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:
March 19, 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. 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`

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 (KM) [41, p. 713] 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 rate. In that original paper, KM [41, 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 [1, 2, 4, 5, 10, 11, 18] and review articles [20–22, 38]. Researchers have primarily focused on ***compartmental models*** like the SIR model, cast either as differential equations following the tradition of KM [41], or as ***stochastic processes*** in the tradition of McKendrick [50] and Bartlett [5]. 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 [28].

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 [6, 7]. Careful consideration of uncertainty is especially important when epidemic models are used for the development and analysis of policy options for infectious disease management [26], 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 [13, 39, 51].

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 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

[1] In the words of KM [41, p. 701], "removed from the number of those who are sick, by recovery or by death".

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 [41] application of their SIR model (1) to the epidemic of plague in Bombay in 1905–1906. The observed data (dots in Figure 1) were weekly numbers of deaths from plague.

Referring to their version of Figure 1, KM [41, 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 [15], and an exact analytical form for this function could not be found, they proceeded to assume [41, 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

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

i.e., Equation (2) is a good approximation as long as the proportion of the population that has been infected and removed is much less than $1/\mathcal{R}_0$.

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

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

In terms of $\mathcal{R}_\mathrm{e}$, γ, S_0 and I_0, the parameters in Equation (2) can be written[4]

$$\omega = \frac{\gamma}{2}\sqrt{(\mathcal{R}_\mathrm{e}-1)^2 + \frac{2I_0}{S_0}\mathcal{R}_\mathrm{e}^2}\,, \tag{6a}$$

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

$$\text{and} \qquad a = \frac{2\,\omega^2 S_0}{\gamma\,\mathcal{R}_\mathrm{e}^2}\,. \tag{6c}$$

[2] We had originally intended to submit this paper to a collection in honour of Fred's memory [44].

[3] $\mathcal{R}_0$ is the expected number of secondary cases resulting from a primary case in a wholly susceptible population [1].

[4] There is a typographical error in equation (31) of KM [41]: their factor $\sqrt{-q}$ should be $(-q)$ in their equivalent of the parameter we call a. Bacaër [3, §3] corrected this error without comment.

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 Figure 1.

3. How to fit the model to the data

The blue curve in Figure 1 does appear to provide a reasonable fit to the data, but KM [41] 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—Equation (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

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

where the observed data are the points $\{(t_\ell, x[t_\ell]) : \ell = 1,\ldots,n_t\}$, $\boldsymbol{\theta}$ 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 Equation (2). (Note that we write $x[\cdot]$ when referring to observations of the variable x and $x(\cdot\,;\cdot)$ when referring to the model.) 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)}
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 red in Figure 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 Equation (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 a very 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 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.* [55] 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

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

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. In Equation (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 Equation (6) can be inverted[6] analytically[7] [3, §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 Equation (10) above we have taken this to be the initial prevalence I_0. From Equation (10), we can compute the transmission rate,

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

and the mean intrinsic generation interval [16],

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

which is the same as the mean infectious period in this simple model [17, 45]. 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.,

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

[8] In his reanalysis of KM's results, Bacaër [3] also retained this conceptual error.

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.

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 shaded region in the top panel of Figure 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 the probability of observing the data $\{x[t_\ell]\}$. We imagine that the model (2) is a perfect representation of reality, and we consider the deviations from the model curve in Figure 1 to be observation errors. We then imagine that 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 of the data given the model is

$$\mathbb{P}(\text{data} \mid \text{model}) = \prod_{\ell=1}^{n}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\big(x(t_\ell;\boldsymbol{\theta})-x[t_\ell]\big)^2}{2\sigma^2}\right)\Delta x[t_\ell]\right]\,. \tag{14}$$

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 [7]. We define the ***likelihood*** $\mathcal{L}$ of a set of parameter values $\boldsymbol{\theta}$ as

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

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

likelihood, yields an estimate,

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

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

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

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

which—lo and behold—agrees exactly with (7), the least squares solution! The standard way of expressing this is to say that the least squares 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 [Equation (1c) or Equation (2)]. This 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. The negative binomial distribution is one way to generalize the Poisson to allow for overdispersion [48].

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 $x[t_\ell]$ 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,

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

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*** $\mathrm{Cov}(\boldsymbol{\theta})$ 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 [7, §6.5] for a heuristic explanation or [62, §§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*** [19, 61] expands $\mathrm{Var}(g(\boldsymbol{\theta}))$ to first order about $\hat{\boldsymbol{\theta}}$, which gives us the variance-covariance matrix of $g(\boldsymbol{\theta})$ [7, §7.5.2] [62, §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 $g(\hat{\boldsymbol{\theta}}) = (\hat{\mathcal{R}}_{\mathrm{e}}, \hat{\gamma}, \hat{S}_0)$. 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 shown in Figure 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 [7, §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 [7, §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 [7, §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*** [55, 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***,

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

Equation (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., Equation (7)], we write $\boldsymbol{x}[t_\ell]$ 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

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

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

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

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

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

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

where we typically assume the probability distribution

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

We have slightly abused notation here, compared with Equation (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.* [55] give a detailed comparison between using the sensitivity equations and computing gradients by finite-difference approximations. (Bjørnstad [6, 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 observed data (`mort`) are assumed to arise from sampling from the negative binomial distribution [`dnbinom`, Equation (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 fitted model trajectory [Equation (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 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

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

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), sta
```

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 Figure 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 Equation (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 [41] 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 [3] argues that the trajectory of the Bombay plague epidemic was primarily governed by seasonality rather than SIR

[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 [42].
[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.

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 [41, 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 [25, 30, 56].

7. Influenza in Philadelphia, October 1918

Deaths caused ultimately by influenza are often attributed to pneumonia [22], 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 [56]. 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 (dots in Figure 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 curve in Figure 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 trajectories of the exact SIR model using `fitode` gives a fit—the solid yellow curve in Figure 2—that is visually similar to the (red) 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 yellow curve in Figure 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 $(\mathcal{R}_0, T_g)$, compute the associated trajectory by solving Equation (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 $(S_0, I_0, \mathcal{R}_0, T_g)$.

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 [2] can be defined by interpreting the individual terms in Equation (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 [27]). Realizations of this discrete-state model can be generated exactly using the Gillespie algorithm [31], or approximately (as we do here) using the "τ-leaping" approach [32], which is implemented in the `adaptivetau` R package [40]. 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 [54]).

In Figure 3, the dots 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 $T_g = \gamma^{-1} = 1$ week. In the top panel, $\mathrm{d}R/\mathrm{d}t$ [Equation (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 yellow. 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 $a\,\mathrm{sech}^2(\omega\, t - \phi)$ [Equation (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 (Figure 3, bottom panel, red 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 curve in Figure 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 (yellow) 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, [41]) 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 [10–12], including non-autonomous models, such as seasonally forced epidemic models [23, 36, 37, 49, 53]. 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 [57]).

We focused here on three illustrative examples of epidemic time series. The first was the reported weekly mortality from plague in Bombay in 1906 (Figure 1), which was examined by KM in their original paper [41]. Although historically important, it is certainly debatable whether 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 §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, Bacaër investigated the longer term pattern of plague mortality in Bombay and found that there were seasonal epidemics every year from 1897 to 1911 [3, Fig. 2], suggesting that the 1906 epidemic was just one in a long sequence of epidemics that were "driven by seasonality" [3, 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 [14] to rank the relative importance of the various mechanisms included in the sequence of models (see, e.g., He *et al.* [35] 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 [8].) 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 to estimate reliably from the reported weekly plague deaths alone (see Figure 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 (Figure 2). Again we fitted the exact 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 reasonable fits (Figure 3) but KM's approximation yields absurd parameter values, whereas `fitode` estimates 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 [24].)

KM's approximation (2) estimates the simulation parameters badly because the assumption on which it is based (4) is strongly violated in the simulation (Figure 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 [57]; a good fit is not, on its own, sufficient to conclude that a model matches the underlying processes of a dynamical system.

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 [52]).

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).

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 [26] 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 [46].

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 [7, §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 [34].)

Even with these extensions, modelers will 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 optimization from multiple starting conditions and drawing likelihood surfaces can help diagnose these problems. Using different optimization methods or reparameterizing the model can also help [9, 55]. We also 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 uncertainties in parameter estimates and confidence bands on the projected dynamics of a system [42, 47, 59]. Popular `R` packages that can fit models with process error include `pomp` [43] and `mcstate` [29].

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 [55], and estimating statistically meaningful uncertainty in those optima is extremely challenging [26, 47].

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

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

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

[3] 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 (2012), pp. 403–422. http://link.springer.com/article/10.1007/s00285-011-0417-5.

[4] N.T.J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd ed., Hafner Press, New York, 1975.

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

[6] O.N. Bjørnstad, *Epidemics: Models and Data Using R*, 1st ed., Springer, New York, NY, 2018 Nov.

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

[8] B. Bolker, *Multimodel approaches are not the best way to understand multifactorial systems* (2023). Available at `https://ecoevorxiv.org/repository/view/5722/`, Publisher: EcoEvoRxiv.

[9] 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 (2013), pp. 501–512. Available at `http://doi.wiley.com/10.1111/2041-210X.12044`.

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

[11] F. Brauer, C. Castillo-Chavez, and Z. Feng, *Mathematical models in epidemiology*, Vol. 32, Springer, 2019.

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

[13] 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 (2021), p. 20210001.

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

[15] M. Campbell-Kelly, *Origin of computing*, Scientific American 301 (2009), pp. 62–69.

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

[17] 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 (2018), pp. 3258–3278. Available at `https://epubs.siam.org/doi/10.1137/18M1186411`.

[18] 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.

[19] R.A. Dorfman, *A note on the δ-method for finding variance formulae.*, The Biometric Bulletin 1 (1938), pp. 129–137.

[20] D.J.D. Earn, *A Light Introduction to Modelling Recurrent Epidemics*, in *Mathematical Epidemiology*, F. Brauer, P. van den Driessche, and J. Wu, eds., Lecture Notes in Mathematics Vol. 1945, Springer, 2008, pp. 3–17. Available at `https://link.springer.com/chapter/10.1007%2F978-3-540-78911-6_1`.

[21] D.J.D. Earn, *Mathematical epidemiology of infectious diseases*, in *Mathematical Biology*,

M.A. Lewis, M.A.J. Chaplain, J.P. Keener, and P.K. Maini, eds., IAS/Park City Mathematics Series Vol. 14, American Mathematical Society, 2009, pp. 151–186. Available at http://www.ams.org/books/pcms/014/.

[22] D.J.D. Earn, J. Dushoff, and S.A. Levin, *Ecology and evolution of the flu*, Trends in Ecology and Evolution 17 (2002), pp. 334–340.

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

[24] 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 (2020), pp. 27703–27711. Available at https://doi.org/10.1073/pnas.2004904117.

[25] 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.

[26] 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 (2006), pp. 15693–15697.

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

[28] 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 (2004), pp. 180–184.

[29] R. FitzJohn, M. Baguelin, E. Knock, L. Whittles, J. Lees, and R. Sonabend, *mcstate: Monte Carlo Methods for State Space Models* (2024). Available at https://github.com/mrc-ide/mcstate, R package version 0.9.20.

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

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

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

[33] 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 (2009), pp. 21825–21829.

[34] L. Grinsztajn, E. Semenova, C.C. Margossian, and J. Riou, *Bayesian workflow for disease transmission modeling in Stan*, Statistics in Medicine 40 (2021), pp. 6209–6234. Available at https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9164, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.9164.

[35] 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 (2013), p. 20131345.

[36] D. He and D.J.D. Earn, *Epidemiological effects of seasonal oscillations in birth rates*, Theoretical Population Biology 72 (2007), pp. 274–291.

[37] D. He and D.J.D. Earn, *The cohort effect in childhood disease dynamics*, Journal of the Royal Society of London, Interface 13 (2016), p. 20160156.

[38] H.W. Hethcote, *The mathematics of infectious diseases*, SIAM Review 42 (2000), pp. 599–653.

[39] 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 (2021), pp. 799–806.

[40] P. Johnson, *adaptivetau: Tau-Leaping Stochastic Simulation* (2023). Available at https://CRAN.R-project.org/package=adaptivetau, R package version 2.3.

[41] 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 (1927), pp. 700–721.

[42] 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 (2015), p. 20150347.

[43] 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 (2016), p. 1–43. Available at `https://www.jstatsoft.org/index.php/jss/article/view/v069i12`.

[44] C.M. Kribs and P. van den Driessche, *Honoring the life and legacy of Fred Brauer*, Journal of Biological Dynamics 17 (2023), p. 2285096. Available at `https://doi.org/10.1080/17513758.2023.2285096`.

[45] 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 (2013), p. 20130098.

[46] N.P. Lemoine, *Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses*, Oikos 128 (2019), pp. 912–928. Available at `https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.05985`, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/oik.05985.

[47] 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 (2018), pp. 1956–1967.

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

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

[50] A.G. McKendrick, *Applications of mathematics to medical problems*, Proc. Edinburgh Math. Soc. 13 (1926), pp. 98–130.

[51] 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 (2022), pp. e738–e747.

[52] 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 (2022), p. 100560.

[53] 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 (2019), p. 20190202. Available at `https://doi.org/10.1098/rsif.2019.0202`.

[54] 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 (2024), p. e2313708120. Available at `https://www.pnas.org/doi/10.1073/pnas.2313708120`.

[55] 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 (2013), p. e74335. Available at `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074335`.

[56] 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.

[57] 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 (2021), p. 20210457. Available at `https://royalsocietypublishing.org/doi/10.1098/rspa.2021.0457`.

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

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

[60] 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 (1907), pp. 693–985. Available at `http://www.jstor.org/stable/4619420`.

[61] J.M. Ver Hoef, *Who Invented the Delta Method?*, The American Statistician 66 (2012), pp. 124–127. Available at `https://doi.org/10.1080/00031305.2012.687494`.

[62] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, Springer, New York, NY, 2010 Dec.

Table 1. Fits of KM's [41] analytical SIR approximation (2) **to Bombay plague** (see Figure 1). The KM column lists the parameter values estimated by KM [41, p. 714]; the `nls` column lists the values estimated by us, using nonlinear least squares with confidence intervals obtained by the Delta method (see §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 [3, p. 408], we assume the population of Bombay was $N = 1$ million.

Estimated parameter	**symbol**	**equation**	**units**	**KM**	`nls`	**95% CI**
peak removal rate	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)
removal rate	γ	(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	(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 Figure 1). Parameter values were estimated using `fitode` to fit trajectories of Equation (1), assuming observation errors were distributed negative binomially (`nbinom`) or normally (`ols`). 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 fits. The `fitode`-fitted trajectories and confidence bands—for both `nbinom` and `ols`—are shown in the lower panel of Figure 1. As in Table 1, a population size N must be assumed to derive $\mathcal{R}_0$ estimates.

Fixed parameter	**symbol**	**units**	`nbinom`	**95% CI**	`ols`	**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)

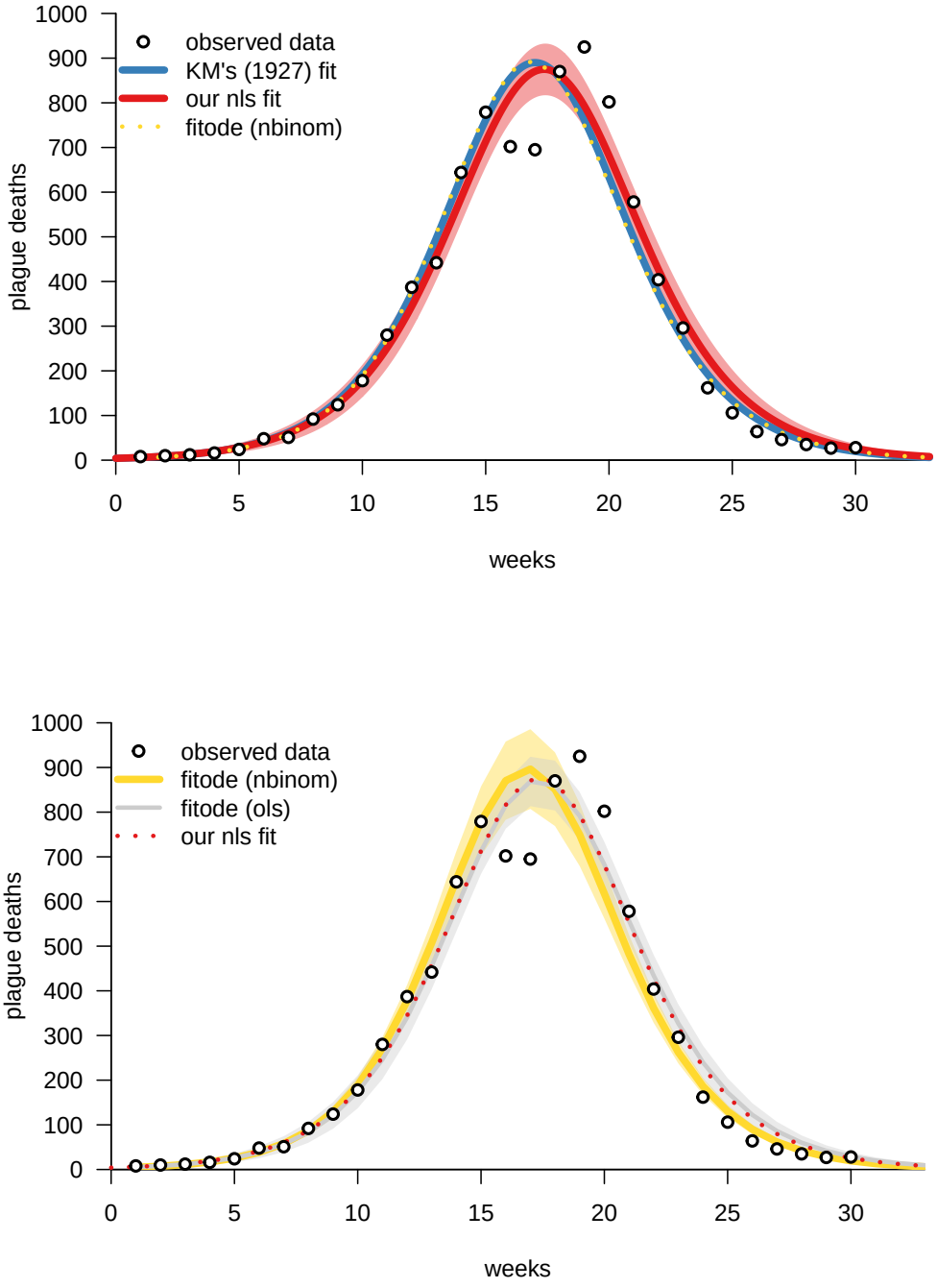

Figure 1. The plague epidemic in Bombay, 17 December 1905 to 21 July 1906, used as an example by KM [41, p. 714]. The data (dots) were digitized from [60, Table IX, p. 753]. *Top panel:* The blue and red curves show the KM approximation (2), as fitted by KM (blue) and by us using `nls` (red, with pink confidence band estimated using the Delta method; see §4). The associated parameter estimates are given in Table 1. The dotted yellow 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)]. *Bottom panel:* The solid yellow curve is identical to the dotted yellow curve in the top panel; the light yellow 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 grey 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 curve is identical to the solid red 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 [41] analytical SIR approximation (2) **to Philadelphia flu** (see Figure 2). Parameter estimates were obtained using nonlinear least squares (`nls`) to fit Equation (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 [58]. The fitted trajectory and confidence band are shown in Figure 2. See §7.

Estimated parameter	**symbol**	**equation**	**units**	`nls`	**95% CI**
peak removal rate	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)
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 Figure 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)

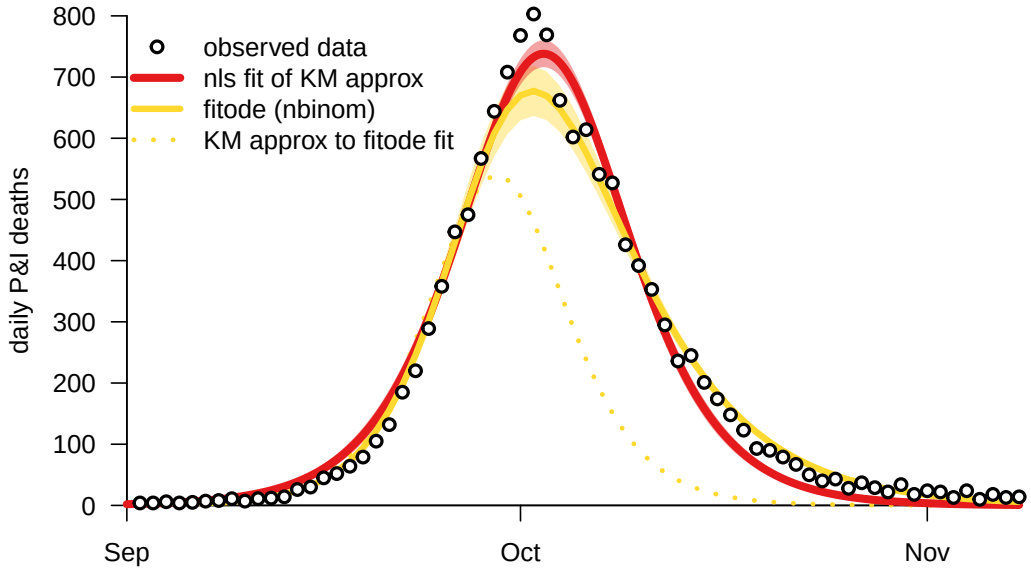

Figure 2. The main wave of the 1918 influenza epidemic in the city of Philadelphia, 1 September 1918 to 31 December 1918 [33, 56]. Reported daily deaths from pneumonia and influenza (P&I) are shown with dots. The red curve and pink confidence band show a nonlinear least squares (`nls`) fit of KM's approximation (2); the parameter estimates are given in Table 3. The solid yellow curve and light yellow confidence band show the `fitode` fit of the SIR model (1), for which the parameter estimates are given in Table 4. The dotted yellow curve shows the KM approximation using the parameters estimated with `fitode`.

Table 5. Fits of KM's [41] analytical SIR approximation (2) **to an epidemic simulated using the standard stochastic SIR model [2]** (see §8 and Figure 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 Equation (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						
peak removal rate	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)
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 [2] (see §8 and Figure 3). Parameter estimates we obtained using `fitode` to fit the SIR model (1) to the simulated data, assuming deviations from the determinsitic 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)

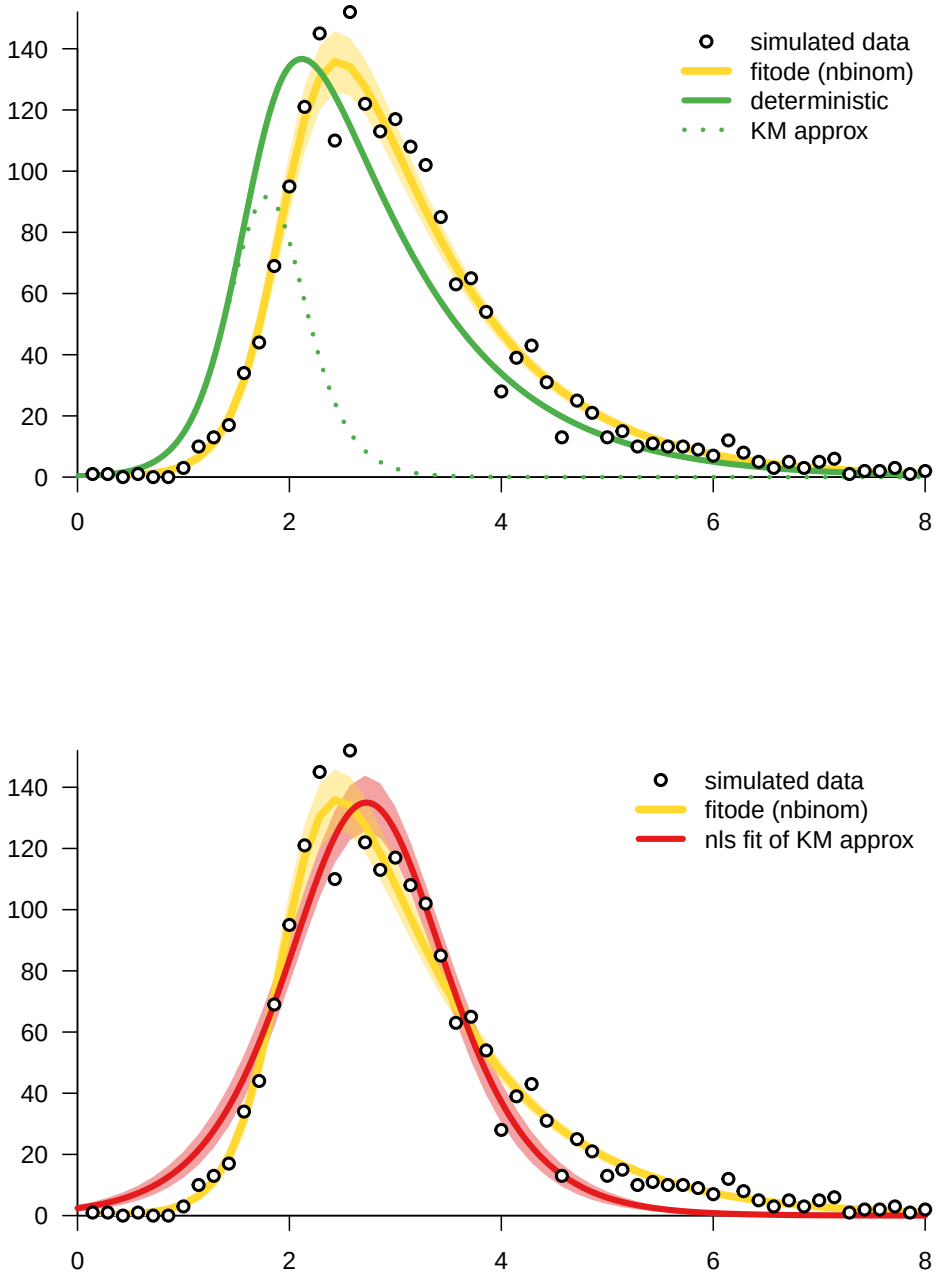

Figure 3. Deterministic fits to daily incidence generated by a stochastic SIR simulation with initial state $(S_0, I_0, R_0) = (1998, 2, 0)$, basic reproduction number $\mathcal{R}_0 = 5$, and mean generation interval $T_g = 1$ week. The simulated data points show the numbers of newly recovered individuals each day. In both panels, the yellow 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 yellow curves arises because there is a random delay until the stochastic trajectory begins to grow exponentially. *Bottom panel:* The red 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.

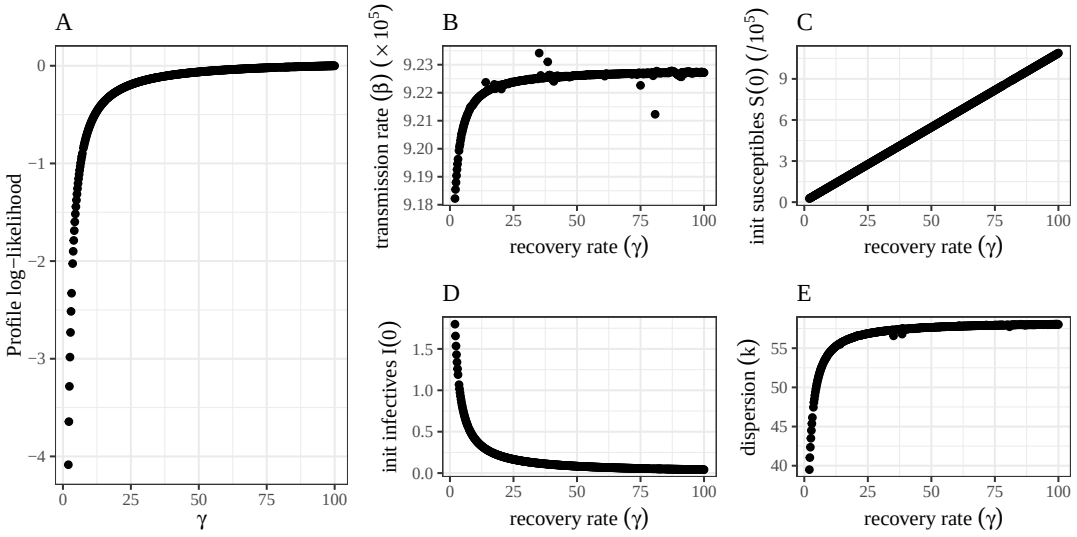

Figure 4. Unidentifiability of the mean generation interval T_g (or, equivalently, the removal rate γ) for the Bombay plague epidemic shown in Figure 1. (A) The profile likelihood—briefly discussed at the end of §4—is calculated by fixing γ to a series of given values and, for each value, maximizing the likelihood by estimating all other parameters [7]. (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.