806, 2021.

E. Howerton, L. Contamin, L. C. Mullany, M. Qin, N. G. Reich, S. Bents, R. K. Borchering, S.-m. Jung, S. L. Loo, C. P. Smith, et al. Evaluation of the US COVID-19 Scenario Modeling Hub for informing pandemic response under uncertainty. *Nature Communications*, 14(1): 7260, 2023.

P. Johnson. *adaptivetau: Tau-Leaping Stochastic Simulation*, 2023. URL `https://CRAN.`

R-project.org/package=adaptivetau. R package version 2.3.

W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A*, 115:700–721, 1927.

T. Kim, B. Lieberman, G. Luta, and E. A. Peña. Prediction intervals for Poisson-based regression models. *WIREs Computational Statistics*, 14(5):e1568, 2022. ISSN 1939-0068. . URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1568. _eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wics.1568.

A. A. King, M. D. de Cellès, F. M. G. Magpantay, and P. Rohani. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proc. R. Soc. B*, 282(1806):20150347, May 2015. ISSN 0962-8452, 1471-2954. .

A. A. King, D. Nguyen, and E. L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(12):1–43, 2016. . URL https://www.jstatsoft.org/index.php/jss/article/view/v069i12.

C. M. Kribs and P. van den Driessche. Honoring the life and legacy of Fred Brauer. *Journal of Biological Dynamics*, 17(1):2285096, 2023. . URL https://doi.org/10.1080/17513758.2023.2285096.

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

N. P. Lemoine. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128(7):912–928, 2019. ISSN 1600-0706. . URL https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.05985. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/oik.05985.

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

A. Lindén and S. Mäntyniemi. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7):1414–1421, 2011.

W. London and J. A. Yorke. Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *American Journal of Epidemiology*, 98(6):453–468, 1973.

A. G. McKendrick. Applications of mathematics to medical problems. *Proc. Edinburgh Math. Soc.*, 13:98–130, 1926.

K. Nixon, S. Jindal, F. Parker, N. G. Reich, K. Ghobadi, E. C. Lee, S. Truelove, and L. Gardner. An evaluation of prospective COVID-19 modelling studies in the USA: from data to science translation. *The Lancet Digital Health*, 4(10):e738–e747, 2022.

S. Nourbakhsh, A. Fazil, M. Li, C. S. Mangat, S. W. Peterson, J. Daigle, S. Langner, J. Shurgold, P. D'Aoust, R. Delatolla, E. Mercier, X. Pang, B. E. Lee, R. Stuart, S. Wijayasri, and D. Champredon. A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities. *Epidemics*, 39:100560, June 2022. ISSN 1755-4365. .

I. Papst and D. J. D. Earn. Invariant predictions of epidemic patterns from radically different forms of seasonal forcing. *Journal of the Royal Society of London, Interface*, 16:20190202, 2019. . URL https://doi.org/10.1098/rsif.2019.0202.

T. Parsons, B. M. Bolker, J. Dushoff, and D. J. D. Earn. The probability of epidemic burnout in the stochastic SIR model with vital dynamics. *PNAS – Proceedings of the National Academy of Sciences of the U.S.A.*, 121(5):e2313708120, 2024. URL https://www.pnas.org/doi/10.1073/pnas.2313708120.

O. G. Pybus, M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey. The epidemic behavior of the hepatitis C virus. *Science*, 292(5525):2323–2325, 2001.

A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller, and J. Timmer. Lessons learned from quantitative dynamical modeling in systems biology. *PLOS ONE*, 8(9):e74335, Sept. 2013. ISSN 1932-6203. . URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074335.

M. Roberts and J. Heesterbeek. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *Journal of Mathematical Biology*, 55(5):803–816, 2007.

S. L. Rogers. *Special Tables of Mortality from Influenza and Pneumonia, in Indiana, Kansas, and Philadelphia, PA*. Department of Commerce, Bureau of the Census, Washington, DC, 1920.

D. P. Rosati, M. H. Woolhouse, B. M. Bolker, and D. J. D. Earn. Modelling song popularity as a contagious process. *Proceedings of the Royal Society of London, Series A*, 477:20210457, 2021. . URL https://royalsocietypublishing.org/doi/10.1098/rspa.2021.0457.

J. K. Taubenberger and D. M. Morens. 1918 influenza: the mother of all pandemics. *Emerging Infectious Diseases*, 12(1):15–22, 2006.

B. P. Taylor, J. Dushoff, and J. S. Weitz. Stochasticity and the limits to confidence when estimating R_0 of Ebola and other emerging infectious diseases. *Journal of theoretical biology*, 408:145–154, 2016.

The Advisory Committee Appointed by the Secretary of State for India, the Royal Society, and the Lister Institute. Reports on Plague Investigations in India. *The Journal of Hygiene*, 7(6):693–985, 1907. ISSN 00221724. URL http://www.jstor.org/stable/4619420.

N. Tuncer and T. T. Le. Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences*, 299:1–18, May 2018. ISSN 0025-5564. .

J. M. Ver Hoef. Who Invented the Delta Method? *The American Statistician*, 66(2):124–127, 2012. . URL https://doi.org/10.1080/00031305.2012.687494.

J. Wallinga and M. Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007.

L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, NY, Dec. 2010. ISBN 978-1-4419-2322-6.

W. Zha, N. Zhou, G. Li, W. Li, H. Zhang, S. Zhang, M. Chen, R. Feng, T. Li, and Y. LV. Assessment and forecasting the spread of SARS-CoV-2 outbreak in Changsha, China: Based on a SEIAR Dynamic Model, Mar. 2020. URL https://www.researchsquare.com/article/rs-16659/v1. ISSN: 2693-5015.

Table 1. Fits of KM's analytical SIR approximation (2) **to Bombay plague** (see Fig. 1). The KM column lists the parameter values estimated by KM [p. 714]; the `nls` column lists the values estimated by us, using nonlinear least squares with confidence intervals obtained by the Delta method (see Sect. 4). Values for the initial prevalence $I(0)$ and population size N are assumed in order to derive estimates of the standard SIR model parameters from the parameters of KM's approximation (using the indicated equations). Like Bacaër [2012, p. 408], we assume the population of Bombay was $N = 1$ million. We emphasize in this table that γ is the *per capita* removal rate, in order to contrast it with a, the *total* removal rate at the epidemic peak; elsewhere we refer to γ simply as the recovery rate.

Estimated parameter	**symbol**	**equation**	**units**	**KM estimate**	**`nls` estimate**	**95% CI**
total removal rate at epidemic peak	a	(6c)	$\frac{1}{\text{weeks}}$	890	875	(816, 935)
outbreak speed	ω	(6a)	$\frac{1}{\text{weeks}}$	0.2	0.19	(0.178, 0.21)
outbreak centre	ϕ	(6b)	–	3.4	3.37	(3.09, 3.67)
Assumed parameter						
initial prevalence	$I(0)$	–	–	1	1	–
population size	N	–	–	10^6	10^6	–
Derived parameter						
peak time	t_p	(8)	weeks	17	17.4	(17.1, 17.7)
effective reproduction number	$\mathcal{R}_\text{e}$	(5), (10a)	–	1.1	1.09	(1.04, 1.15)
per capita removal rate	γ	(1), 10b	$\frac{1}{\text{weeks}}$	3.96	4.11	(1.95, 6.31)
initial susceptibles	$S(0)$	(10c)	–	53300	57400	(26000, 88800)
transmission rate	β	(1), (11)	$\frac{1}{\text{years}}$	0.00425	0.00407	(0.00372, 0.00443)
mean generation interval	T_g	(12)	days	1.77	1.7	(0.802, 2.59)
basic reproduction number	$\mathcal{R}_0$	(3)	–	20.6	19	(7.65, 30.5)

Table 2. Fits of numerical SIR model solutions to Bombay plague (see Fig. 1). Parameter values were estimated using `fitode` to fit trajectories of Eq. (1), assuming observation errors were negative binomially (`nbinom`) or normally (`ols`) distributed. The recovery rate γ was fixed rather than fitted due to parameter unidentifiability (see Footnote 11); we fixed γ to the value obtained from our `nls` fit of the KM approximation [Table 1] in order to compare fits fairly. The `fitode`-fitted trajectories and confidence bands—for both `nbinom` and `ols`—are shown in the lower panel of Fig. 1. As in Table 1, a population size N must be assumed to derive $\mathcal{R}_0$ estimates.

Fixed parameter	**symbol**	**units**	`nbinom` estimate	**95% CI**	`ols` estimate	**95% CI**
recovery rate	γ	$\frac{1}{\text{weeks}}$	4.11	–	4.11	–
Estimated parameter						
transmission rate	β	$\frac{1}{\text{years}}$	0.004784	(0.00449, 0.00510)	0.0044564	(0.00393, 0.00506)
initial susceptibles	$S(0)$	–	49200	(46200, 52400)	52600	(46700, 59300)
initial prevalence	$I(0)$	–	0.941	(0.76, 1.17)	1.05	(0.627, 1.77)
overdispersion parameter	k	–	48.8	(24.4, 97.7)	–	–
Assumed parameter						
population size	N	–	10^6	–	10^6	–
Derived parameter						
effective reproduction number	$\mathcal{R}_\text{e}$	–	1.1	(1.1, 1.11)	1.1	(1.09, 1.1)
mean generation interval	T_g	days	1.7	–	1.7	–
basic reproduction number	$\mathcal{R}_0$	–	22.4	(21, 23.8)	20.9	(18.2, 23.5)

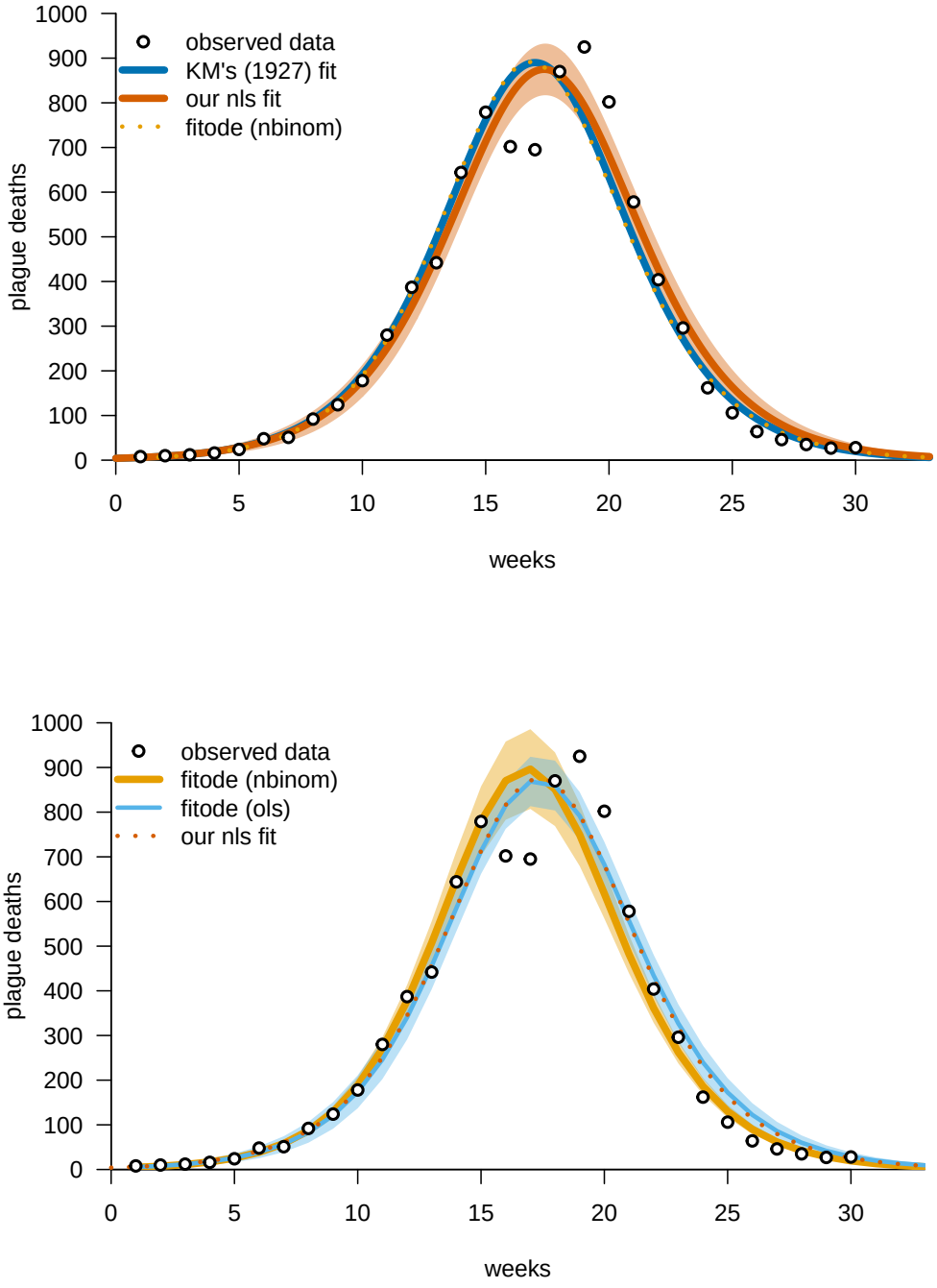


Figure 1. The plague epidemic in Bombay, 17 December 1905 to 21 July 1906, used as an example by KM [p. 714]. The data (large dots) were digitized from The Advisory Committee Appointed by the Secretary of State for India, the Royal Society, and the Lister Institute [1907, Table IX, p. 753]. *Top panel:* The KM approximation (2), as fitted by KM (blue curve) and by us using `nls` (orange curve, with confidence band estimated using the Delta method; see Sect. 4). The associated parameter estimates are given in Table 1. The dotted gold curve shows the `fitode` fit of the SIR model (1), for which the associated parameter estimates are given in Table 2 [observation errors are assumed to be negative binomially distributed (17)]; this curve happens to coincide almost exactly with KM's fit. *Bottom panel:* The solid gold curve is identical to the dotted gold curve in the top panel; its confidence band is the `fitode` confidence band obtained by the Delta method [the band is shown as a linear interpolation between successive observation times because the model (1) is fitted to incidence at discrete time points rather than to a continuous curve representation of the instantaneous death rate]. The light blue curve shows the `fitode` fit obtained by minimizing the ordinary least squares (7) [i.e., assuming observation errors are normally (14) distributed with variance σ^2 estimated from the residuals across all observation times]. The dotted orange curve is identical to the solid orange curve in the top panel. We have separated the two panels because the confidence band overlap would make the plots difficult to interpret.

Table 3. Fits of KM's analytical SIR approximation (2) **to Philadelphia flu** (see Fig. 2). Parameter estimates were obtained using nonlinear least squares (`nls`) to fit Eq. (2) to the reported daily pneumonia and influenza (P&I) mortality during the main wave of the pandemic in 1918. In order to derive estimates of the standard epidemiological parameters, we assumed the initial prevalence had the value estimated by `fitode` for the SIR model (see Table 4). We do not use the raw population size in our estimate of $\mathcal{R}_0$; instead, we account for the fact that reported deaths are roughly equal to incidence times the case fatality proportion (CFP) by taking N to be the size of population that would eventually die if everyone in the city were infected, i.e., the product of the population size of Philadelphia in 1918 (1,768,825) and an assumed CFP of 0.025 [Taubenberger and Morens, 2006]. The fitted trajectory and confidence band are shown in Fig. 2. See Sect. 7.

Estimated parameter	**symbol**	**equation**	**units**	`nls`	**95% CI**
total removal rate at epidemic peak	a	(6c)	$\frac{1}{\text{years}}$	738	(715, 761)
outbreak speed	ω	(6a)	$\frac{1}{\text{years}}$	42	(40.3, 43.5)
outbreak centre	ϕ	(6b)	–	3.64	(3.51, 3.79)
Assumed parameter					
initial prevalence	$I(0)$	–	–	3.05	–
effective population size	N	–	–	44,221	–
Derived parameter					
peak time	t_p	(8)	weeks	4.524	(4.49, 4.56)
effective reproduction number	$\mathcal{R}_\text{e}$	(5), (10a)	–	128	(90.8, 165)
per capita removal rate	γ	(10b)	$\frac{1}{\text{years}}$	0.66	(0.492, 0.831)
initial susceptibles	$S(0)$	(10c)	–	2270	(1660, 2880)
transmission rate	β	(11)	$\frac{1}{\text{years}}$	0.0372	(0.0285, 0.0459)
mean generation interval	T_g	(12)	years	1.52	(1.12, 1.9)
basic reproduction number	$\mathcal{R}_0$	(3)	–	2490	(2416, 2571)

Table 4. Fits of numerical SIR model solutions to Philadelphia flu (see Fig. 2). Parameter estimates are based on `fitode` fits of the SIR model (1) to reported P&I mortality during the main wave of the 1918 influenza pandemic in the city of Philadelphia. As in Table 3, in order to derive an estimate of $\mathcal{R}_0$, we assume an effective population size that accounts for the data representing deaths rather than cases.

Estimated parameter	**symbol**	**units**	`nbinom`	**95% CI**
transmission rate	β	$\frac{1}{\text{years}}$	0.0124	(0.0119, 0.0128)
recovery rate	γ	$\frac{1}{\text{years}}$	85.6	(75.9, 96.5)
initial susceptibles	$S(0)$	–	15300	(14500, 16200)
initial prevalence	$I(0)$	–	3.05	(2.32, 4.01)
overdispersion parameter	k	–	157	(44.2, 557)
Assumed parameter				
effective population size	N	–	44,221	–
Derived parameter				
effective reproduction number	$\mathcal{R}_\text{e}$	–	2.21	(2.02, 2.4)
mean generation interval	T_g	days	4.27	(3.75, 4.78)
basic reproduction number	$\mathcal{R}_0$	–	6.38	(5.53, 7.24)

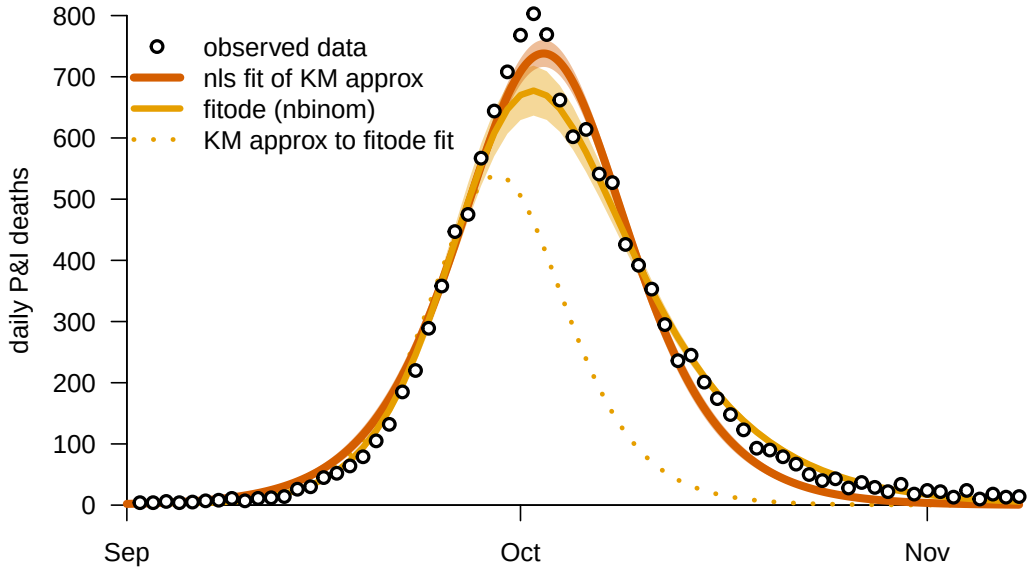


Figure 2. The main wave of the 1918 influenza epidemic in the city of Philadelphia, 1 September 1918 to 31 December 1918 [Rogers, 1920, Goldstein et al., 2009]. Reported daily deaths from pneumonia and influenza (P&I) are shown with large dots. The orange curve and corresponding confidence band show a nonlinear least squares (`nls`) fit of KM's approximation (2); the parameter estimates are given in Table 3. The solid gold curve and corresponding confidence band show the `fitode` fit of the SIR model (1), for which the parameter estimates are given in Table 4. The dotted gold curve shows the KM approximation using the parameters estimated with `fitode`.

Table 5. Fits of KM's analytical SIR approximation (2) **to an epidemic simulated using the standard stochastic SIR model [Andersson and Britton, 2000]** (see Sect. 8 and Fig. 3). The parameter values in the "true" column are those used to generate the stochastic simulation ($S(0)$, $I(0)$, $\mathcal{R}_0$ and T_g) and the values of other parameters derived from these true parameter values using the indicated equations. The `nls` column lists our estimates and confidence intervals obtained by fitting Eq. (2) to the simulated data using nonlinear least squares and the Delta method.

Assumed parameter	**symbol**	**equation**	**units**	**true**	`nls`	**95% CI**
initial prevalence	$I(0)$	–	–	2	2	–
Estimated parameter						
total removal rate at epidemic peak	a	(6c)	$\frac{1}{\text{weeks}}$	641	135	(125, 144)
outbreak speed	ω	(6a)	$\frac{1}{\text{weeks}}$	2	0.99	(0.907, 1.08)
outbreak centre	ϕ	(6b)	–	3.58	2.7	(2.48, 2.93)
Derived parameter						
peak time	t_p	(8)	weeks	1.79	2.72	(2.66, 2.78)
effective reproduction number	$\mathcal{R}_\mathrm{e}$	(5), (10a)	–	5	2.62	(1.79, 3.44)
per capita removal rate	γ	(10b)	$\frac{1}{\text{weeks}}$	1	1.21	(0.7, 1.77)
initial susceptibles	$S(0)$	(10c)	–	2000	571	(518, 624)
transmission rate	β	(11)	$\frac{1}{\text{years}}$	0.13	0.289	(0.233, 0.346)
mean generation interval	T_g	(12)	days	7	5.77	(3.2, 8.12)
basic reproduction number	$\mathcal{R}_0$	(3)	–	5	9.18	(6.95, 11.4)

Table 6. Fits of numerical (deterministic) SIR model solutions to an epidemic simulated using the standard stochastic SIR model [Andersson and Britton, 2000] (see Sect. 8 and Fig. 3). Parameter estimates we obtained using `fitode` to fit the SIR model (1) to the simulated data, assuming deviations from the deterministic curve were generated by negative binomially (17) distributed observation errors.

Estimated parameter	**symbol**	**units**	**true**	`nbinom`	**95% CI**
transmission rate	β	$\frac{1}{\text{years}}$	0.13	0.131	(0.119, 0.144)
recovery rate	γ	$\frac{1}{\text{weeks}}$	1	0.971	(0.884, 1.07)
initial susceptibles	$S(0)$	–	1998	2000	(1900, 2110)
initial prevalence	$I(0)$	–	2	0.605	(0.306, 1.2)
overdispersion parameter	k	–	–	251	(19.6, 3226.6)
Derived parameter					
mean generation interval	T_g	days	7	7.21	(7.92, 6.56)
effective reproduction number	$\mathcal{R}_\text{e}$	–	4.995	5.2	(4.44, 5.95)
basic reproduction number	$\mathcal{R}_0$	–	5	5.19	(4.37, 6.01)

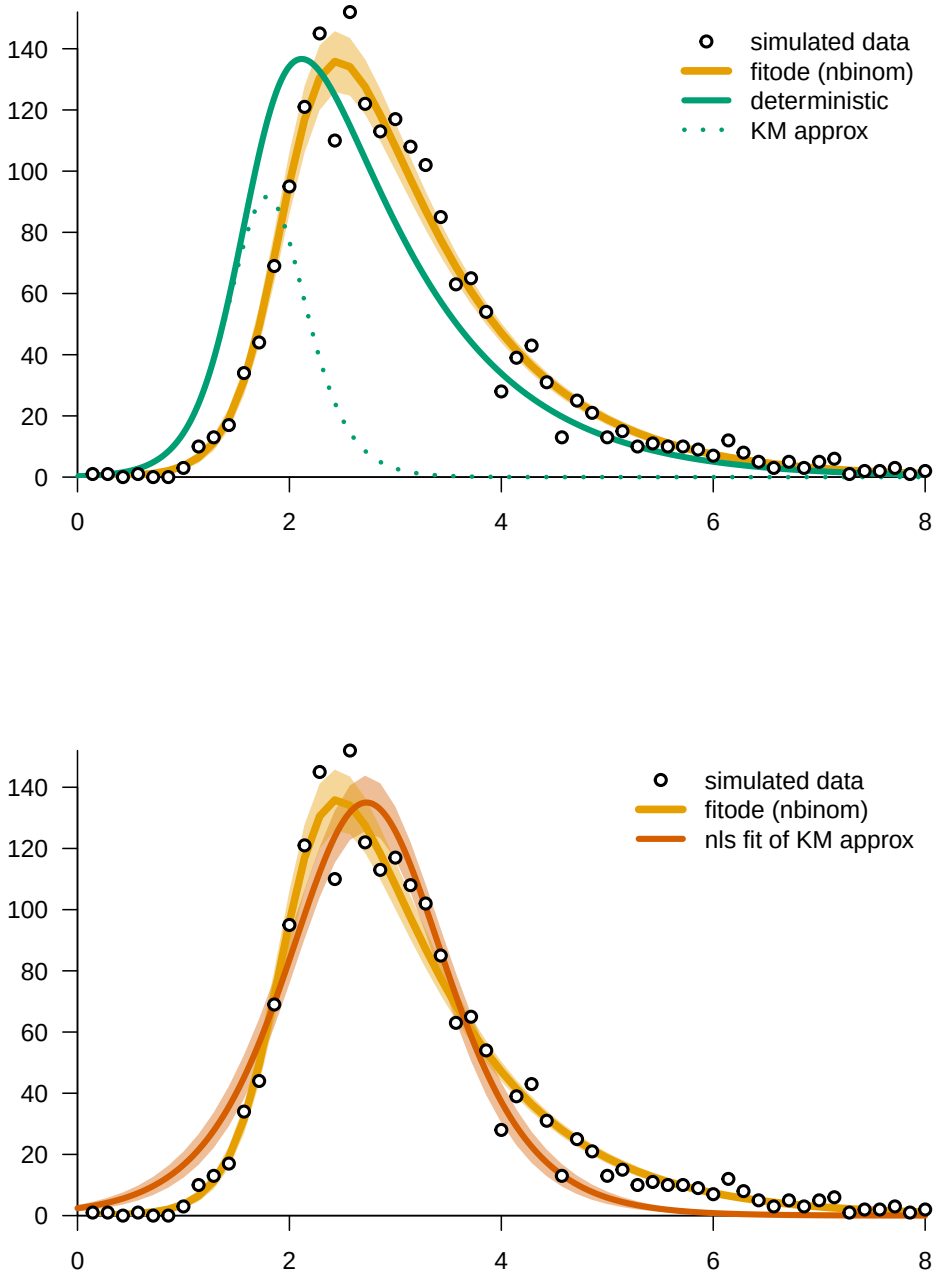


Figure 3. Deterministic fits to daily incidence generated by a stochastic SIR simulation with initial state $(S(0), I(0), R(0)) = (1998, 2, 0)$, basic reproduction number $\mathcal{R}_0 = 5$, and mean generation interval $T_g = 1$ week. The simulated data points show the numbers of newly recovered individuals each day. In both panels, the gold curve and confidence band show the `fitode` fit to the simulated data. *Top panel:* The solid green curve shows the solution of deterministic SIR model (1) with the initial conditions and parameters used for the stochastic simulation. The dotted green curve shows the KM approximation (2) to this deterministic trajectory. The time shift between the green and gold curves arises because there is a random delay until the stochastic trajectory begins to grow exponentially. *Bottom panel:* The orange curve shows the KM approximation (2), fitted to the stochastic simulation using `nls`. Since the KM approximation is symmetric about its maximum, it is impossible to obtain a good fit in situations like this, where the rise of the epidemic is faster than the fall.

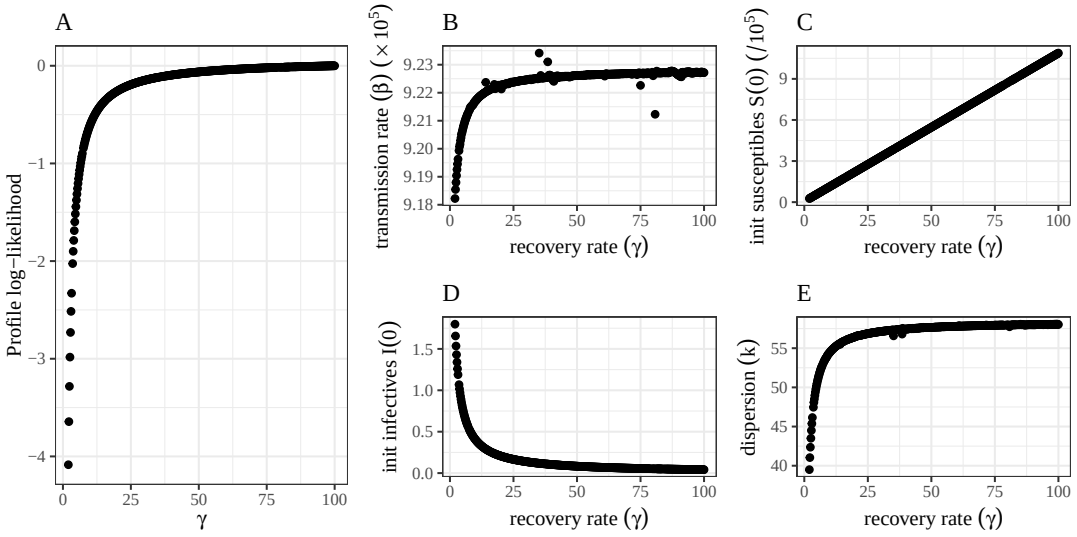


Figure 4. Unidentifiability of the mean generation interval T_g (or, equivalently, the *per capita* removal rate γ) for the Bombay plague epidemic shown in Fig. 1. (A) The profile likelihood—briefly discussed at the end of Sect. 4—is calculated by fixing γ to a series of given values and, for each value, maximizing the likelihood by estimating all other parameters [Bolker, 2008]. (The maximum value is shifted to 0 without loss of generality.) A flat profile-likelihood surface indicates parameter unidentifiability, meaning that we can obtain very similar fits across a wide range of values of the focal parameter (γ). (B–E) The corresponding best parameter estimates for a given value of γ.

Fred Brauer was an eminent mathematician who studied dynamical systems, especially differential equations. He made many contributions to mathematical epidemiology, a field that is strongly connected to data, but he always chose to avoid data analysis. Nevertheless, he recognized that fitting models to data is usually necessary when attempting to apply infectious disease transmission models to real public health problems. He was curious to know how one goes about fitting dynamical models to data, and why it can be hard. Initially in response to Fred's questions, we developed a user-friendly \code{R} package, \code{fitode}, that facilitates fitting ordinary differential equations to observed time series. Here, we use this package to provide a brief tutorial introduction to fitting compartmental epidemic models to a single observed time series. We assume that, like Fred, the reader is familiar with dynamical systems from a mathematical perspective, but has limited experience with statistical methodology or optimization techniques.

ARTICLE TYPE: EDUCATION

Fitting epidemic models to data – a tutorial in memory of Fred Brauer

David J. D. Earn[a], Sang Woo Park[b] and Benjamin M. Bolker[a,c],

[a]Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada, L8S 4K1;

[b]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544;

[c]Department of Biology, McMaster University, Hamilton, Ontario, Canada, L8S 4K1

Article compiled:
June 4, 2024

ABSTRACT

Fred Brauer was an eminent mathematician who studied dynamical systems, especially differential equations. He made many contributions to mathematical epidemiology, a field that is strongly connected to data, but he ~~had no desire to touch data~~ always chose to avoid data analysis. Nevertheless, he recognized that fitting models to data is usually necessary when attempting to apply infectious disease transmission models to real public health problems. He was curious to know how one goes about fitting dynamical models to data, and why it can be hard. Initially in response to Fred's questions, we developed a user-friendly `R` package, `fitode`, that facilitates fitting ordinary differential equations to observed time series. Here, we use this package to provide a brief tutorial introduction to fitting compartmental epidemic models to a single observed time series. We assume that, like Fred, the reader is familiar with dynamical systems from a mathematical perspective, but has limited experience with statistical methodology or optimization techniques.

KEYWORDS

epidemic models; infectious diseases; ordinary differential equations; parameter estimation; maximum likelihood; `fitode`

Submitted to *Bulletin of Mathematical Biology*

1. Introduction

In their landmark 1927 paper, ~~Kermack and McKendrick~~ Kermack and McKendrick [1927, p. 713] (KM) ~~[Kermack and McKendrick, 1927, p. 713]~~

D.J.D. Earn. earn@math.mcmaster.ca. ORCID: 0000-0002-7562-1341
S.W. Park. swp2@princeton.edu. ORCID 0000-0003-2202-3361,
B.M. Bolker. bolkerb@mcmaster.ca. ORCID: 0000-0002-2127-0443

introduced the now-standard susceptible-infected-removed (SIR) epidemic model,

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI\,, \tag{1a}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI - \gamma I\,, \tag{1b}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I\,, \tag{1c}$$

where S, I and R represent the numbers of individuals who are susceptible, infected or removed[1], β is the transmission rate, and γ is the ~~recovery~~ removal (or recovery) rate. In that original paper, ~~[Kermack and McKendrick, 1927, p. 714]~~ KM [p. 714] also fit their model to plague mortality data from an epidemic in Bombay (now Mumbai) that occurred about 20 years before their paper was written.

In the century that has elapsed since publication of KM's initial paper, the field of mathematical epidemiology has expanded and matured, and has been the subject of many books ~~Bartlett [1960], Bailey [1975], Anderson and May [1991], Andersson and Britton [2000], Diekmann and H~~ [Bartlett, 1960, Bailey, 1975, Anderson and May, 1991, Andersson and Britton, 2000, Diekmann and Hee and review articles ~~Hethcote [2000], Earn et al. [2002], Earn [2008, 2009]~~ [Hethcote, 2000, Earn et al., 2002, Earn, 2008, 2009]. Researchers have primarily focused on ***compartmental models*** like the SIR model, cast either as differential equations following the tradition of KM ~~Kermack and McKendrick [1927]~~, or as ***stochastic processes*** in the tradition of ~~McKendrick McKendrick [1926] and Bartlett Bartlett [1960]~~McKendrick [1926] and Bartlett [1960]. In recent years, as the power of computers has grown, some researchers have turned to ***agent-based models***, which represent each individual as a separate unit that can have unique properties ~~Eubank et al. [2004]~~[Eubank et al., 2004].

Throughout the history of the subject, and regardless of the modelling frameworks they have exploited, mathematical epidemiologists have frequently attempted to fit—or at least to compare—their models to observed infectious disease data. Such fits have often been naïve, with limited consideration of their quality. Over the years, however, there has been a trend towards greater sophistication and statistical rigour in parameter estimation for infectious disease models; books that explain these methods have begun to appear in recent decades ~~Bolker [2008], Bjørnstad [2018]~~ [Bolker, 2008, Bjørnstad, 2018]. Careful consideration of uncertainty is especially important when epidemic models are used for the development and analysis of policy options for infectious disease management ~~Elderd et al. [2006]~~[Elderd et al., 2006], a challenge that began to absorb the attention of many mathematical epidemiologists as soon as the emergence of SARS-CoV-2 ignited the COVID-19 pandemic ~~Brooks-Pollock et al. [2021], Hillmer et al. [2021], Nixon et al. [2022]~~ [Brooks-Pollock et al., 2021, Hillmer et al., 2021, Nixon et al., 2022, Howerton et al., 2023].

While visiting the University of British Columbia in 2014–2015, one of us (DE) had many conversations with Fred Brauer about epidemic models and how they can be used in practical applications. While he had no desire to analyze data himself, Fred was acutely aware that fitting to data is essential if one wishes to apply epidemic models

[1] In the words of ~~[Kermack and McKendrick, 1927, p. 701]~~KM [p. 701] , "removed from the number of those who are sick, by recovery or by death".

to real public health problems, and he did want to understand what was involved in doing so.

Fred's curiosity inspired us to develop user-friendly software for fitting ordinary differential equation (ODE) models to observed time series, with the goal of illustrating the process and challenges of model fitting to Fred and others like him, i.e., individuals who are comfortable with mathematical analysis of ODEs but have little or no experience with statistics and parameter estimation. Unfortunately, we have lost the opportunity to present our work to Fred, but it seems fitting (!) to highlight Fred's role in the history of this work, and to dedicate this tutorial to his memory.[2]

2. Kermack and McKendrick's fit

We begin by revisiting KM's ~~Kermack and McKendrick [1927]~~ application of their SIR model (1) to the epidemic of plague in Bombay in 1905–1906. The observed data (dots in Fig. 1) were weekly numbers of deaths from plague.

Referring to their version of Fig. 1, ~~[Kermack and McKendrick, 1927, p. 714]~~ KM [p. 714] argued that "As at least 80 to 90 per cent. of the cases reported terminate fatally, the ordinate may be taken as approximately representing $[\mathrm{d}R/\mathrm{d}t]$ as a function of t." Since (non-human) computers did not yet exist ~~Campbell-Kelly [2009]~~ [Campbell-Kelly, 2009], and an exact analytical form for this function could not be found, they proceeded to assume ~~[Kermack and McKendrick, 1927, p. 713]~~ [KM, p. 713] that $\frac{\beta}{\gamma}R(t) \ll 1$, which yields the approximate analytical form,

$$\frac{\mathrm{d}R}{\mathrm{d}t} \approx a\,\mathrm{sech}^2(\omega\,t-\phi)\,. \tag{2}$$

Noting that the ***basic reproduction number*** is[3]

$$\mathcal{R}_0\mathcal{R}_0 = \frac{N\beta}{\gamma}\,, \tag{3}$$

where N is the total population size, the assumption that yields KM's approximation (2) can be written

$$\frac{R(t)}{N}\frac{R(t)-R(0)}{N} \ll \frac{1}{\mathcal{R}_0}\frac{1}{\mathcal{R}_0}\,, \tag{4}$$

~~i.e.~~ (KMassumed $R(0)=0$); thus, Eq. (2) is a good approximation as long as the proportion of the population that has been infected and removed since the initial time is much less than ~~$1/\mathcal{R}_0$.~~ $1/\mathcal{R}_0$.

Given Eq. (3), the ***effective reproduction number*** at time $t=0$ is

$$\mathcal{R}_\mathrm{e} = \frac{S_0\beta}{\gamma}\frac{S(0)\beta}{\gamma}\,. \tag{5}$$

[2] We had originally intended to submit this paper to a collection in honour of Fred's memory ~~Kribs and van den Driessche [2023]~~[Kribs and van den Driessche, 2023].

[3] ~~$\mathcal{R}_0$~~ $\mathcal{R}_0$ is the expected number of secondary cases resulting from a primary case in a wholly susceptible population ~~Anderson and May [1991]~~[Anderson and May, 1991].

In terms of $\mathcal{R}_e$, γ, ~~S_0 and I_0~~$S(0)$ and $I(0)$, the parameters in Eq. (2) can be written[4]

$$\omega = \frac{\gamma}{2}\sqrt{(\mathcal{R}_e - 1)^2 + \frac{2I_0}{S_0}\mathcal{R}_e^2}\sqrt{(\mathcal{R}_e - 1)^2 + \frac{2I(0)}{S(0)}\mathcal{R}_e^2}\,, \tag{6a}$$

$$\phi = \operatorname{arctanh}\left(\frac{\mathcal{R}_e - 1}{2\,\omega/\gamma}\right)\,, \tag{6b}$$

$$\text{and} \qquad a = \frac{2\,\omega^2 S_0}{\gamma\,\mathcal{R}_e^2}\frac{2\,\omega^2 S(0)}{\gamma\,\mathcal{R}_e^2}\,. \tag{6c}$$

The values of these parameters that KM estimated for the Bombay plague epidemic are listed in the KM column of Table 1. Using these values, KM plotted their "calculated" curve, which we have reproduced in blue in Fig. 1.

3. How to fit the model to the data

The blue curve in Fig. 1 does appear to provide a reasonable fit to the data, but KM ~~Kermack and McKendrick [1927]~~ gave no indication of how their parameter estimates were obtained. Whatever their process, they must have engaged in some sort of ***trajectory matching***, i.e., adjusting parameter values until the model—Eq. (2) in their case—is, by some measure, close to the observed data points. The most obvious metric for this purpose is the Euclidean distance between the model curve and the data. Thus, a natural ***objective function*** to minimize is

$$\sum_{\ell=1}^{n_t}\big(x(t_\ell;\boldsymbol{\theta}) - x[t_\ell]x_{\text{obs}}(t_\ell)\big)^2\,, \tag{7}$$

where the observed data are the points ~~$\{(t_\ell, x[t_\ell]) : \ell = 1,\ldots,n_t\}$~~$\{(t_\ell, x_{\text{obs}}(t_\ell)) : \ell = 1,\ldots,n_t\}$, $\boldsymbol{\theta}$ is the vector of parameters, and $x(t;\boldsymbol{\theta})$ is the model; for KM's problem, the parameter vector is $\boldsymbol{\theta} = (a, \omega, \phi)$ and the model is given by Eq. (2). (Note that we write ~~$x[\cdot]$~~ $x_{\text{obs}}(\cdot)$ when referring to observations of the variable x and $x(\cdot\,;\cdot)$ when referring to the model.) Choosing this objective function is equivalent to assuming that the $x_{\text{obs}}(t_\ell)$ values are direct (but noisy) observations of the state variable $x(t)$. When the connection between the dynamical system and our observations is more complicated, we need to define an explicit ***observation process***; see Sect. 4. Minimizing (7) with respect to $\boldsymbol{\theta}$ would have required some heroic arithmetic with a pencil and paper in 1927, but it is a simple task with the aid of a modern computer.

In the following segment of R code, we fit equation (2) to the Bombay plague data (which are included in the `fitode` package that we describe below, as a data frame with columns `week` and `mort`). We exploit R's nonlinear least squares function (`nls`), which attempts to minimize the distance (7) to the data, starting from an initial guess (`start`).

[4] There is a typographical error in equation (31) of KM ~~Kermack and McKendrick [1927]~~: their factor $\sqrt{-q}$ should be $(-q)$ in their equivalent of the parameter we call a. ~~Bacaër [Bacaër, 2012, §3]~~ Bacaër [2012, §3] corrected this error without comment.

```
sech <- function(x) {1/cosh(x)}
KM_approx <- function(t, a, omega, phi) {a * sech(omega*t - phi)^2}
KM.parameters <- c(a = 890, omega = 0.2, phi = 3.4)
nlsfit <- nls(mort ~ KM_approx(week, a, omega, phi),
              data = fitode::bombay,
              start = KM.parameters)
nls.parameters <- coef(nlsfit)
print(nls.parameters)

##           a       omega         phi
## 874.7545749   0.1935916   3.3720557
```

Above, we chose as our starting value the fitted parameter values of KM. Our least squares parameter values differ from KM's by a few percent (see Table 1). The least squares fitted function is shown in orange in Fig. 1.

Starting from someone else's fit is not a great way to test the method, but fortunately the least squares fit for this problem is not very sensitive to the starting value. To pick reasonable starting values, it often helps to think about the meaning of parameters. For example, in the case of Eq. (2), it is useful to note that a is the maximum of the function, and if write $\omega t - \phi$ as $\omega(t - t_\mathrm{p})$ then

$$t_\mathrm{p} = \frac{\phi}{\omega} \tag{8}$$

is the ***peak time*** (at which the maximum occurs); both a and t_p can be approximated by looking at the plotted data. Assuming ~~$I_0/S_0 \ll 1$~~$I(0)/S(0) \ll 1$, ω is half the initial exponential growth rate[5], so it can be approximated easily by plotting the data on a log scale, estimating the initial slope, and dividing by 2. Very rough guesses for a, t_p and ω are sufficient to converge on the same fit:

```
a.guess <- 1000    # crude "by eye" estimate of peak value,
tpeak.guess <- 15  # peak time,
omega.guess <- 1   # and half the initial growth rate
phi.guess <- omega.guess * tpeak.guess
nlsfit <- nls(mort ~ KM_approx(week, a, omega, phi),
              data = fitode::bombay,
              start = c(a = a.guess, omega = omega.guess,
                        phi = phi.guess))
print(nls.parameters <- coef(nlsfit))

##           a       omega         phi
## 874.7550490   0.1935918   3.3720589
```

However, if you experiment with starting values, you will find that if you pick sufficiently *bad* starting values, then `nls` will fail. For example, starting from $a = 2000$, $t_\mathrm{p} = 5$, and $\omega = 0.1$ yields a "singular gradient" error. More interestingly, starting from $a = 500$, $t_\mathrm{p} = 5$, and $\omega = 0.1$ yields $a = 869$, $\omega = -0.19$, $\phi = -3.48$, which is far from our fitted values and illustrates ~~a very~~ an important fact: there is ~~*not necessarily a unique best fit set of parameters!*~~ not necessarily a unique best fit set of parameters!

[5]From Eq. (1b), the initial exponential growth rate is ~~$\beta S_0 - \gamma = \gamma(\mathcal{R}_\mathrm{e} - 1)$~~$\beta S(0) - \gamma = \gamma(\mathcal{R}_\mathrm{e} - 1)$.

In this case, the alternative solution exists because $\operatorname{sech}^2(x)$ is symmetric about the y axis, but in general, there can be multiple local minima that cause nonlinear optimizers to converge to points that may or may not represent equally good fits to the data. The potential existence of multiple local optima makes fitting to data hard; you need to be cautious, and use common sense, in interpreting the solutions found by your software (always plot the solutions!). ~~Raue *et al.* Raue et al. [2013]~~ Raue et al. [2013] give suggestions for how to diagnose and handle multiple optima.

If you know that your parameters should be in a certain range, then you can exclude values outside that range. For example, to ensure that all the parameters are non-negative (and exclude the alternative fit above), you would add the `nls` option

```
lower = c(a = 0, omega = 0, phi = 0)
```

which would prevent convergence to negative ω and ϕ. Alternatively, you could write

$$a = e^A, \quad \omega = e^\Omega, \quad \phi = e^\Phi, \tag{9}$$

and fit A, Ω, and Φ, which would guarantee positive a, ω, and ϕ without having to constrain the values of the fitted parameters. While this last suggestion may just seem like a cute trick, there is more to it than that. Many more optimization algorithms are available for unconstrained fitting; numerical parameter values of very small magnitude can also lead to numerical instability, so it is advantageous to link parameters that must lie in a given range to unconstrained parameters that can be fit more easily [Bolker, 2008, pp. 328–329]. In Eq. (9), the ***link function*** that converts the parameters to the unconstrained scale is $\log(x)$. Another common link function is $\operatorname{logit}(x) = \log(x/(1-x))$ (the log-odds function, or the inverse of the logistic function), which converts the unit interval $(0, 1)$ to $(-\infty, \infty)$, and is convenient when parameters represent proportions or probabilities. (Requiring positivity is so common that `fitode` uses a log link for all parameters by default.)

If we accept our fit as satisfactory, what can we infer about the dynamics of plague that KM were attempting to capture with the SIR model (1)? We need to convert the parameters of KM's approximation (6) back to the original parameters that are directly related to the mechanism of disease spread formalized by the model (i.e., β and γ, and initial conditions ~~S_0 and I_0~~$S(0)$ and $I(0)$).

The nonlinear algebraic relationships specified by Eq. (6) can be inverted[6] analytically[7] ~~[Bacaër, 2012, §3]~~[Bacaër, 2012, §3], to obtain

$$\mathcal{R}_\mathrm{e} = 1 + \frac{2\,\omega I_0 \sinh(\phi)\cosh(\phi)}{a}\,\frac{2\,\omega I(0)\sinh(\phi)\cosh(\phi)}{a}, \tag{10a}$$

$$\gamma = \frac{2\,\omega \tanh\phi}{\mathcal{R}_\mathrm{e} - 1}, \tag{10b}$$

$$S_0 S(0) = \frac{2\,\mathcal{R}_\mathrm{e}^2\, I_0 \sinh^2\phi}{(\mathcal{R}_\mathrm{e}-1)^2}\,\frac{2\,\mathcal{R}_\mathrm{e}^2\, I(0)\sinh^2\phi}{(\mathcal{R}_\mathrm{e}-1)^2}. \tag{10c}$$

[6] Our expressions are slightly different from those of ~~Bacaër [Bacaër, 2012, eq. (3)]~~ Bacaër [2012, eq. (3)] because we have corrected a minor error. At the start of §3 of ~~Bacaër [2012]~~Bacaër [2012], in the expression for Q, the term ~~$2Ry_0/x_0$~~ $2Ry_0/x_0$ should be $2R^2 y_0/x_0$~~and ;~~ this missing square is propagated through to the inversion formulae.

[7] In (common) situations in which nonlinear algebraic equations cannot be solved analytically, they can still be solved numerically, for example with the `nleqslv` package in R.

Since there are four original parameters (β, γ, ~~S_0, I_0~~$S(0)$, $I(0)$) and only three parameters in KM's approximation (2) (a, ω, ϕ), one of the four original parameters needs to be specified separately; in Eq. (10) above we have taken this to be the initial prevalence ~~I_0~~$I(0)$. From Eq. (10), we can compute the transmission rate,

$$\beta = \frac{\mathcal{R}_\mathrm{e}\gamma}{S_0}\frac{\mathcal{R}_\mathrm{e}\gamma}{S(0)}\,, \tag{11}$$

and the mean intrinsic generation interval ~~Champredon and Dushoff [2015]~~ [Champredon and Dushoff, 2015],

$$T_\mathrm{g} = \frac{1}{\gamma}\,, \tag{12}$$

which is the same as the mean infectious period in this simple model ~~Krylova and Earn [2013], Champredon et al. [2018]. Table 1~~ [Pybus et al., 2001, Roberts and Heesterbeek, 2007, Wallinga and Lipsitch, 2007, Krylova and Earn, 2013]. Table 1 lists the values of the parameters as estimated by KM and by us using `nls`.

Correctly handling weekly mortality. We have glossed over the fact that we have fitted observed weekly mortality to the *instantaneous* rate, $\mathrm{d}R/\mathrm{d}t$ (2), which is not observed. We did this because it is what KM did, and we wanted to be able to compare formal nonlinear least squares fits to KM's results[8]. Weekly mortality reported at time t_ℓ should really be modelled as the aggregation of $\mathrm{d}R/\mathrm{d}t$ over the preceding week, i.e., it would be better to define

$$x(t_\ell;\boldsymbol{\theta}) = \int_{t_{\ell-1}}^{t_\ell} \frac{\mathrm{d}R}{\mathrm{d}t}\,\mathrm{d}t \tag{13a}$$

$$= \int_{t_{\ell-1}}^{t_\ell} a\,\mathrm{sech}^2(\omega\,t-\phi)\,\mathrm{d}t \tag{13b}$$

$$= \frac{a}{\omega}\Big(\tanh(\omega\,t_\ell-\phi)-\tanh(\omega\,t_{\ell-1}-\phi)\Big)\,. \tag{13c}$$

Indeed, whether we are fitting to mortality or incidence or another instantaneous rate, we should be integrating over the observation interval, which is precisely what we do below when fitting to the ODEs directly. In addition, we really ought to consider the fact that not all infections end in death—we have followed KM in assuming that the ***infection fatality proportion*** is 100%. Similarly, when analyzing incidence data, the ***reporting proportion*** ought to be taken into account.

4. Uncertainty

To this point, we have addressed only an optimization problem. We solved it using the method of nonlinear least squares, which yields estimates of the values of the parameters of the model (2). But our best estimates are just that: *estimates*, not known values of the parameters.

[8] In his reanalysis of KM's results, ~~Bacaër Bacaër [2012]~~ Bacaër [2012] also retained this conceptual error.

To quantify uncertainty in our estimates, we need a statistical framework. The typical output of such a framework is a ***confidence interval*** (CI) within which our best estimate lies. For example, the final column of Table 1 lists 95% CIs on our `nls` parameter estimates, and the ~~pink~~ light orange shaded region in the top panel of Fig. 1 is a 95% ***confidence band***, which shows CIs for each point of the fitted model curve.

To understand how to estimate CIs, we will start by thinking about our observation model, the probability of observing the data ~~$\{x[t_\ell]\}$~~ $\{x_{\text{obs}}(t_\ell)\}$ given the model trajectory $x(t;\boldsymbol{\theta})$. We imagine that the ~~model (2) is~~ model—for now, KM's approximation (2)—is a perfect representation of reality, and we consider the deviations from the model curve in Fig. 1 to be observation errors. ~~We then imagine that~~ A simple observation model assumes that the observation error for each data point is independent and identically distributed (iid), and drawn from a Normal distribution with zero mean and standard deviation σ equal to the standard deviation of the residuals (the differences between the model curve and the observed data). Then the ~~probability~~ joint probability density $\mathbb{p}$ of the data given the model is

$$\mathbb{P}(\text{data} \mid \text{model}) = \prod_{\ell=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\big(x(t_\ell;\boldsymbol{\theta}) - x[t_\ell]\big)^2}{2\sigma^2} \right) \Delta x[t_\ell] \right] .$$

$$\mathbb{p}(\text{data} \mid \text{model}) = \mathbb{p}(\{x_{\text{obs}}(t_\ell)\} \mid \boldsymbol{\theta}) \tag{14a}$$

$$= \prod_{\ell=1}^{n} \left[\lim_{\Delta x_\ell \to 0} \frac{\mathbb{P}\big(x(t_\ell;\boldsymbol{\theta}) \le x_{\text{obs}}(t_\ell) < x(t_\ell;\boldsymbol{\theta}) + \Delta x_\ell\big)}{\Delta x_\ell} \right] \tag{14b}$$

$$= \prod_{\ell=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\big(x(t_\ell;\boldsymbol{\theta}) - x_{\text{obs}}(t_\ell)\big)^2}{2\sigma^2} \right) \right] . \tag{14c}$$

Note that we write $\mathbb{P}$ for the probability measure and $\mathbb{p}$ for the probability density above. We use a probability density function here because the Normal is a continuous distribution; we would use a probability mass function for a discrete response distribution such as the Poisson. In practice, we don't have to worry about this distinction when we are estimating the parameters of an epidemic model (the elements Δx_ℓ will always appear as constant multipliers or divisors and don't affect any of our conclusions). Consequently, in the interests of brevity, below we interpret $\mathbb{p}$ as either probability mass or probability density, depending on whether the associated distribution is discrete or continuous, and refer simply to "probability".

Using these assumptions we can adopt a ~~*maximum likelihood*~~ ***maximum likelihood*** framework, where we consider parameter values that maximize the probability of observing the data (14) to be the best ~~Bolker [2008]~~[Bolker, 2008]. We define the ***likelihood*** $\mathcal{L}$ of a set of parameter values $\boldsymbol{\theta}$ as

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{P}\mathbb{p}(\{x[t_\ell] x_{\text{obs}}(t_\ell)\} \mid \boldsymbol{\theta}) . \tag{15}$$

Maximizing $\mathcal{L}$ with respect to $\boldsymbol{\theta}$ or, equivalently, minimizing the negative log-

likelihood, yields an estimate,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \tag{16a}$$

$$= \arg\min_{\boldsymbol{\theta}} \big(-\log \mathcal{L}(\boldsymbol{\theta})\big) \tag{16b}$$

$$= \arg\min_{\boldsymbol{\theta}} \Big(\sum_{\ell=1}^{n_t} \big(x(t_\ell;\boldsymbol{\theta}) - x[t_\ell] x_{\text{obs}}(t_\ell)\big)^2 \;+\; \text{constant} \Big) \tag{16c}$$

$$= \arg\min_{\boldsymbol{\theta}} \sum_{\ell=1}^{n_t} \big(x(t_\ell;\boldsymbol{\theta}) - x[t_\ell] x_{\text{obs}}(t_\ell)\big)^2 , \tag{16d}$$

which—lo and behold—agrees exactly with (7), the ~~least squares~~ ***ordinary least squares*** (OLS) solution! The standard way of expressing this is to say that the ~~least squares~~ OLS solution $\hat{\boldsymbol{\theta}}$ is the ***maximum likelihood estimate*** (MLE) of $\boldsymbol{\theta}$, under the assumption of independent, identically distributed (i.e., mean-zero, constant-variance) Normal observation errors in the time series.

Having introduced the idea of maximum likelihood, we can do better by making a more realistic assumption about the error distribution. We will then end up with a different likelihood function to maximize, and obtain a different $\hat{\boldsymbol{\theta}}$, but the basic idea is the same.

So what is a better assumption about the observation error distribution, and how can we use the likelihood function to estimate uncertainty in $\hat{\boldsymbol{\theta}}$ and on the fitted trajectory?

Our data are actually non-negative, discrete counts of deaths (or cases in other epidemiological contexts), so a continuous, real-valued Normal distribution is somewhat unrealistic. More importantly, we expect (and can see in the plots of our fitted curves) that the magnitude of error in the observations will vary over the course of the epidemic; the error might be ± 2 at the beginning of the epidemic when mortality is low and ± 50 at the peak.

We could address both of these problems by using a Poisson distribution of observations with mean equal to the fitted model trajectory [Eq. (1c) or Eq. (2)]. This approach handles discrete observations and allows the variance to change as a function of the mean. However, the Poisson distribution assumes ***equidispersion***—the variance is equal to the mean—while typical observation errors are ***overdispersed***, meaning that the variance is greater than the mean. Ignoring overdispersion will underestimate the uncertainty in the parameters and lead to overly narrow confidence intervals on parameters and predictions [Li et al., 2018]. The negative binomial distribution is ~~one~~ the most common way to generalize the Poisson to allow for overdispersion [Lindén and Mäntyniemi, 2011], although other distributions such as the generalized Poisson are occasionally used [Brooks et al., 2019, Kim et al., 2022].

The probability mass function for the ***negative binomial distribution*** (for counts $x = 0, 1, 2, \ldots$) is

$$\texttt{NB}(x;\mu,k) = \frac{\Gamma(k+x)}{\Gamma(k)x!} \left(\frac{k}{k+\mu}\right)^k \left(\frac{\mu}{k+\mu}\right)^x . \tag{17}$$

The predicted variance of a particular observation ~~$x[t_\ell]$~~ $x_{\text{obs}}(t_\ell)$ is given by $\mu_\ell(1+\mu_\ell/k)$, where $\mu_\ell(\boldsymbol{\theta}) = x(t_\ell;\boldsymbol{\theta})$ is the model evaluated at the ℓ^{th} observed data point [*cf.* (7)

and (13)]. The maximum likelihood estimate is, therefore,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{\ell=1}^{n_t} \Big(-\log\Gamma\big(x_{\text{obs}}(t_\ell)+k\big) + \log\Gamma(k) + \log(x_{\text{obs}}(t_\ell)!) \\ - k\log\left(\frac{k}{k+\mu_\ell(\boldsymbol{\theta})}\right) - x_{\text{obs}}(t_\ell)\log\left(\frac{\mu_\ell(\boldsymbol{\theta})}{k+\mu_\ell(\boldsymbol{\theta})}\right)\Big). \tag{18}$$

Here, the overdispersion parameter k also needs to be estimated alongside $\hat{\boldsymbol{\theta}}$ to maximize the likelihood. This is different from the likelihood associated with Normal errors, where σ^2 can be either computed as the variance of the residuals across the full time series or estimated jointly with model parameters.

Regardless of the form of the likelihood function, we can use it to obtain CIs on the MLE $\hat{\boldsymbol{\theta}}$. A relatively simple approach is to use the the curvature of $-\log\mathcal{L}(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$ to infer parameter values of a multivariate Normal distribution for $\boldsymbol{\theta}$. At $\hat{\boldsymbol{\theta}}$, the shape of $-\log\mathcal{L}$ is described by its ***Hessian matrix*** (the matrix of second order partial derivatives of $-\log\mathcal{L}$, also known as the ***Fisher information matrix***), and the inverse of the Hessian is the ***variance-covariance matrix*** $\text{Cov}(\boldsymbol{\theta})$ that specifies the desired multivariate Normal with mean $\hat{\boldsymbol{\theta}}$. This relationship between $\text{Cov}(\boldsymbol{\theta})$ and the Hessian of $-\log\mathcal{L}$ is, admittedly, not obvious! See ~~[Bolker, 2008, §6.5]~~ Bolker [2008, §6.5] for a heuristic explanation or ~~[Wasserman, 2010, §§9.7, 9.10]~~ Wasserman [2010, §§9.7, 9.10] for a rigorous (if terse) explanation.

The diagonal elements of $\text{Cov}(\boldsymbol{\theta})$ are the (estimated) variances of the parameter estimates, so we can take their (positive) square roots to get the standard error (SE) and compute approximate 95% confidence intervals by adding $\pm 1.96\,\text{SE}$ to $\hat{\boldsymbol{\theta}}$ (± 1.96 represents a range containing 95% of the probability of a standard Normal distribution). To obtain CIs on *functions of the fitted parameters* (e.g., ~~$\mathcal{R}_0$~~ $\mathcal{R}_0$ or γ if our model is KM's approximation (2)), we build on the idea that if the error in a parameter a is Δa, then the associated error in a (differentiable) function $g(a)$ is $\Delta g \approx g'(a)\Delta a$. Given a (smooth) nonlinear function $g(\boldsymbol{\theta})$ of the parameters, the ***Delta Method*** ~~Dorfman [1938], Ver Hoef [2012]~~ [Dorfman, 1938, Ver Hoef, 2012] expands $\text{Var}(g(\boldsymbol{\theta}))$ to first order about $\hat{\boldsymbol{\theta}}$, which gives us the variance-covariance matrix of $g(\boldsymbol{\theta})$ ~~[Bolker, 2008, §7.5.2] [Wasserman, 2010, §9.9]~~[Bolker, 2008, §7.5.2] [Wasserman, 2010, §9.9]. In particular, the variance of $g(\boldsymbol{\theta})$ is

$$\begin{aligned}
\text{Var}(g(\boldsymbol{\theta})) &\approx \text{Var}\big[g(\hat{\boldsymbol{\theta}}) + (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})\cdot(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\big] && \text{(19a)}\\
&= \text{Var}\big[(\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})\cdot(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\big] && \text{(19b)}\\
&= \mathbb{E}\big[\big((\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})\cdot(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\big)^2\big] && \text{(19c)}\\
&= \mathbb{E}\big[(\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})^{\mathsf{T}}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathsf{T}}(\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})\big] && \text{(19d)}\\
&= (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})^{\mathsf{T}}\,\mathbb{E}\big[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathsf{T}}\big]\,(\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}}) && \text{(19e)}\\
&= (\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}})^{\mathsf{T}}\,\text{Cov}(\boldsymbol{\theta})\,(\nabla_{\boldsymbol{\theta}} g)(\hat{\boldsymbol{\theta}}) && \text{(19f)}
\end{aligned}$$

We can again get the 95% CIs by taking square roots and computing $g(\hat{\boldsymbol{\theta}}) \pm 1.96\,\text{SE}$.

Given a fit of KM's approximation (2) to the time series data, which yields $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{\omega}, \hat{\phi})$, we can apply the Delta method (19) to the nonlinear relationships (10) to obtain CIs on ~~$g(\hat{\boldsymbol{\theta}}) = (\hat{\mathcal{R}}_e, \hat{\gamma}, \hat{S}_0)$~~$g(\hat{\boldsymbol{\theta}}) = (\widehat{\mathcal{R}_e}, \widehat{\gamma}, \widehat{S(0)})$. This is precisely how we obtained

the CIs on the derived parameters listed in Table 1. Perhaps less obviously, we can also use the Delta method to obtain CIs on the fitted trajectory at each observation time t_ℓ (and hence obtain a confidence band) by considering $g(\boldsymbol{\theta}) = x(t_\ell; \boldsymbol{\theta})$. This is how we obtained the ~~pink confidence band~~ confidence band for the nonlinear least squares fit (light orange) shown in Fig. 1.

Better confidence intervals can be obtained using the ***profile likelihood***, which is calculated by fixing a set of model parameters to specific values and fitting the remaining parameters to maximize the likelihood ~~[Bolker, 2008, §7.5.1]~~[Bolker, 2008, §7.5.1] . By calculating the profile likelihood across a range of parameter values, we obtain the profile likelihood surface, from which confidence intervals can be estimated using the likelihood ratio test ~~[Bolker, 2008, §6.4.1.1]~~[Bolker, 2008, §6.4.1.1]. While profile likelihoods generally give more accurate estimates of confidence intervals, calculating the profile likelihood can be challenging, if not practically impossible, for derived parameters or epidemic trajectories ~~[Bolker, 2008, §7.5.1.2]~~[Bolker, 2008, §7.5.1.2]. Consequently, we rely on the Delta Method here.

5. Fitting the ODE

Until now, we have focused on fitting KM's approximation (2) rather than actual solutions of the SIR model (1). If we had an exact analytical solution of the SIR ODE (1) then we could proceed as above, replacing the approximate analytical expression (2) with the exact formula. Since we do not have an exact solution, we ~~must~~ instead rely on numerical solutions of the ODE. Fitting numerical solutions of ODEs to data introduces significant coding/computational challenges, but conceptually the problem is the same as if we did have an analytical formula. We can still use the Delta method (19) to estimate uncertainty, but calculating the gradient $(\nabla_{\boldsymbol{\theta}} g)(\boldsymbol{\theta})$ is not straightforward if g is a numerical solution of an ODE; we must simultaneously solve a set of ***sensitivity equations*** ~~[Raue et al., 2013, Eq. (6)]~~ [Raue et al., 2013, Eq. (6)] alongside the main differential equations. Sensitivity equations define the time derivatives of the gradients of trajectories with respect to the parameters. They can easily be derived using the chain rule; if we write a generic, autonomous ODE for $\boldsymbol{x}(t; \boldsymbol{\theta})$ as

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \,, \qquad \boldsymbol{x}(0, \boldsymbol{\theta}) = \boldsymbol{x}_0(\boldsymbol{\theta}), \tag{20}$$

then the sensitivity equations are

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\nabla_{\boldsymbol{\theta}} \boldsymbol{x}(t; \boldsymbol{\theta})\Big) = \nabla_{\boldsymbol{\theta}}\Big(\frac{\mathrm{d}\boldsymbol{x}(t; \boldsymbol{\theta})}{\mathrm{d}t}\Big) = \nabla_{\boldsymbol{\theta}}\Big(\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})\Big) \tag{21a}$$

$$= \nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})\, \nabla_{\boldsymbol{\theta}} \boldsymbol{x}(t; \boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \,. \tag{21b}$$

If $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are $n_{\bar{x}}$ and $n_{\bar{\theta}}$dimensional, respectively, then the $n_x n_\theta$ ***sensitivities*** $\mathsf{S}_{ij}(t)$ are given by the $n_x \times n_\theta$ ***sensitivity matrix***,

$$\mathbf{S}(t) = \nabla_{\boldsymbol{\theta}} \boldsymbol{x}(t; \boldsymbol{\theta}) \,. \tag{22}$$

Eq. (21) defines a set of $n_x n_\theta$ differential equations for S_{ij},

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \left[\nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})\right] \mathbf{S} + \left[\nabla_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})\right], \tag{23a}$$

which can be solved jointly with the original ODEs (20) for the state variables ($\boldsymbol{x}$) by specifying initial conditions

$$\mathbf{S}(0) = \nabla_{\boldsymbol{\theta}} \big(\boldsymbol{x}_0(\boldsymbol{\theta})\big). \tag{23b}$$

We can then use a further chain-rule step to compute the (total) derivative of the log-likelihood of the observations with respect to the parameters. To get this right, it helps to make explicit the dependence on the trajectory ($\boldsymbol{x}$) versus dependence on the parameters ($\boldsymbol{\theta}$, by which we will now mean all parameters, including parameters of the trajectory model and of the observation process model). For a general function $\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{\theta})$, the total derivative with respect to $\boldsymbol{\theta}$ is

$$\frac{\mathrm{d}\boldsymbol{\Phi}}{\mathrm{d}\boldsymbol{\theta}} = \nabla_{\boldsymbol{x}} \boldsymbol{\Phi}\ \nabla_{\boldsymbol{\theta}} \boldsymbol{x} + \nabla_{\boldsymbol{\theta}} \boldsymbol{\Phi}. \tag{24}$$

To apply this to the log-likelihood, it is helpful to make dependence on the trajectory $\boldsymbol{x}$ explicit. Consistent with our notation above [e.g., Eq. (7)], we write ~~$\boldsymbol{x}[t_\ell]$~~ $\boldsymbol{x}_{\mathrm{obs}}(t_\ell)$ for the observations at times $t_\ell \in \{t_1, t_2, \ldots, t_{n_t}\}$, making it easier to distinguish them from the fitted model trajectory evaluated at these times, $\boldsymbol{x}(t_\ell; \boldsymbol{\theta})$. Then

$$\frac{\mathrm{d} \log \mathcal{L}(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \Big(\log \mathbb{P}_{\mathbb{p}}\big(\{\boldsymbol{x}_{\mathrm{obs}}(t_\ell) : \ell = 1, \ldots, n_t\} \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) \Big) \tag{25a}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \Big(\log \prod_{\ell=1}^{n_t} \mathbb{P}_{\mathbb{p}}\big(\boldsymbol{x}_{\mathrm{obs}}(t_\ell) \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) \Big) \tag{25b}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \sum_{\ell=1}^{n_t} \Big(\log \mathbb{P}_{\mathbb{p}}\big(\boldsymbol{x}_{\mathrm{obs}}(t_\ell) \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) \Big) \tag{25c}$$

$$= \sum_{\ell=1}^{n_t} \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \Big(\log \mathbb{P}_\ell \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \Big) \qquad [\text{abbreviating } \mathbb{P}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \mathbb{P}\big(\boldsymbol{x}[t_\ell] \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) \text{abbreviating } \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \mathbb{p}\big(\boldsymbol{x}_{\mathrm{obs}}(t_\ell) \mid \tag{25d}$$

$$= \sum_{\ell=1}^{n_t} \frac{1}{\mathbb{P}_\ell(\boldsymbol{x}, \boldsymbol{\theta})} \frac{1}{\mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})} \Big(\nabla_{\boldsymbol{x}} \mathbb{P}_\ell \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})\, \nabla_{\boldsymbol{\theta}} \boldsymbol{x} + \nabla_{\boldsymbol{\theta}} \mathbb{P}_\ell \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \Big) \Big|_{\boldsymbol{x} = \boldsymbol{x}(t_\ell; \boldsymbol{\theta})} \tag{25e}$$

$$= \sum_{\ell=1}^{n_t} \frac{1}{\mathbb{P}_\ell(\boldsymbol{x}, \boldsymbol{\theta})} \frac{1}{\mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})} \Big(\nabla_{\boldsymbol{x}} \mathbb{P}_\ell \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta})\, \mathbf{S}(t_\ell) + \nabla_{\boldsymbol{\theta}} \mathbb{P}_\ell \mathbb{p}_\ell(\boldsymbol{x}, \boldsymbol{\theta}) \Big) \Big|_{\boldsymbol{x} = \boldsymbol{x}(t_\ell; \boldsymbol{\theta})}, \tag{25f}$$

where we typically assume the probability distribution

$$\mathbb{P}_{\mathbb{p}}\big(\boldsymbol{x}_{\mathrm{obs}}(t_\ell) \mid \boldsymbol{x}(t_\ell; \boldsymbol{\theta}),\, \boldsymbol{\theta}\big) = \prod_{i=1}^{n_x} \mathtt{NB}(x_i[t_\ell] x_{\mathrm{obs},i}(t_\ell);\, x_i(t_\ell, \boldsymbol{\theta}),\, \boldsymbol{\theta}). \tag{26}$$

We have slightly abused notation here, compared with Eq. (17); we have written $\boldsymbol{\theta}$ rather than k as the final argument of the negative binomial distribution, since there might be a different k for each observed variable x_i, and we collect all parameters into the single vector $\boldsymbol{\theta}$. (The examples we discuss in this paper involve only a single observed time series, so $n_x = 1$.)

Integrating the sensitivity equations (23) in parallel with the ODEs (20) is a computationally efficient and numerically stable way to calculate the overall gradients of the log-likelihood with respect to the parameters, which makes nonlinear estimation more robust and efficient. We can also use these gradients to calculate CIs using the Delta method. ~~Raue *et al.*~~ ~~Raue et al. [2013]~~ Raue et al. [2013] give a detailed comparison between using the sensitivity equations and computing gradients by finite-difference approximations. (~~Bjørnstad [Bjørnstad, 2018, Chapter 9]~~ Bjørnstad [2018, Chapter 9] also gives an introduction to trajectory matching.)

The `fitode` package[9] does all of this computational work under the hood, and makes it as easy for a user to fit an ODE to data as it was for us to use `nls` above to fit a curve based on an analytical formula. We begin illustrating the use of the package by fitting the SIR model (1) to the Bombay plague epidemic.

We first load the package

```
library(fitode)
```

and define a model object:

```
SIR_model <- odemodel(
    name="SIR model",
    model=list(
        S ~ - beta * S * I,
        I ~ beta * S * I - gamma * I,
        R ~ gamma * I
    ),
    observation = list(
        mort ~ dnbinom(mu = R, size = k)
    ),
    diffnames="R",
    initial=list(
        S ~ S0,
        I ~ I0,
        R ~ 0
    ),
    par=c("beta", "gamma", "S0", "I0", "k")
)
```

In the model definition above:

`model` specifies the vector field given by the ODE (1).

`observation` specifies ~~that the~~ the observation model: the observed data (`mort`) are assumed to arise from sampling from the negative binomial distribution [`dnbinom`, Eq. (17)] with overdispersion parameter k. Ordinary least squares (with normally distributed observation errors) can be implemented by changing the `observation` argument to `mort ~ ols(mean = R)`. The mean of the distribu-

[9] `fitode` is available on CRAN, and can be installed via `install.packages("fitode")`.

tion is given by the incidence derived from the fitted model trajectory [Eq. (13a)],

$$\mu(t_\ell) = \int_{t_{\ell-1}}^{t_\ell} \frac{\mathrm{d}R}{\mathrm{d}t}\,\mathrm{d}t = R(t_\ell) - R(t_{\ell-1})\,, \tag{27}$$

Fitting to such differences, useful whenever the observations represent accumulated values of processes (such as infections, recoveries, or deaths) between observation times, is implemented by using the `diffnames` argument to specify the state variable for which consecutive differences are to be used (so, if the focal variable is R then `fitode` fits to $R(t_\ell) - R(t_{\ell-1})$ rather than $R(t_\ell)$).

`initial` conditions are expressed as numbers of individuals.

`par` refers to the parameters to be fitted: β, γ, initial conditions $S(0)$ and $I(0)$, and the overdispersion parameter k.

Since we are taking the difference $\mu(t_\ell) = R(t_\ell) - R(t_{\ell-1})$ to calculate the mortality trajectory,[10] we have to add an extra row representing t_0 to the data set in order to compute $\mu(t_1) = R(t_1) - R(t_0)$:

```
bombay2 <- rbind(
    c(times=bombay$week[1] -
          diff(bombay$week)[1], mort=NA),
    bombay
)
```

Taking our previous parameter estimates from `nls` as starting values (and choosing a starting value for k), we can fit the model by calling the `fitode` function:

```
SIR_start <- c(beta=beta.nls, gamma=gamma.nls,
               I0=I0.KM, S0=S0.nls, k=50)
SIR_fit <- fitode( model = SIR_model, data = bombay2,
                   fixed = list(gamma=gamma.nls),
                   start = SIR_start, tcol = "week" )
```

In the fitting function above:

`model` specifies the ODE model to be fitted.

`data` specifies the data.

`fixed` specifies parameter values to be fixed (and therefore not estimated); above, we chose to assume that the recovery rate γ is known (due to parameter unidentifiability[11]).

`start` specifies the starting parameter set for the optimization[12].

[10] Modelers often fit trajectories to cumulative curves. However, doing so is ill-advised because points in a cumulative time series are not independent, making it difficult to define CIs ~~King et al. [2015]~~[King et al., 2015].

[11] In short, unidentifiability of γ means that we can obtain nearly identical fits across a wide range of γ. While it is possible to fit the model without fixing γ, the resulting estimates are sensitive to starting conditions and numerically unstable, preventing a reliable calculation of the Hessian matrix and therefore precluding estimation of confidence intervals. These issues could be addressed alternatively by fixing a different parameter instead and estimating γ. We typically choose to fix γ because the mean duration of infection ($1/\gamma$) can often be estimated from independent data sources; here, to make comparisons of fits easier to interpret, we have fixed γ to the value we estimated via `nls` fits of the KM approximation (2).

[12] In general, worse models (providing a poorer or less identifiable fit to the data) and worse data (fewer data points and more noise) will increase the sensitivity of fits to the starting values.

`tcol` specifies the name of the time column of the data frame.

The resulting fits are plotted in Fig. 1 and summarized in Table 2. The estimated parameter values (the ***coefficients*** of the model) can be obtained via `coef(SIR_fit)`. The coefficients together with associated confidence intervals are obtained via `confint(SIR_fit)`, which can also provide confidence intervals for derived parameters using the Delta method. Note that `fitode` gives discrete predictions (rather than smooth curves) because we are calculating mortality at discrete (weekly) time intervals using Eq. (27).

6. Cautionary remarks concerning fits to Bombay plague

We have highlighted the Bombay plague data because of their prominent role in KM's paper ~~Kermack and McKendrick [1927]~~ [Kermack and McKendrick, 1927] and, consequently, for the history of mathematical epidemiology. However, while they provide an interesting example with which to illustrate the process of fitting an epidemiological model to data, modelling plague dynamics with the simple SIR model is, at best, difficult to justify: ~~Bacaër Bacaër [2012]~~ Bacaër [2012] argues that the trajectory of the Bombay plague epidemic was primarily governed by seasonality rather than SIR dynamics. Indeed, KM ~~-~~themselves recognized that their model involves a sequence of optimistic assumptions, which they admitted were not "strictly" satisfied:

> "We are, in fact, assuming that plague in [humans] is a reflection of plague in rats, and that with respect to the rat (1) the uninfected population was uniformly susceptible; (2) that all susceptible rats in the island had an equal chance of being infected; (3) that the infectivity, recovery, and death rates were of constant value throughout the course of sickness of each rat; (4) that all cases ended fatally or became immune; and (5) that the flea population was so large that the condition approximated to one of contact infection. None of these assumptions are strictly fulfilled and consequently the numerical equation can only be a very rough approximation. A close fit is not to be expected, and deductions as to the actual values of the various constants should not be drawn."
>
> — KM [p. 715]

Given the mental gymnastics required to motivate applying the SIR model to plague transmission, it is surprising that KM did not choose to examine a more obviously suitable disease. The surprise is especially extreme given that the most salient infectious disease epidemic in the 1920s would have been the 1918 influenza pandemic, which did involve direct human-to-human transmission, and for which much more detailed data were available at the time ~~Rogers [1920], Frost [1920], Eichel [1923]~~ [Rogers, 1920, Frost, 1920, Eichel, 1923].

7. Influenza in Philadelphia, October 1918

Deaths caused ultimately by influenza are often attributed to pneumonia ~~Earn et al. [2002]~~[Earn et al., 2002], so influenza mortality studies typically combine pneumonia and influenza (P&I). Among published tables summarizing P&I mortality during the 1918 pandemic, a particularly valuable example concerns the main wave in the city of Philadelphia ~~Rogers [1920]~~[Rogers, 1920]. These data are exceptional because they are restricted to a single, large city, and because they provide *daily* counts that capture the detailed temporal pattern (large dots in Fig. 2).

As for Bombay plague, we can fit KM's approximation (2) to the Philadelphia influenza epidemic using nonlinear least squares, which yields the ~~red~~ orange curve in Fig. 2. While this `nls` fit does not look unreasonable at a glance, the fitted parameter values (Table 3) are absurd, including a basic reproduction number ~~$\mathcal{R}_0 \approx 2500$~~ $\mathcal{R}_0 \approx 2500$ and a mean generation interval $T_g \approx 1.5$ years.

Matching numerically computed trajectories of the exact SIR model using `fitode` gives a fit—the solid ~~yellow~~ gold curve in Fig. 2—that is visually similar to the (~~red~~orange) fit of KM's approximation, but provides much more realistic parameter estimates (Table 4); in particular, ~~$\mathcal{R}_0 \approx 6.4$~~ $\mathcal{R}_0 \approx 6.4$ and $T_g \approx 4.3$ days.

If we convert the `fitode` estimates of the SIR parameters to the parameters of KM's approximation, we obtain the dotted ~~yellow~~ gold curve in Fig. 2, which grossly underestimates the magnitude of the epidemic (the epidemic peak occurs much too soon). The KM approximation (2) is good initially, but becomes poorer and poorer over time as the underlying assumption on which it is based (4) becomes less and less valid.

8. Fitting the deterministic SIR model to stochastic simulations

The most compelling tests of estimation methods involve fitting models to data that have been generated from a known model, so we know the true underlying values of the parameters we are trying to estimate.

The most basic test is essentially a consistency check: in the context of the SIR model, we choose initial conditions ~~(S_0, I_0)~~ $(S(0), I(0))$ and parameter values (~~$\mathcal{R}_0, T_g$~~$\mathcal{R}_0, T_g$), compute the associated trajectory by solving Eq. (1) numerically, and then use `fitode` to estimate the parameters. At least if we choose starting values reasonably close to the correct underlying values, `fitode` should converge to those values.

The next level of testing is to take our numerically computed solution and artificially "observe" it with error, i.e., using a noise distribution that we specify. For example, we could take observation errors to be negative binomially distributed with overdispersion parameter k, and then use `fitode` to estimate k together with the other parameters (~~$S_0, I_0, \mathcal{R}_0, T_g$~~$S(0), I(0), \mathcal{R}_0, T_g$).

A still more stringent test is to simulate data from a model that is more complex and realistic than the idealized model that we want to fit, and then see if we can nevertheless recover parameters that correspond to those of our idealized model (e.g., ~~$\mathcal{R}_0$~~ $\mathcal{R}_0$ and T_g for the SIR model). We will take a step in this direction in this section by fitting the deterministic SIR model (1) to data generated by a fully stochastic version of the model.

The standard stochastic SIR model ~~Andersson and Britton [2000]~~ [Andersson and Britton, 2000] can be defined by interpreting the individual terms in Eq. (1) as event rates for stochastic processes in a population of N individuals (in the limit $N \to \infty$ the stochastic model approaches the ODEs (1); see ~~Ethier and Kurtz [1986]~~Ethier and Kurtz [1986]). Realizations of this discrete-state model can be generated exactly using the Gillespie algorithm ~~Gillespie [1976]~~ [Gillespie, 1976], or approximately (as we do here) using the "τ-leaping" approach ~~Gillespie [2001]~~[Gillespie, 2001], which is implemented in the `adaptivetau` R package ~~Johnson [2023]~~[Johnson, 2023]. The demographic stochasticity that these algorithms simulate is essential to capture real effects that occur when the number of infected individuals is small (especially the possibility that an epidemic can burn out

Parsons et al. [2024][Parsons et al., 2024]).

In Fig. 3, the dots simulated data points show a single realization of the stochastic SIR model with initial state $(S_0, I_0, R_0) = (1998, 2, 0)$$(S(0), I(0), R(0)) = (1998, 2, 0)$, basic reproduction number $\mathcal{R}_0 = 5$$\mathcal{R}_0 = 5$, and mean generation interval $T_g = \gamma^{-1} =$ 1 week. In the top panel, dR/dt [Eq. (1c)] with the correct initial conditions and parameter values is shown with solid greengreen , and the KM approximation (2) based on those parameter values is shown with dotted greengreen . The `fitode` fit [based on $\int (dR/dt) dt$$\int (dR/dt)\, dt$] and confidence band are shown in yellowgold . The time shift between the deterministic solution and the stochastic realization arises because the stochastic model captures the demographic noise (which causes a randomly distributed delay until the tipping point is reached, i.e., until the epidemic takes off in a roughly deterministic fashion).

As expected, with the correct parameter values, KM's approximation (2) fails once the requirement (4) that $R(t)/N \ll 1/\mathcal{R}_0$$R(t)/N \ll 1/\mathcal{R}_0$ is violated. We can, of course, find values of the parameters (a, ω, ϕ) such that the function $a\,\mathrm{sech}^2(\omega\, t - \phi)$ [Eq. (2)] more closely matches the shape of the full simulated epidemic. Using nonlinear least squares (`nls`) as in previous sections, we obtain visually reasonable agreement (Fig. 3, bottom panel, red orange curve; Table 5). This `nls` fit cannot be improved further because the function we are fitting (2) is symmetric about its peak, whereas the rise is steeper than the fall in the simulated epidemic. It is also worth emphasizing that the parameter values that yield the red orange curve in Fig. 3 are far from the true parameters that were used in the simulation (Table 5).

The excellent fit of the deterministic trajectory that `fitode` finds (yellowgoldin Fig. 3) is achieved by estimating an initial prevalence that is only a third of the true initial prevalence, thereby mimicking the stochastic delay with the deterministic model; all other parameter estimates are nearly identical to the true parameter values used to generate the stochastic trajectory (Table 6).

9. Discussion

We have presented a basic theoretical and practical introduction to standard methods for fitting dynamical models to time series, in the context of infectious disease epidemiology. We explained how to use nonlinear least squares (`nls`) to fit a given function to a time series, and illustrated the approach using the Kermack and McKendrick (KM , Kermack and McKendrick [1927]) analytical approximation (2) to the solution of the standard SIR model (1). We also explained how to fit solutions of ordinary differential equations (ODEs) to a time series—using our R package `fitode`—and obtain parameter estimates and confidence intervals, regardless of whether analytical solutions of the ODEs are available.

`fitode` is flexible enough to handle most compartmental epidemiological and ecological models Brauer and Castillo-Chavez [2001], Brauer and Kribs [2016], Brauer et al. [2019] [Brauer and Castillo-Chavez, 2001, Brauer and Kribs, 2016, Brauer et al., 2019] , including non-autonomous models, such as seasonally forced epidemic models London and Yorke [1973], Earn et al. [2000], He and Earn [2007, 2016], Papst and Earn [2019] [London and Yorke, 1973, Earn et al., 2000, He and Earn, 2007, 2016, Papst and Earn, 2019] . We hope the package will be useful for many readers, not only as a pedagogical tool but also to fit models to novel data. Potential applications abound (we have ourselves used `fitode`'s predecessor, `fitsir`, to study music popularity Rosati et al. [2021]

[Rosati et al., 2021]).

We focused here on three illustrative examples of epidemic time series. The first was the reported weekly mortality from plague in Bombay in 1906 (Fig. 1), which was examined by KM in their original paper ~~Kermack and McKendrick [1927]~~ [Kermack and McKendrick, 1927]. Although historically important, it is certainly debatable whether we can trust any inferences we might draw from fitting the simple SIR model (1) to these plague data~~can be trusted~~. As we quoted at the end of Sect. 5, to justify the application of their SIR model to these data, KM highlighted five implicit assumptions, any or all of which might be violated. ~~In addition~~Furthermore, Bacaër ~~investigated~~ found that over the longer term ~~pattern of plague mortality in Bombay and found that there were seasonal epidemics~~ seasonal epidemics of plague occurred in Bombay every year from 1897 to 1911 ~~[Bacaër, 2012, Fig. 2]~~[Bacaër, 2012, Fig. 2], suggesting that the 1906 epidemic was just one in a long sequence of epidemics that were "driven by seasonality" ~~[Bacaër, 2012, p. 403]~~[Bacaër, 2012, p. 403]. Of course, other mechanisms (e.g., heterogeneity in contact patterns) might play a role as well.

To obtain a deeper understanding of the Bombay plague epidemic, we could formulate a variety of models, fit them to the data using `fitode` or other software, and use a statistical framework for model selection ~~Burnham and Anderson [2002]~~ [Burnham and Anderson, 2002] to rank the relative importance of the various mechanisms included in the sequence of models (see, e.g., ~~He *et al.* He et al. [2013]~~ He et al. [2013] for an example of using this approach to understand the occurrence of three distinct waves in the 1918 influenza pandemic~~. ()~~. Alternatively, we could formulate one model that included *all* of the processes and attempt to measure their relative importance by comparing the magnitudes of parameters ~~Bolker [2023].)~~ [Bolker, 2023] . We have not attempted such a study here, since our goal was simply to explain and illustrate the fitting methodology. However, it is worth highlighting that our analysis using the SIR model did reveal a computational challenge that—in the absence of additional information about the Bombay plague outbreak—would likely limit how much can be learned from a model selection exercise: the mean generation interval (T_g) appears to be ***unidentifiable***, i.e., ~~impossible~~ difficult to estimate reliably from the reported weekly plague deaths alone (see Fig. 4).

Our second example was the main wave of the 1918 influenza pandemic in the city of Philadelphia, for which daily mortality from pneumonia and influenza (P&I) was reported (Fig. 2). Again we fitted ~~the exact~~ a numerical solution of the SIR model (1) using `fitode`, and KM's analytical approximation (2), but found—unlike the situation for Bombay plague—that only the `fitode` fit yielded plausible parameter estimates (see Tables 3 and 4).

Finally, we conducted a kind of test that truly makes most sense to perform *before* fitting to a real, empirically observed time series: we fit models to a simulation that we ran, so we knew the parameter values used to generate the simulated "observations". The simulation was a realization of the stochastic SIR model, to which, again, we fit both the deterministic SIR model (1) using `fitode` and KM's analytical approximation (2) using `nls`. ~~Both provide visually~~ At a glance, both provide visually roughly reasonable fits (Fig. 3, bottom panel) but KM's approximation cannot represent the asymmetry about the peak in the epidemic curve and yields absurd parameter values, whereas `fitode` estimates ~~the correct~~ an epidemic curve with the correct shape and the correct values of the underlying disease-related parameters (Tables 5 and 6). (We did find a discrepancy in the estimates of initial conditions; this was driven by the failure of the stochastic outbreak simulation to take off immediately. A lower initial prevalence is the only mechanism by which the deterministic model can capture

the delayed onset of the epidemic. In practice, modelers fitting to epidemic time series by trajectory matching usually pick an "epidemic window" that corresponds to the part of the epidemic that can be reasonably captured by a deterministic model ~~Earn et al. [2020]~~[Earn et al., 2020].)

KM's approximation (2) estimates the simulation parameters badly because the assumption on which it is based (4) is strongly violated in the simulation (Fig. 3). Consequently, the parameters of the KM approximation cannot be interpreted biologically or mechanistically. More generally, a purely phenomenological model with the same number of parameters can sometimes fit a stochastic simulation just as closely or even closer than the deterministic limit of the model that generated the data ~~Rosati et al. [2021]~~[Rosati et al., 2021]; a good fit is not, on its own, sufficient to conclude that a model matches the underlying processes of a dynamical system.

While `fitode` provides a relatively easy way to specify ODEs and estimate their parameters from data, any programming language will work to implement the steps we have outlined above, including both free general-purpose languages such as Python [Batista and da Silva, 2022, Gupta, 2023] or commercial, domain-specific tools such as MATLAB [Chowell, 2017] or Berkeley Madonna [Zha et al., 2020]. As long as a language provides tools for integrating arbitrary sets of ODEs (e.g., MATLAB's `ode45`) and optimizing nonlinear functions (e.g., MATLAB's `fminunc` or `lsqnonlin`), it can be used to estimate parameters of ODEs. However, `fitode`'s simple interface, automatic derivation of sensitivity equations, flexible specification of observation models, and provision of confidence intervals make it both convenient and powerful.

Beyond the basics that we have discussed here, `fitode` contains a number of useful advanced features. In particular, `fitode` can

fit to multiple data streams: `fitode` is not limited to fitting a trajectory to a single state variable, such as incidence or prevalence of infected individuals. For example, during the later stages of the COVID-19 pandemic modelers often had access to time series of case reports, hospitalization reports, and wastewater sampling for the same geographic region. If we build a model that includes state variables for hospitalized individuals and for virus concentrations in wastewater, `fitode` can fit the model's parameters using all of the available data (as in ~~Nourbakhsh et al. [2022]).~~ Nourbakhsh et al. [2022]).

compute confidence intervals via importance sampling: While the Delta method can compute confidence intervals for derived quantities such as predicted trajectories, it rests on strong and sometimes unreliable assumptions. A more accurate but computationally expensive approach starts by sampling parameter sets randomly from a multivariate normal distribution with a mean and covariance matrix drawn from the maximum likelihood fit. For each set of parameters in the ensemble, `fitode` computes the likelihood and a predicted trajectory (or some quantity such as the total size of the epidemic); an average value and confidence intervals are derived from weighted moments (means) or quantiles (medians or extremes such as 10th and 90th percentiles).

specify priors and apply Bayesian inference: Unlike maximum likelihood approaches, which seek to estimate the best-fitting parameter set, Bayesian methods aim to estimate a distribution of parameters (also known as the posterior distribution) that are consistent with our previous knowledge about the system (encapsulated in ~~*prior distributions*~~***prior distributions***) as well as the observed data. These approaches are generally better at handling parameter uncertainties [Elderd et al., 2006] but are usually much more computationally expensive.

`fitode` allows the user to specify prior distributions on parameters; these priors can either reflect previous knowledge of a disease system, or can be used to ~~*regularize*~~ ***regularize*** a fitting procedure by downweighting extreme values of parameters ~~Lemoine [2019]~~[Lemoine, 2019], which can help mitigate problems with identifiability (see below).

Bayesian modelers typically use ~~*Markov chain Monte Carlo*~~ ***Markov chain Monte Carlo*** algorithms to explore the parameter space and approximate the target distribution. `fitode` implements a simple ~~*Metropolis-Hasting*~~ ***Metropolis-Hasting*** sampler [Bolker, 2008, §7.3.1]. (The Stan platform provides a much more powerful Bayesian sampling algorithm using sensitivity equations, built on top of a fully general system for specifying ODEs; however, this tool requires significantly more computational and statistical background to use effectively ~~Grinsztajn et al. [2021]~~[Grinsztajn et al., 2021].)

Even with these extensions, modelers ~~will~~ may face many challenges when fitting ODEs to data with the `fitode` package, as with fitting any nonlinear model to data. For example, it is often difficult to ensure that the model has converged properly or reached its true maximum. ~~Performing~~ More generally, when they first start attempting to fit models to data, naïve and optimistic epidemic modelers often run into problems of ***structural identifiability*** (the impossibility of estimating particular sets of parameters from data, regardless of how much data is available [Tuncer and Le, 2018, Chowell et al., 2023]) and ***practical identifiability*** (the impossibility of reliably estimating parameters from a particular small, noisy data set [Gallo et al., 2022, Chowell et al., 2023]). In addition to the rigorous methods described by Chowell et al. [2023], using a ***multistart method*** (performing optimization from multiple starting conditions ~~and drawing likelihood surfaces~~[Raue et al., 2013]), or plotting likelihood surfaces, can help diagnose these problems. Using different optimization methods or reparameterizing the model can also help ~~Raue et al. [2013], Bolker et al. [2013]. We also~~ [Raue et al., 2013, Bolker et al., 2013]. We encourage users of `fitode` who encounter these or other fitting challenges to open issues via the `fitode` GitHub repository (`https://github.com/parksw3/fitode`).

As its name suggests, `fitode` is limited to fitting ODEs to time series. Consequently, by design, `fitode` ignores ***process error***, i.e., random variability that affects both current and future steps of the trajectory—as opposed to ***observation error***, which arises from imperfect measurements or reporting and is usually assumed to be independent of the trajectory itself. A key component of process error is the demographic stochasticity that is inherent to the discrete-state stochastic SIR model discussed above (and to any real host-pathogen system). Parameters of models can also be subject to process error; for example, the transmission rate might depend on random fluctuations in weather. Properly accounting for process error can be critical for ~~estimating~~ accurately quantifying uncertainties in parameter estimates and ~~confidence bands on the projected dynamics of a system [King et al., 2015, Taylor et al., 2016, Li et al., 2018]~~ model forecasts [King et al., 2015, Taylor et al., 2016, Li et al., 2018]; however, the required statistical and computational procedures are significantly more challenging than the approaches considered here. Popular `R` packages that can fit models with process error include `pomp` [King et al., 2016] and `mcstate` [FitzJohn et al., 2024].

10. Closing remarks: from Fred Brauer to `fitode`

The idea of digging into ~~to~~ data seemed like punishment to Fred Brauer, but while he never—to our knowledge—did any data analysis himself, he did develop a sincere appreciation for the value of data in epidemiological research. Fred's curiosity—about how dynamical models can be fit to data, and why it is hard—convinced us that it would be worth writing a paper (and accompanying software) that could draw more dynamicists working on epidemic models into the world of data.

We have provided two answers to Fred's question of "how" to fit models to data (via `nls` or `fitode`), and through examples we have hinted at some of the reasons "why" such fitting can be very difficult. A true understanding of "why it is hard" is something that builds over time with experience, but the key points are that finding optima of a complex multi-dimensional function is hard enough on its own [Raue et al., 2013], and estimating statistically meaningful uncertainty in those optima is extremely challenging [Elderd et al., 2006, Li et al., 2018].

Fred would never have used `fitode`, but would have delighted in seeing it demonstrated and in discussing the theoretical background on model fitting that we have presented in this paper. We hope that others like him, as well as students and researchers who actually do want to dig into data, will benefit from this exposition.

References

R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1991.

H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis*, volume 151 of *Lecture notes in statistics*. Springer-Verlag, New York, 2000.

N. Bacaër. The model of Kermack and McKendrick for the plague epidemic in Bombay and the type reproduction number with seasonality. *Journal of Mathematical Biology*, 64(3):403–422, Feb. 2012. ISSN 0303-6812, 1432-1416. . http://link.springer.com/article/10.1007/s00285-011-0417-5.

N. T. J. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Hafner Press, New York, second edition, 1975.

M. S. Bartlett. *Stochastic population models in ecology and epidemiology*, volume 4 of *Methuen's Monographs on Applied Probability and Statistics*. Spottiswoode, Ballantyne & Co. Ltd., London, 1960.

A. A. Batista and S. H. da Silva. An epidemiological compartmental model with automated parameter estimation and forecasting of the spread of COVID-19 with analysis of data from Germany and Brazil. *Frontiers in Applied Mathematics and Statistics*, 8, Apr. 2022. ISSN 2297-4687. . URL `https://www.frontiersin.org/articles/10.3389/fams.2022.645614`. Publisher: Frontiers.

O. N. Bjørnstad. *Epidemics: Models and Data Using R*. Springer, New York, NY, 1st ed. 2018 edition edition, Nov. 2018. ISBN 978-3-319-97486-6.

B. Bolker. Multimodel approaches are not the best way to understand multifactorial systems. July 2023. URL `https://ecoevorxiv.org/repository/view/5722/`. Publisher: EcoEvoRxiv.

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

B. M. Bolker, B. Gardner, M. Maunder, C. W. Berg, M. Brooks, L. Comita, E. Crone, S. Cubaynes, T. Davies, P. de Valpine, J. Ford, O. Gimenez, M. Kéry, E. J. Kim, C. Lennert-Cody, A. Magnusson, S. Martell, J. Nash, A. Nielsen, J. Regetz, H. Skaug, and E. Zipkin. Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods in Ecology and Evolution*, 4(6):501–512, June 2013. ISSN 2041210X. . URL

http://doi.wiley.com/10.1111/2041-210X.12044.
F. Brauer and C. Castillo-Chavez. *Mathematical models in population biology and epidemiology*, volume 40 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2001.
F. Brauer and C. Kribs. *Dynamical systems for biological modeling: An introduction*. CRC press, 2016.
F. Brauer, C. Castillo-Chavez, and Z. Feng. *Mathematical models in epidemiology*, volume 32. Springer, 2019.
M. E. Brooks, K. Kristensen, M. R. Darrigo, P. Rubim, M. Uriarte, E. Bruna, and B. M. Bolker. Statistical modeling of patterns in annual reproductive rates. *Ecology*, 100(7):e02706, 2019. ISSN 1939-9170. .
E. Brooks-Pollock, L. Danon, T. Jombart, and L. Pellis. Modelling that shaped the early COVID-19 pandemic response in the UK. *Philosophical Transactions of the Royal Society B*, 376(1829):20210001, 2021.
K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: A practical information-theoretic approach*. Springer, New York, 2nd edition, 2002.
M. Campbell-Kelly. Origin of computing. *Scientific American*, 301(3):62–69, 2009.
D. Champredon and J. Dushoff. Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821):20152026, 2015.
D. Champredon, J. Dushoff, and D. J. D. Earn. Equivalence of the Erlang SEIR epidemic model and the renewal equation. *SIAM Journal on Applied Mathematics*, 78(6):3258–3278, 2018. . URL https://epubs.siam.org/doi/10.1137/18M1186411.
G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, Aug. 2017. ISSN 2468-0427. .
G. Chowell, S. Dahal, Y. R. Liyanage, A. Tariq, and N. Tuncer. Structural identifiability analysis of epidemic models based on differential equations: A tutorial-based primer. *Journal of Mathematical Biology*, 87(6):79, Nov. 2023. ISSN 1432-1416. .
O. Diekmann and J. A. P. Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Wiley Series in Mathematical and Computational Biology. John Wiley & Sons, LTD, New York, 2000.
R. A. Dorfman. A note on the δ-method for finding variance formulae. *The Biometric Bulletin*, 1:129–137, 1938.
D. J. D. Earn. A Light Introduction to Modelling Recurrent Epidemics. In F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, volume 1945 of *Lecture Notes in Mathematics*, pages 3–17. Springer, 2008. . URL https://link.springer.com/chapter/10.1007%2F978-3-540-78911-6_1.
D. J. D. Earn. Mathematical epidemiology of infectious diseases. In M. A. Lewis, M. A. J. Chaplain, J. P. Keener, and P. K. Maini, editors, *Mathematical Biology*, volume 14 of *IAS/-Park City Mathematics Series*, pages 151–186. American Mathematical Society, 2009. . URL http://www.ams.org/books/pcms/014/.
D. J. D. Earn, P. Rohani, B. M. Bolker, and B. T. Grenfell. A simple model for complex dynamical transitions in epidemics. *Science*, 287(5453):667–670, 2000. . URL http://science.sciencemag.org/content/287/5453/667.
D. J. D. Earn, J. Dushoff, and S. A. Levin. Ecology and evolution of the flu. *Trends in Ecology and Evolution*, 17(7):334–340, 2002. .
D. J. D. Earn, J. Ma, H. Poinar, J. Dushoff, and B. M. Bolker. Acceleration of plague outbreaks in the second pandemic. *PNAS – Proceedings of the National Academy of Sciences of the U.S.A.*, 117(44):27703–27711, 2020. ISSN 0027-8424. . URL https://doi.org/10.1073/pnas.2004904117.
O. R. Eichel. *A Special Report on the Mortality From Influenza in New York State During the Epidemic of 1918–19*. New York State Department of Health, Albany, NY, 1923.
B. D. Elderd, V. M. Dukic, and G. Dwyer. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proceedings of the National*

Academy of Sciences, 103(42):15693–15697, 2006.
S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley and Sons, New York, 1986.
S. Eubank, H. Guclu, V. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429 (6988):180–184, 2004.
R. FitzJohn, M. Baguelin, E. Knock, L. Whittles, J. Lees, and R. Sonabend. *mcstate: Monte Carlo Methods for State Space Models*, 2024. URL `https://github.com/mrc-ide/mcstate`. R package version 0.9.20.
W. H. Frost. Statistics of influenza morbidity: with special reference to certain factors in case incidence and case fatality. *Public Health Reports*, 35:584–597, 1920.
L. Gallo, M. Frasca, V. Latora, and G. Russo. Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models. *Science Advances*, 8(3):eabg5234, Jan. 2022. ISSN 2375-2548. .
D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.
D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.
E. Goldstein, J. Dushoff, J. Ma, J. Plotkin, D. J. D. Earn, and M. Lipsitch. Reconstructing influenza incidence by deconvolution of daily mortality time series. *PNAS – Proceedings of the National Academy of Sciences of the U.S.A.*, 106(51):21825–21829, 2009. .
L. Grinsztajn, E. Semenova, C. C. Margossian, and J. Riou. Bayesian workflow for disease transmission modeling in Stan. *Statistics in Medicine*, 40(27):6209–6234, 2021. ISSN 1097-0258. . URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9164`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.9164.
N. Gupta. On the Calibration of Compartmental Epidemiological Models. Master's thesis, New York University Tandon School of Engineering, United States – New York, 2023. URL `https://www.proquest.com/docview/2820207706/abstract/C69EF306BFF041E0PQ/1`. ISBN: 9798379583248.
D. He and D. J. D. Earn. Epidemiological effects of seasonal oscillations in birth rates. *Theoretical Population Biology*, 72:274–291, 2007. .
D. He and D. J. D. Earn. The cohort effect in childhood disease dynamics. *Journal of the Royal Society of London, Interface*, 13:20160156, 2016. .
D. He, J. Dushoff, T. Day, J. Ma, and D. J. D. Earn. Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proceedings of the Royal Society of London, Series B*, 280(1766):20131345, 2013. .
H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.
M. P. Hillmer, P. Feng, J. R. McLaughlin, V. K. Murty, B. Sander, A. Greenberg, and A. D. Brown. Ontario's COVID-19 Modelling Consensus Table: mobilizing scientific expertise to support pandemic response. *Canadian Journal of Public Health*, 112(5):799–806, 2021.
E. Howerton, L. Contamin, L. C. Mullany, M. Qin, N. G. Reich, S. Bents, R. K. Borchering, S.-m. Jung, S. L. Loo, C. P. Smith, et al. Evaluation of the US COVID-19 Scenario Modeling Hub for informing pandemic response under uncertainty. *Nature Communications*, 14(1): 7260, 2023.
P. Johnson. *adaptivetau: Tau-Leaping Stochastic Simulation*, 2023. URL `https://CRAN.R-project.org/package=adaptivetau`. R package version 2.3.
W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A*, 115:700–721, 1927.
T. Kim, B. Lieberman, G. Luta, and E. A. Peña. Prediction intervals for Poisson-based regression models. *WIREs Computational Statistics*, 14(5):e1568, 2022. ISSN 1939-0068. . URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1568`. _eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wics.1568.
A. A. King, M. D. de Cellès, F. M. G. Magpantay, and P. Rohani. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proc. R.*

Soc. B, 282(1806):20150347, May 2015. ISSN 0962-8452, 1471-2954. .

A. A. King, D. Nguyen, and E. L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(12):1–43, 2016. . URL https://www.jstatsoft.org/index.php/jss/article/view/v069i12.

C. M. Kribs and P. van den Driessche. Honoring the life and legacy of Fred Brauer. *Journal of Biological Dynamics*, 17(1):2285096, 2023. . URL https://doi.org/10.1080/17513758.2023.2285096.

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

N. P. Lemoine. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128(7):912–928, 2019. ISSN 1600-0706. . URL https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.05985. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/oik.05985.

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

A. Lindén and S. Mäntyniemi. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7):1414–1421, 2011.

W. London and J. A. Yorke. Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *American Journal of Epidemiology*, 98(6):453–468, 1973.

A. G. McKendrick. Applications of mathematics to medical problems. *Proc. Edinburgh Math. Soc.*, 13:98–130, 1926.

K. Nixon, S. Jindal, F. Parker, N. G. Reich, K. Ghobadi, E. C. Lee, S. Truelove, and L. Gardner. An evaluation of prospective COVID-19 modelling studies in the USA: from data to science translation. *The Lancet Digital Health*, 4(10):e738–e747, 2022.

S. Nourbakhsh, A. Fazil, M. Li, C. S. Mangat, S. W. Peterson, J. Daigle, S. Langner, J. Shurgold, P. D'Aoust, R. Delatolla, E. Mercier, X. Pang, B. E. Lee, R. Stuart, S. Wijayasri, and D. Champredon. A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities. *Epidemics*, 39:100560, June 2022. ISSN 1755-4365. .

I. Papst and D. J. D. Earn. Invariant predictions of epidemic patterns from radically different forms of seasonal forcing. *Journal of the Royal Society of London, Interface*, 16:20190202, 2019. . URL https://doi.org/10.1098/rsif.2019.0202.

T. Parsons, B. M. Bolker, J. Dushoff, and D. J. D. Earn. The probability of epidemic burnout in the stochastic SIR model with vital dynamics. *PNAS – Proceedings of the National Academy of Sciences of the U.S.A.*, 121(5):e2313708120, 2024. URL https://www.pnas.org/doi/10.1073/pnas.2313708120.

O. G. Pybus, M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey. The epidemic behavior of the hepatitis C virus. *Science*, 292(5525):2323–2325, 2001.

A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller, and J. Timmer. Lessons learned from quantitative dynamical modeling in systems biology. *PLOS ONE*, 8(9):e74335, Sept. 2013. ISSN 1932-6203. . URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074335.

M. Roberts and J. Heesterbeek. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *Journal of Mathematical Biology*, 55(5):803–816, 2007.

S. L. Rogers. *Special Tables of Mortality from Influenza and Pneumonia, in Indiana, Kansas, and Philadelphia, PA*. Department of Commerce, Bureau of the Census, Washington, DC, 1920.

D. P. Rosati, M. H. Woolhouse, B. M. Bolker, and D. J. D. Earn. Modelling song popularity as a contagious process. *Proceedings of the Royal Society of London, Series A*, 477:20210457, 2021. . URL https://royalsocietypublishing.org/doi/10.1098/rspa.2021.0457.

J. K. Taubenberger and D. M. Morens. 1918 influenza: the mother of all pandemics. *Emerging Infectious Diseases*, 12(1):15–22, 2006.

B. P. Taylor, J. Dushoff, and J. S. Weitz. Stochasticity and the limits to confidence when estimating R_0 of Ebola and other emerging infectious diseases. *Journal of theoretical biology*, 408:145–154, 2016.

The Advisory Committee Appointed by the Secretary of State for India, the Royal Society, and the Lister Institute. Reports on Plague Investigations in India. *The Journal of Hygiene*, 7(6):693–985, 1907. ISSN 00221724. URL `http://www.jstor.org/stable/4619420`.

N. Tuncer and T. T. Le. Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences*, 299:1–18, May 2018. ISSN 0025-5564. .

J. M. Ver Hoef. Who Invented the Delta Method? *The American Statistician*, 66(2):124–127, 2012. . URL `https://doi.org/10.1080/00031305.2012.687494`.

J. Wallinga and M. Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007.

L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, NY, Dec. 2010. ISBN 978-1-4419-2322-6.

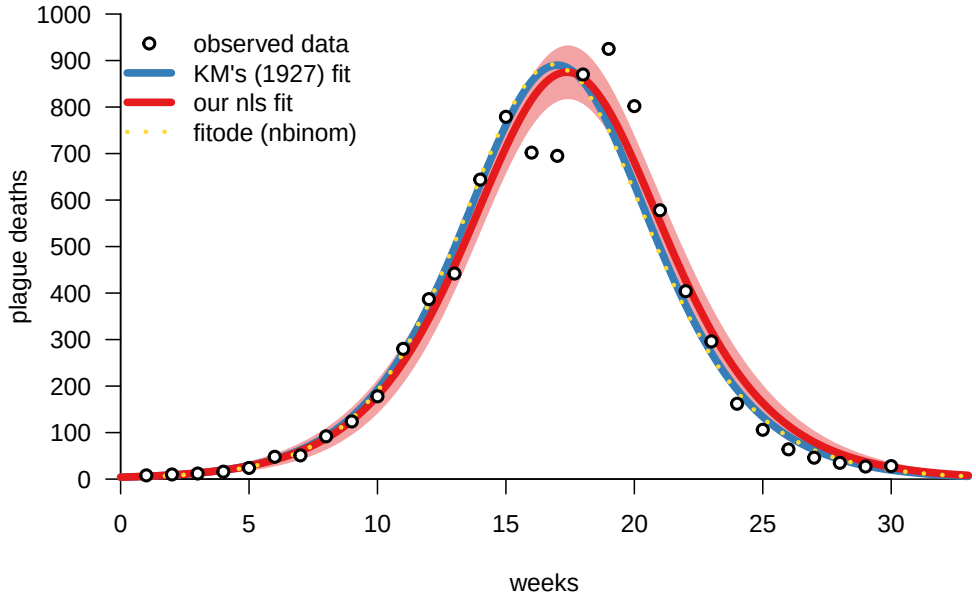
W. Zha, N. Zhou, G. Li, W. Li, H. Zhang, S. Zhang, M. Chen, R. Feng, T. Li, and Y. LV. Assessment and forecasting the spread of SARS-CoV-2 outbreak in Changsha, China: Based on a SEIAR Dynamic Model, Mar. 2020. URL `https://www.researchsquare.com/article/rs-16659/v1`. ISSN: 2693-5015.

Table 1. ~~Fits of 's Kermack and McKendrick [1927] analytical SIR approximation (2) to Bombay plague~~ **Fits of KM's analytical SIR approximation (2) to Bombay plague** (see Fig. 1). The KM column lists the parameter values estimated by ~~[Kermack and McKendrick, 1927, p. 714]~~KM [p. 714] ; the `nls` column lists the values estimated by us, using nonlinear least squares with confidence intervals obtained by the Delta method (see Sect. 4). Values for the initial prevalence ~~I_0~~ $I(0)$ and population size N are assumed in order to derive estimates of the standard SIR model parameters from the parameters of KM's approximation (using the indicated equations). Like ~~Bacaër [Bacaër, 2012, p. 408]~~Bacaër [2012, p. 408], we assume the population of Bombay was $N = 1$ million. We emphasize in this table that γ is the *per capita* removal rate, in order to contrast it with a, the *total* removal rate at the epidemic peak; elsewhere we refer to γ simply as the recovery rate.

Estimated parameter	**symbol**	**equation**	**units**	~~KM~~ **KM estimate**	**`nls` estimate**	**95% CI**
~~peak removal rate~~ total removal rate at epidemic peak	a	(6c)	$\frac{1}{\text{weeks}}$	890	875	(816, 935)
outbreak speed	ω	(6a)	$\frac{1}{\text{weeks}}$	0.2	0.19	(0.178, 0.21)
outbreak centre	ϕ	(6b)	–	3.4	3.37	(3.09, 3.67)
Assumed parameter						
initial prevalence	~~I_0~~ $I(0)$	–	–	1	1	–
population size	N	–	–	10^6	10^6	–
Derived parameter						
peak time	t_p	(8)	weeks	17	17.4	(17.1, 17.7)
effective reproduction number	$\mathcal{R}_\text{e}$	(5), (10a)	–	1.1	1.09	(1.04, 1.15)
per capita removal rate	γ	~~(10b)~~ (1), 10b	$\frac{1}{\text{weeks}}$	3.96	4.11	(1.95, 6.31)
initial susceptibles	~~S_0~~ $S(0)$	(10c)	–	53300	57400	(26000, 88800)
transmission rate	β	(1), (11)	$\frac{1}{\text{years}}$	0.00425	0.00407	(0.00372, 0.00443)
mean generation interval	T_g	(12)	days	1.77	1.7	(0.802, 2.59)
basic reproduction number	~~$\mathcal{R}_0$~~ $\mathcal{R}_0$	(3)	–	20.6	19	(7.65, 30.5)

Table 2. Fits of numerical SIR model solutions to Bombay plague (see Fig. 1). Parameter values were estimated using `fitode` to fit trajectories of Eq. (1), assuming observation errors were ~~distributed~~ negative binomially (`nbinom`) or normally (`ols`) distributed. The recovery rate γ was fixed rather than fitted due to parameter unidentifiability (see Footnote 11); we fixed γ to the value obtained from our `nls` fit of the KM approximation [Table 1] ~~for the purpose of fair comparison of the~~ in order to compare fits fairly. The `fitode`-fitted trajectories and confidence bands—for both `nbinom` and `ols`—are shown in the lower panel of Fig. 1. As in Table 1, a population size N must be assumed to derive ~~$\mathcal{R}_0$~~ $\mathcal{R}_0$ estimates.

Fixed parameter	**symbol**	**units**	`nbinom` estimate	**95% CI**	`ols` estimate	**95% CI**
recovery rate	γ	$\frac{1}{\text{weeks}}$	4.11	–	4.11	–
Estimated parameter						
transmission rate	β	$\frac{1}{\text{years}}$	0.004784	(0.00449, 0.00510)	0.0044564	(0.00393, 0.00506)
initial susceptibles	~~S_0~~ $S(0)$	–	49200	(46200, 52400)	52600	(46700, 59300)
initial prevalence	~~I_0~~ $I(0)$	–	0.941	(0.76, 1.17)	1.05	(0.627, 1.77)
overdispersion parameter	k	–	48.8	(24.4, 97.7)	–	–
Assumed parameter						
population size	N	–	10^6	–	10^6	–
Derived parameter						
effective reproduction number	$\mathcal{R}_e$	–	1.1	(1.1, 1.11)	1.1	(1.09, 1.1)
mean generation interval	T_g	days	1.7	–	1.7	–
basic reproduction number	~~$\mathcal{R}_0$~~ $\mathcal{R}_0$	–	22.4	(21, 23.8)	20.9	(18.2, 23.5)



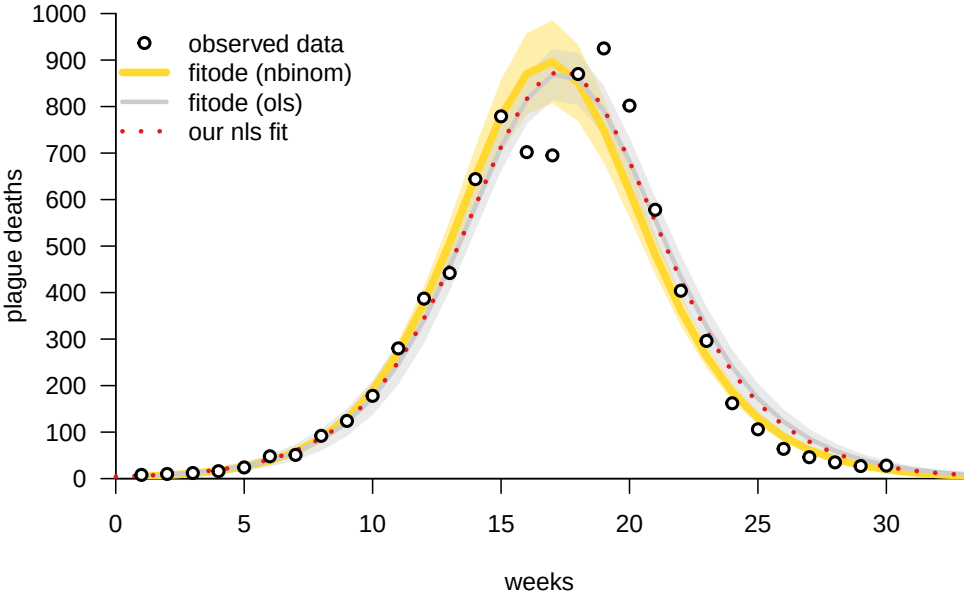


Figure 1. The plague epidemic in Bombay, 17 December 1905 to 21 July 1906, used as an example by ~~KM [Kermack and McKendrick, 1927, p. 714]~~KM [p. 714] . The data (large dots) were digitized from ~~[The Advisory Committee Appointed by the Secretary of State for India, the Royal Society, and the Lister Institute, 1907, Table IX, p.~~ The Advisory Committee Appointed by the Secretary of State for India, the Royal Society, and the Lister Institute [1907, Table IX, p. . *Top panel:* The ~~and red curves show the~~ KM approximation (2), as fitted by KM (blue curve) and by us using `nls` (~~red~~orangecurve, with ~~pink~~ confidence band estimated using the Delta method; see Sect. 4). The associated parameter estimates are given in Table 1. The dotted ~~yellow~~ gold curve shows the `fitode` fit of the SIR model (1), for which the associated parameter estimates are given in Table 2 [observation errors are assumed to be negative binomially distributed (17)]; this curve happens to coincide almost exactly with KM's fit. *Bottom panel:* The solid ~~yellow~~ gold curve is identical to the dotted ~~yellow~~ gold curve in the top panel; ~~the light yellow~~ its confidence band is the `fitode` confidence band obtained by the Delta method [the band is shown as a linear interpolation between successive observation times because the model (1) is fitted to incidence at discrete time points rather than to a continuous curve representation of the instantaneous death rate]. The light blue curve shows the `fitode` fit obtained by minimizing the ordinary least squares (7) [i.e., assuming observation errors are normally (14) distributed with variance σ^2 estimated from the residuals across all observation times]. The dotted ~~red~~ orange curve is identical to the solid ~~red~~ orange curve in the top panel. We have separated the two panels because the confidence band overlap would make the plots difficult to interpret.

Table 3. ~~Fits of 's Kermack and McKendrick [1927] analytical SIR approximation (2) to Philadelphia flu~~ Fits of KM's analytical SIR approximation (2) to Philadelphia flu (see Fig. 2). Parameter estimates were obtained using nonlinear least squares (`nls`) to fit Eq. (2) to the reported daily pneumonia and influenza (P&I) mortality during the main wave of the pandemic in 1918. In order to derive estimates of the standard epidemiological parameters, we assumed the initial prevalence had the value estimated by `fitode` for the SIR model (see Table 4). We do not use the raw population size in our estimate of ~~$\mathcal{R}_0$~~ $\mathcal{R}_0$; instead, we account for the fact that reported deaths are roughly equal to incidence times the case fatality proportion (CFP) by taking N to be the size of population that would eventually die if everyone in the city were infected, i.e., the product of the population size of Philadelphia in 1918 (1,768,825) and an assumed CFP of 0.025 ~~Taubenberger and Morens [2006]~~[Taubenberger and Morens, 2006]. The fitted trajectory and confidence band are shown in Fig. 2. See Sect. 7.

Estimated parameter	**symbol**	**equation**	**units**	`nls`	**95% CI**
~~peak removal rate~~ total removal rate at epidemic peak	a	(6c)	$\frac{1}{\text{years}}$	738	(715, 761)
outbreak speed	ω	(6a)	$\frac{1}{\text{years}}$	42	(40.3, 43.5)
outbreak centre	ϕ	(6b)	–	3.64	(3.51, 3.79)
Assumed parameter					
initial prevalence	~~I_0~~ $I(0)$	–	–	3.05	–
effective population size	N	–	–	44,221	–
Derived parameter					
peak time	t_p	(8)	weeks	4.524	(4.49, 4.56)
effective reproduction number	$\mathcal{R}_\text{e}$	(5), (10a)	–	128	(90.8, 165)
per capita removal rate	γ	(10b)	$\frac{1}{\text{years}}$	0.66	(0.492, 0.831)
initial susceptibles	~~S_0~~ $S(0)$	(10c)	–	2270	(1660, 2880)
transmission rate	β	(11)	$\frac{1}{\text{years}}$	0.0372	(0.0285, 0.0459)
mean generation interval	T_g	(12)	years	1.52	(1.12, 1.9)
basic reproduction number	~~$\mathcal{R}_0$~~ $\mathcal{R}_0$	(3)	–	2490	(2416, 2571)

Table 4. Fits of numerical SIR model solutions to Philadelphia flu (see Fig. 2). Parameter estimates are based on `fitode` fits of the SIR model (1) to reported P&I mortality during the main wave of the 1918 influenza pandemic in the city of Philadelphia. As in Table 3, in order to derive an estimate of ~~$\mathcal{R}_0$~~$\mathcal{R}_0$, we assume an effective population size that accounts for the data representing deaths rather than cases.

Estimated parameter	**symbol**	**units**	`nbinom`	**95% CI**
transmission rate	β	$\frac{1}{\text{years}}$	0.0124	(0.0119, 0.0128)
recovery rate	γ	$\frac{1}{\text{years}}$	85.6	(75.9, 96.5)
initial susceptibles	~~S_0~~ $S(0)$	–	15300	(14500, 16200)
initial prevalence	~~I_0~~ $I(0)$	–	3.05	(2.32, 4.01)
overdispersion parameter	k	–	157	(44.2, 557)
Assumed parameter				
effective population size	N	–	44,221	–
Derived parameter				
effective reproduction number	$\mathcal{R}_\text{e}$	–	2.21	(2.02, 2.4)
mean generation interval	T_g	days	4.27	(3.75, 4.78)
basic reproduction number	~~$\mathcal{R}_0$~~ $\mathcal{R}_0$	–	6.38	(5.53, 7.24)

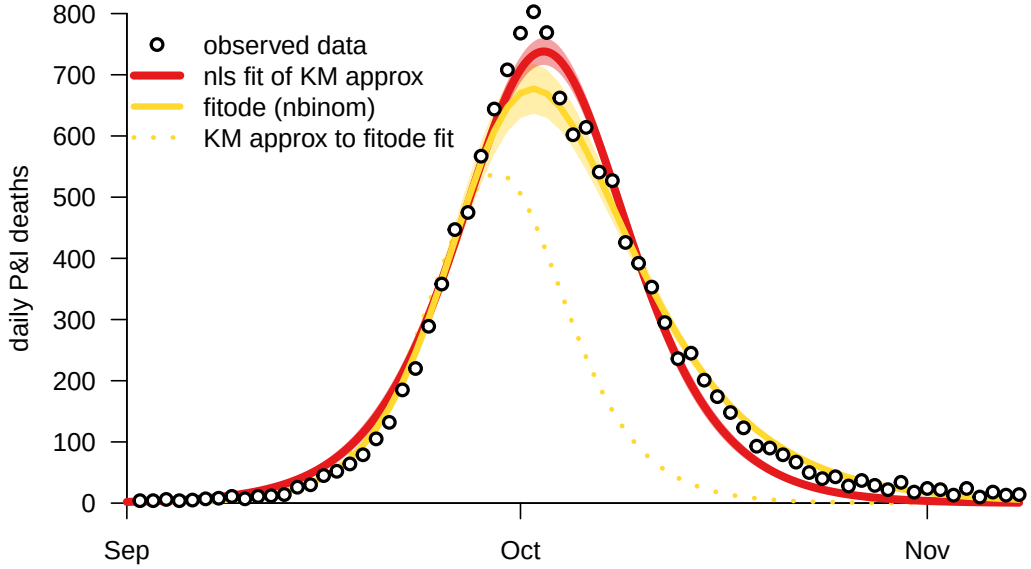


Figure 2. The main wave of the 1918 influenza epidemic in the city of Philadelphia, 1 September 1918 to 31 December 1918 ~~Rogers [1920], Goldstein et al. [2009]~~[Rogers, 1920, Goldstein et al., 2009]. Reported daily deaths from pneumonia and influenza (P&I) are shown with large dots. The ~~red~~ orange curve and ~~pink~~ corresponding confidence band show a nonlinear least squares (`nls`) fit of KM's approximation (2); the parameter estimates are given in Table 3. The solid ~~yellow~~ gold curve and ~~light yellow~~ corresponding confidence band show the `fitode` fit of the SIR model (1), for which the parameter estimates are given in Table 4. The dotted ~~yellow~~ gold curve shows the KM approximation using the parameters estimated with `fitode`.

Table 5. ~~Fits of 's Kermack and McKendrick [1927] analytical SIR approximation (2) to an epidemic simulated using the standard stochastic SIR model Andersson and Britton [2000]~~ **Fits of KM's analytical SIR approximation (2) to an epidemic simulated using the standard stochastic SIR model [Andersson and Britton, 2000]** (see Sect. 8 and Fig. 3). The parameter values in the "true" column are those used to generate the stochastic simulation (~~S_0~~ $S(0)$, ~~I_0~~ $I(0)$, ~~$\mathcal{R}_0$~~ $\mathcal{R}_0$ and T_g) and the values of other parameters derived from these true parameter values using the indicated equations. The `nls` column lists our estimates and confidence intervals obtained by fitting Eq. (2) to the simulated data using nonlinear least squares and the Delta method.

Assumed parameter	**symbol**	**equation**	**units**	**true**	`nls`	**95% CI**
initial prevalence	~~I_0~~ $I(0)$	–	–	2	2	–
Estimated parameter						
~~peak removal rate~~ total removal rate at epidemic peak	a	(6c)	$\frac{1}{\text{weeks}}$	641	135	(125, 144)
outbreak speed	ω	(6a)	$\frac{1}{\text{weeks}}$	2	0.99	(0.907, 1.08)
outbreak centre	ϕ	(6b)	–	3.58	2.7	(2.48, 2.93)
Derived parameter						
peak time	t_p	(8)	weeks	1.79	2.72	(2.66, 2.78)
effective reproduction number	$\mathcal{R}_e$	(5), (10a)	–	5	2.62	(1.79, 3.44)
per capita removal rate	γ	(10b)	$\frac{1}{\text{weeks}}$	1	1.21	(0.7, 1.77)
initial susceptibles	~~S_0~~ $S(0)$	(10c)	–	2000	571	(518, 624)
transmission rate	β	(11)	$\frac{1}{\text{years}}$	0.13	0.289	(0.233, 0.346)
mean generation interval	T_g	(12)	days	7	5.77	(3.2, 8.12)
basic reproduction number	~~$\mathcal{R}_0$~~ $\mathcal{R}_0$	(3)	–	5	9.18	(6.95, 11.4)

Table 6. ~~Fits of numerical (deterministic) SIR model solutions to an epidemic simulated using the standard stochastic SIR model Andersson and Britton [2000]~~ Fits of numerical (deterministic) SIR model solutions to an epidemic simulated using the standard stochastic SIR model [Andersson and Britton, 2000] (see Sect. 8 and Fig. 3). Parameter estimates we obtained using `fitode` to fit the SIR model (1) to the simulated data, assuming deviations from the ~~determinsitic~~ deterministic curve were generated by negative binomially (17) distributed observation errors.

Estimated parameter	**symbol**	**units**	**true**	`nbinom`	**95% CI**
transmission rate	β	$\frac{1}{\text{years}}$	0.13	0.131	(0.119, 0.144)
recovery rate	γ	$\frac{1}{\text{weeks}}$	1	0.971	(0.884, 1.07)
initial susceptibles	~~S_0~~ $S(0)$	–	1998	2000	(1900, 2110)
initial prevalence	~~I_0~~ $I(0)$	–	2	0.605	(0.306, 1.2)
overdispersion parameter	k	–	–	251	(19.6, 3226.6)
Derived parameter					
mean generation interval	T_g	days	7	7.21	(7.92, 6.56)
effective reproduction number	$\mathcal{R}_\text{e}$	–	4.995	5.2	(4.44, 5.95)
basic reproduction number	~~$\mathcal{R}_0$~~ $\mathcal{R}_0$	–	5	5.19	(4.37, 6.01)

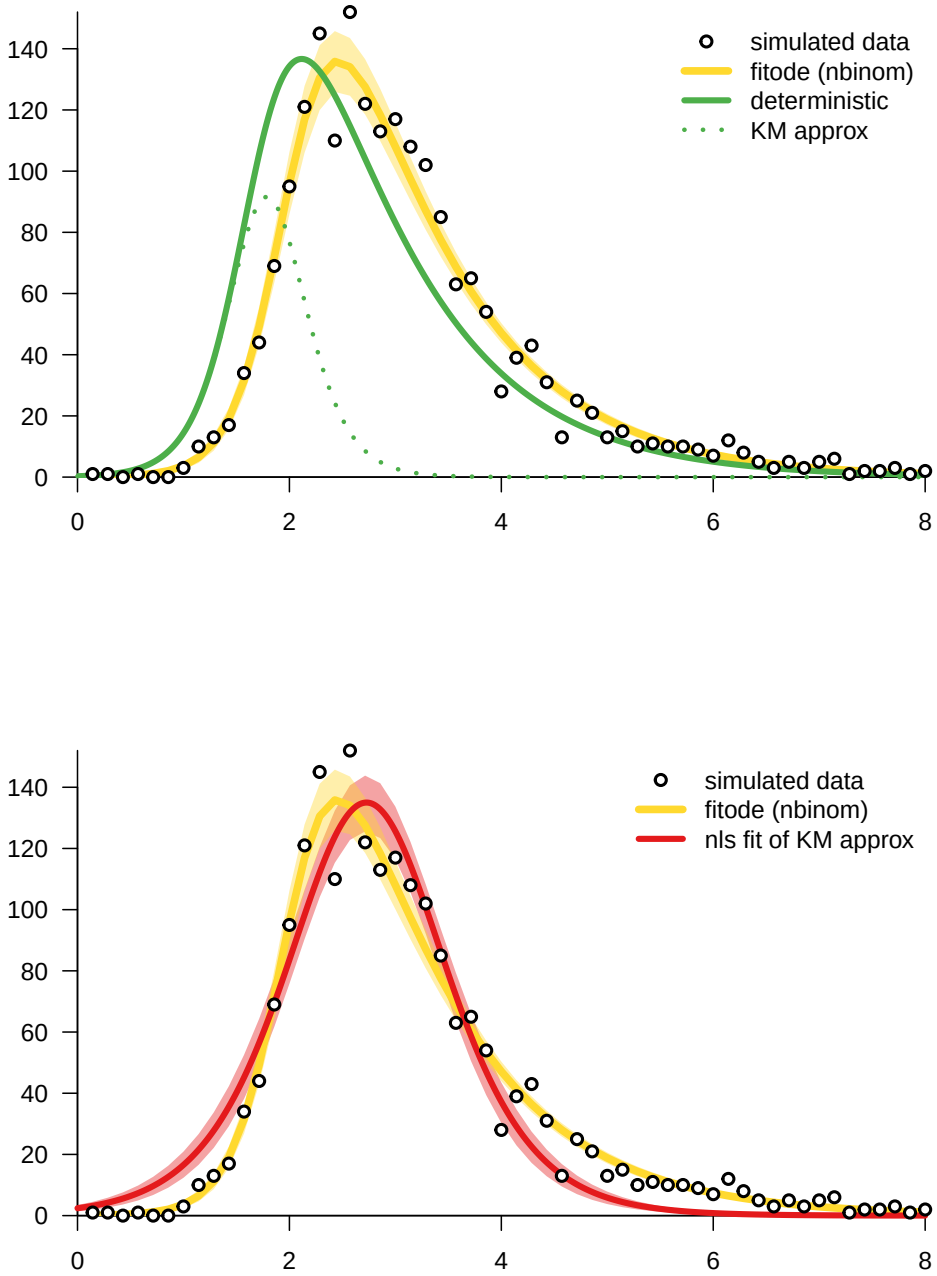


Figure 3. Deterministic fits to daily incidence generated by a stochastic SIR simulation with initial state ~~$(S_0, I_0, R_0) = (1998, 2, 0)$~~$(S(0), I(0), R(0)) = (1998, 2, 0)$, basic reproduction number ~~$\mathcal{R}_0 = 5$~~$\mathcal{R}_0 = 5$, and mean generation interval $T_g = 1$ week. The simulated data points show the numbers of newly recovered individuals each day. In both panels, the ~~yellow~~ gold curve and confidence band show the `fitode` fit to the simulated data. *Top panel:* The solid ~~green~~ green curve shows the solution of deterministic SIR model (1) with the initial conditions and parameters used for the stochastic simulation. The dotted ~~green~~ green curve shows the KM approximation (2) to this deterministic trajectory. The time shift between the ~~green~~ green and ~~yellow~~ gold curves arises because there is a random delay until the stochastic trajectory begins to grow exponentially. *Bottom panel:* The ~~red~~ orange curve shows the KM approximation (2), fitted to the stochastic simulation using `nls`. Since the KM approximation is symmetric about its maximum, it is impossible to obtain a good fit in situations like this, where the rise of the epidemic is faster than the fall.

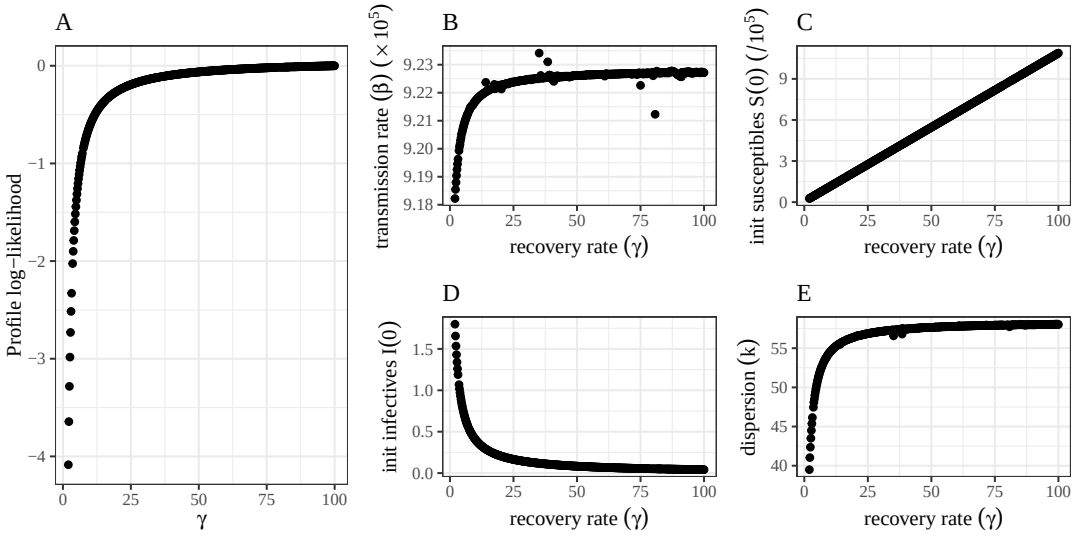


Figure 4. ~~Unidentifiability of the mean generation interval T_g (or, equivalently, the removal rate γ) for the Bombay plague epidemic shown in Fig. 1.~~ **Unidentifiability of the mean generation interval T_g (or, equivalently, the *per capita* removal rate γ) for the Bombay plague epidemic shown in Fig. 1.** (A) The profile likelihood—briefly discussed at the end of Sect. 4—is calculated by fixing γ to a series of given values and, for each value, maximizing the likelihood by estimating all other parameters ~~Bolker [2008]~~ [Bolker, 2008]. (The maximum value is shifted to 0 without loss of generality.) A flat profile-likelihood surface indicates parameter unidentifiability, meaning that we can obtain very similar fits across a wide range of values of the focal parameter (γ). (B–E) The corresponding best parameter estimates for a given value of γ.

Response to reviewers of Earn et al (2024) Fitting epidemic models to data – a tutorial in memory of Fred Brauer

We are very grateful to the reviewers for their very positive feedback and helpful suggestions for minor improvements, which we have attempted to implement.

Reviewer 1

The manuscript "Fitting epidemic models to data – a tutorial in memory of Fred Brauer" was a pleasure to read, and will make an excellent contribution to the literature. It is well written, appropriately balances breadth and depth, introduces a new computational tool, and it and fills a sizable information gap that exists right now for mathematicians (with limited or no background in applied statistics) – and many statisticians – who are interested in modern approaches to fitting nonlinear dynamic models to time series data. I commend the authors for this contribution, and for their remebrance of Fred.

I did not find any major faults with the paper, but have suggested some edits below that I feel should improve the paper. The first four comments are most substantive.

1. I urge the authors to be more explicit about the "observation process model"/"observation model" as an important link in connecting ODEs to data. In practice, data are often not direct observations of state variable, e.g., they might be the number of reported cases which might be proportional to the true case count, as already mentioned by the authors. It is therefore common practice, when building your likelihood function, to create an "observation model" of your data. This is typically done using a joint PDF/PMF parameterized by ODE model outputs, as this let's us define a likelihood function to use in an MLE framework. The authors first mention an "observation process model" in section 5, well after the framework has been introduced, and the authors' package (fitode) takes an observation model argument as illustrated in the code below line 357 and as described above equation (27). Earlier, explicit mention of the observation model as a concept would go a long way to put those existing

references into appropriate context.

I trust the authors judgement in making the above modification, but just in case it's helpful, here are my suggestions for how to more explicitly introduce the "observation model" concept: First, introduce the idea briefly with one or two sentences early in section 3. Something as simple as the following might suffice: "Here we assume our data are direct observations of one of our state variables. However, this is frequently not the case. When a more nuanced relationship defines the link between model and data, we can specify an ***observation model*** that describes how our data values relate to the ODE trajectories. We will revisit this concept in the next section." Then, in section 4, mention it again in the context of constructing the likelihood function: Minor adjustments to the text preceding eq. (14) could reframe that derivation using the concept of an observation model more explicitly. With those two modifications in place, then the text leading up to the "observation" option in in fitode (see eq. (27)) could be modified to refer to the observation model as an explicit concept.

Done.

2. Notation: Somewhat related to observation model comments above is the use of "x()" for ODE state variable values, and "x[]" for data values. I find this to be atypical and probably not helpful to most readers. I strongly encourage the authors to consider notation that is easier to grasp. My recommendation: Using the upper case "X()" for data crossed my mind, however I think a better choice might be "y()", which greatly improves the readability compared to discerning "()" from "[]". While a lower level undergrad might find this use of "x vs y" notation a little problematic, I think the target audience here will find it much easier notation to follow than the current notation. Moreover, y is a commonly used symbol for the response values in statistics, so this use of "y()" will not be terribly foreign to readers familiar with existing statistical theory notation.

Rather than $x[]$, we now use $x_{\mathbf{obs}}$ to denote observed values of the variable x.

3. I'm confused by the $\Delta x[t_\ell]$ term at the end of eq. (14), which doesn't quite follow from the assumptions in the preceding text. It seems as if the authors were at one point assuming that the probability of an observation x[t] can be approximated with a discrete distribution, letting $P(x = x[t]) = f(x[y], \sigma)\Delta x$ and then this discretized Normal distribution could be used to construct the likelihood function. The rest of the text, however, does not reflect this, nor does it otherwise explain what the Δx term represents. If this was the intention, then there needs to be (1) clarification in the preceding text that this is what is being assumed, and (2) some follow-up text after eq. (17) clarifying how this Δx term turns into a constant in the log-likelihood function and can ultimately be ignored in the optimization step that follows. However, as it stands, I'm not convinced that this is the most practical way to introduce the concept

of likehood and MLE. Alternatively, please consider dropping the Δx from eq. (14) and changing the left side of equation (14) to indicate that it's a joint density function, not a probability, and then define the likelihood from there with a followup sentence or footnote about how likelihoods are defined the discrete case usng PMFs instead of PDFs. This approach (defining the likelihood using the joint PDF or PMF) is a common way to introduce the concept of a likelihood, and in this context, that approach might be preferable.

Done

I should add that, for all three comments above, there is value in using notational conventions and other formalisms familiar to statisticians, where possible. This would help mathematicians who read this paper to more easily see how it relates to existing statistics literature, and likewise would make the paper more accessible to statisticians who have limited exposure to dynamic models and these techniques.

4. The discussion could include a bit more guidance regarding identifiability issues. For example, line 533, the parenthetical discription for "unidentifiable" might be reconsidered, and replaced with something that elaborates a bit more on the problem of (a) structural unidentifiability and (b) practical identifiability, each with their own description touching upon the fact that many ODE models (used in this context) are overparameterized and yield non-unique parameter estimates, even under ideal data assumptions. The second issue can arise for even strucuturally identifiable models which may be practically unidentifiable due to, e.g., a lack of data from certain parts of state space. A simple verbal example to illustrate this last point (if you wanted to go into that much detail) is to ask the reader to consider fitting the logistic growth model $dx/dt = rx(1 - x/K)$, with known initial value x_0, to time series data that ends while the trajectory is only in the exponential growth phase. In that case, the parameter r will be confidently estimated, but K will not since all sufficiently large K values will appear to give equally good fits, which reflects that the data contain no information about that steady state value, K. Some or all of this might be better placed in the paragraph starting at line 600, where you discuss convergence issues. Consider also adding a brief mention of identifiability issues somewhere in section 4, and please consider adding some references for dealing with the different types of identifiability issues.

5. Related to the above comment, on line 603, it might be useful to call this procedure the "multistart method" (as it is sometimes called) and to mention that it is not only useful for diagnosing convergence issues (where you would see dissimilar "best" parameter sets with differing likelihood values), but it's also useful for detecting identifiability issues (here, differing best parameter sets would show very similar likelihood values).

Done

6. Consider mentioning some other distribution options above eq. (17), around line 242-244, like the various Generalized Poisson distributions. Include references (some of these are available in other R packages that might be worth mentioning). I find that mathematicians often are unaware that these other options exist, and so they only consider Normal, Poisson, and Negative Binomial.

Done

7. Minor inconsistencies in referring to equations: These include model (2), Equation (2), equation (2), distance (7), approximation (6), dR/dt (2), etc. I don't recall this journal having strict guidelines for equation referencing, so at a minimum I would ask the authors to consider dropping the instances of capital-E "Equation" to lower case.

We have now attempted to follow BMB style for everything.

8. There are a few places where table captions and R code run off into the margins.

We adjusted these slightly, but this seems best to deal with at the copy-editing stage.

9. Generally speaking, there might be room to include a few more references throughout, especially given the tutorial nature of this manuscript. Readers looking to apply this approach would benefit from some extra guidance finding literature related to some of the concepts or methods that are only briefly mentioned here.

We have added a few references.

Reviewer 2

This manuscript introduces the software package fitode, an R-based tool developed to aid the fitting of ODE models to observed time series data, particularly for epidemiological applications. The software serves as a practical response to the historical curiosity about how dynamic models fit to empirical data, a question posed by the late mathematical biologist Fred Brauer. The manuscript not only shows the functionality of fitode through examples involving compartmental epidemic models but also provides a tutorial approach to guide readers, presumably with a background in mathematics but less familiarity with statistical methods and optimization techniques. The discussion delves into the technical details of model fitting, parameter estimation, and the challenges inherent in this domain.

Major Comments:

1. The authors acknowledge the issue of non-uniqueness in parameter values achieving the minimum error in model fitting. This is a critical limitation

as it affects the reliability of the model predictions and the interpretability of the results. To address this challenge, the authors might consider incorporating regularization techniques which can help in constraining the parameter space and reducing the likelihood of overfitting. It would also be helpful if the manuscript could discuss strategies to identify and handle the impact of parameter correlations, which often contribute to this non-uniqueness.

We have added a paragraph on identifiability, and expanded a comment about regularization.

2. While fitode is presented as a user-friendly tool tailored for epidemiologists, a systematic comparison with existing methods available in platforms like MATLAB or Berkeley Madonna is missing. Each of these tools has its strengths and limitations which could significantly influence user choice depending on their specific needs. For instance, MATLAB offers a broad range of built-in functions for optimization and model fitting along with high computational power, but it may not be as accessible due to licensing costs. On the other hand, Berkeley Madonna is known for its ease of use and speed in running complex dynamic models but might lack some of the advanced statistical tools provided by R. The authors should elaborate on the comparative advantages of fitode, possibly in terms of its ease of integration with other statistical methods in R, its specific utility for epidemiological models, or any unique features that address the nuances of fitting disease transmission models to data.

We have added some comments about MATLAB and Berkeley Madonna.