



Cite this article: House T, Ford A, Lan S, Bilson S, Buckingham-Jeffery E, Girolami M. 2016 Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry. *J. R. Soc. Interface* **13**: 20160279. <http://dx.doi.org/10.1098/rsif.2016.0279>

Received: 8 April 2016
Accepted: 25 July 2016

Subject Category:
Life Sciences—Mathematics interface

Subject Areas:
computational biology, biomathematics

Keywords:
shedding, Markov chain Monte Carlo, compartmental model

Author for correspondence:
Thomas House
e-mail: thomas.house@manchester.ac.uk

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2016.0279> or via <http://rsif.royalsocietypublishing.org>.

Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry

Thomas House^{1,2}, Ashley Ford⁵, Shiwei Lan³, Samuel Bilson², Elizabeth Buckingham-Jeffery^{2,4} and Mark Girolami³

¹School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK

²Warwick Infectious Disease Epidemiology Research Centre (WIDER), Warwick Mathematics Institute,

³Department of Statistics, and ⁴Complexity Science Doctoral Training Centre, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

⁵School of Mathematics, University of Bristol, Bristol BS8 1TW, UK

TH, 0000-0001-5835-8062

Infectious diseases exert a large and in many contexts growing burden on human health, but violate most of the assumptions of classical epidemiological statistics and hence require a mathematically sophisticated approach. Viral shedding data are collected during human studies—either where volunteers are infected with a disease or where existing cases are recruited—in which the levels of live virus produced over time are measured. These have traditionally been difficult to analyse due to strong, complex correlations between parameters. Here, we show how a Bayesian approach to the inverse problem together with modern Markov chain Monte Carlo algorithms based on information geometry can overcome these difficulties and yield insights into the disease dynamics of two of the most prevalent human pathogens—influenza and norovirus—as well as Ebola virus disease.

1. Introduction

Infectious diseases continue to pose a major threat to human health, with transmission dynamic models playing a key role in developing scientific understanding of their spread and informing public health policy [1]. These models typically require many parameters to make accurate predictions, which interact with data in complex, nonlinear ways. It is seldom possible to perform a series of individual controlled experiments to calibrate these models, meaning that the use of multiple datasets is often necessary [2].

At the population level, infectious disease models most frequently involve a ‘compartmental’ approach in which individuals progress between different discrete disease states, usually at constant rates [3]. We note that alternatives to such a compartmental approach exist, for example, use of a deterministic time-varying infectivity, or allowing for non-Markovian state transitions in a stochastic context [4]. There are theoretical differences between these and compartmental models for predictive epidemic modelling (see [5] for additional discussion), but from the point of view of inference they pose very similar challenges and therefore the rest of this paper will consider a compartmental modelling framework.

At the individual level, there are three separate events that can be represented using different compartments: (i) an individual receiving an infectious dose of a pathogen, (ii) the individual becoming infectious and able to infect new cases, and (iii) the individual ceasing to be infectious and becoming unable to infect new cases. For some pathogens, there are also events relating to changes in disease progression during the infectious period. The times elapsed between

these events are important for understanding the epidemiology of infectious diseases as well as designing control and mitigation strategies [6,7].

Here, we consider how the rates of moving between compartments in standard epidemic models can be estimated from shedding data generated either when individuals are given an infectious dose of a virus under controlled conditions, or when existing cases are enrolled in a study, and the level of live virus that they produce (shed) is measured over time. The nature of these models and data means that there are not simple optimal values for the rate parameters, but instead the data constrain the parameters to lie close to a nonlinear curve in parameter space. We show that modern computationally intensive Bayesian methods that make use of information geometry can be used to calculate the posterior density for models of influenza, norovirus and Ebola. We then use forward modelling based on this posterior knowledge to show that some epidemiological conclusions are robust under the remaining uncertainty, but others require additional information to determine. In particular, for influenza, we show that the predicted effectiveness of quarantine-type interventions is unaffected by the remaining uncertainty, but antiviral-like interventions have a bimodal uncertainty structure. For norovirus, we show that the frequency of epidemics is predictable under the remaining uncertainty, but the severity and timing of each seasonal epidemic is not. For Ebola, we are able to distinguish between high- and low-viraemic infectious disease progression, giving results that are consistent with population-level observations of the case fatality ratio.

2. Methods

2.1. Overview

Our methodological approach involves three related components. We start by defining the different compartmental disease models that we use in §2.2. These are defined in terms of each case's *natural history*, which we represent mathematically as a continuous-time Markov chain. We also show how these models can be used to make population-level predictions by assuming, for example, that an infectious individual will cause new cases at a rate β .

Then in §2.3 we consider how the natural history parameters can be estimated from shedding data. An important distinction will be between the expected rate at which individuals infect in the population (quantified using a rate like β above) and the measured intensity of shedding (quantified using log titre). We will generally assume a simple linear relationship between these using a scaling parameter that we call τ .

Finally in §2.4, we present the Bayesian approach to uncertainty quantification as well as the algorithms necessary to implement this for the complex posterior distributions that arise in the analysis of shedding data.

2.2. Compartmental models of infectious disease

In a compartmental approach, we model the state of an individual who has been infected with a pathogen at time $t = 0$ as an integer random variable $X(t)$. We label the possible states after infection $i, j, \dots \in \{1, \dots, m\}$; while more general structures are needed for other pathogens, for both influenza and norovirus we consider the 'linear chain' case where an individual spends an exponentially distributed period of time with mean $1/\gamma_i$ in each state $i < m$ before progressing to state $i + 1$, and where m is the 'recovered' state, corresponding to the end of the infection,

from which the individual does not move. In the electronic supplementary material, we provide a general solution for the probability of being in disease compartment i at time t ,

$$\pi_i(t; \gamma) := \Pr[X(t) = i | X(0) = 1; \gamma], \quad (2.1)$$

and its derivatives with respect to the rates $\{\gamma_i\}$. For Ebola, we consider a chain that branches, which introduces parameters relating to the probability of following one path or another in addition to rate constants.

We also assume that an individual has 'infection rate' λ_i in state i (with $\lambda_m = 0$), which is proportional to the rate at which they would infect new cases in a population-level epidemic model. A key quantity is the expected infectiousness of an individual over time

$$\Lambda(t; \boldsymbol{\theta}) := \sum_{i=1}^m \lambda_i \pi_i(t; \boldsymbol{\gamma}), \quad (2.2)$$

where the model parameters are $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\lambda}) = (\theta_a)$. Perhaps the most important quantity in any epidemiological model of infections is the basic reproductive ratio, R_0 , defined as the expected number of secondary infections produced by a typical infectious individual early in the epidemic [4]. Under the simplifying (but frequently made) assumption of a homogeneous population, this quantity is given by

$$R_0(\boldsymbol{\theta}) = \int_{t=0}^{\infty} \Lambda(t; \boldsymbol{\theta}) dt. \quad (2.3)$$

Note that the constants of proportionality $\boldsymbol{\lambda}$ depend on the nature and strength of interactions in the population and therefore cannot be determined from measurements of individuals alone.

2.2.1. The SIR model

One of the simplest models in mathematical epidemiology is the SIR model, in which individuals are susceptible, infectious or removed. We use this model as a simple example of the methodology we propose. An individual infected at time $t = 0$ spends an exponentially distributed period of time in the infectious class, with rate γ , before recovering, and has infectiousness β . Therefore,

$$\pi_1(t; \gamma) = e^{-\gamma t} \quad \text{and} \quad \Lambda(t; \beta, \gamma) = \beta e^{-\gamma t}. \quad (2.4)$$

At the population level, supposing that we can ignore demographic processes such as births and deaths so the population size is fixed at N , we have a set of ordinary differential equations describing the evolution of an epidemic:

$$\frac{dS}{dt} = -\beta \frac{SI}{N}, \quad \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I \quad \text{and} \quad \frac{dR}{dt} = \gamma I. \quad (2.5)$$

Here $S(t)$ is the expected number of susceptible individuals in the population, $I(t)$ is the expected number of infectious cases and $R(t)$ is the expected number of removed individuals; we will use a similar notation below generalized in a natural way. In this work, we will consider how to fit expressions such as (2.4) to shedding data in such a way that population-level models such as (2.5) can be parametrized accounting for uncertainty.

2.2.2. Influenza

Influenza is commonly modelled using the 'SEIIR' or 'SE₂I₂R' framework, for example, in the work that was used to inform vaccination policy during the recent H1N1 pandemic [8]. Here $m = 5$ and individuals spend a 2-Erlang distributed period of time with mean $1/\omega$ in a non-infectious 'exposed' state, and then a 2-Erlang distributed period of time with mean $1/\gamma$ in the infectious state, before recovering.

To show the possible different impacts of uncertainty in parameter values, we will consider two models for delayed intervention (implemented after a time period of length d after infection) during an influenza pandemic. The first of these assumes that the intervention is ‘quarantine-like’ and completely halts transmission after administration, leading to an epidemic with reproduction number

$$R_d^{(1)} = \int_{t=0}^d \Lambda(t; \theta) dt. \quad (2.6)$$

The second, however, assumes that the individual must not have progressed from the ‘latent’ to the ‘infectious’ class and is therefore similar to some models of administration of antiviral medication like Tamiflu [9,10], leading to an epidemic with reproduction number

$$R_d^{(2)} = (1 - \Pr[X(d) < 3])R_0. \quad (2.7)$$

We note from standard results in mathematical epidemiology [4] that if the infection rate is β , and the mean infectious period is $1/\gamma$, then the basic reproductive ratio is $R_0 = \beta/\gamma$, and for an epidemic with reproduction number R the epidemic final size Z is given by the solution to the transcendental equation

$$Z = 1 - e^{-RZ}, \quad (2.8)$$

which we solve numerically for the two different reproduction numbers $R_d^{(1)}$ and $R_d^{(2)}$ above, a range of delays, and fitted values of the parameters (ω, γ) at a fixed value of $R_0 = 1.4$ (as in [8]) to make a direct comparison, although it would be straightforward to place a distribution on R_0 . We note that while all parameter values agree on the value of $R_0 = \lim_{d \rightarrow \infty} R_d$, and that $\lim_{d \rightarrow 0} R_d = 0$, at finite non-zero d the uncertainty in parameters will lead to posterior variability in R_d and hence Z .

Temporal features of an influenza epidemic are often more relevant for policy than the final size [11,12] and are typically considered using systems of differential equations. As we are considering an intervention with fixed delay, we couple the standard ODE system [8] to a set of terms modelling the delayed intervention leading to the delay-differential equation system

$$\left. \begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N} S(t)I(t), \\ \frac{dE_1}{dt} &= \frac{\beta}{N} S(t)I(t) - 2\omega E_1(t) - \pi_1(d; \omega, \gamma) \frac{\beta}{N} S(t-d)I(t-d), \\ \frac{dE_2}{dt} &= 2\omega(E_1(t) - E_2(t)) - \pi_2(d; \omega, \gamma) \frac{\beta}{N} S(t-d)I(t-d), \\ \frac{dI_1}{dt} &= 2(\omega E_2(t) - \gamma I_1(t)) - \epsilon \pi_3(d; \omega, \gamma) \frac{\beta}{N} S(t-d)I(t-d) \\ \text{and } \frac{dI_2}{dt} &= 2\gamma(I_1(t) - I_2(t)) - \epsilon \pi_4(d; \omega, \gamma) \frac{\beta}{N} S(t-d)I(t-d), \end{aligned} \right\} \quad (2.9)$$

where the variable ϵ is 1 if the intervention works for any infected individual and 0 if it only works during the latent period, and d is the intervention delay. $\pi_i(t; \omega, \gamma)$ is the probability of being in state i time t after starting in state 1 at time 0 as in (2.1) above and is also determined by the parameters in the shedding model.

2.2.3. Norovirus

Norovirus is usually assumed to follow the ‘SEIRS’ framework [13], where after infection an individual spends an exponentially distributed period of time with mean $1/\omega$ in a ‘latent’ class, then an exponentially distributed period of time with mean $1/\gamma$ in the infectious class. In contrast to influenza, individuals move from the ‘recovered’ class back to the ‘susceptible’ class with rate μ ; this loss of immunity is a relatively slow process that does not impact on the analysis of shedding data. It does, however,

influence the population-level disease dynamics of norovirus, which can be described using a set of ordinary differential equations, where $S(t)$ stands for the expected number of people who are susceptible and similarly for other compartments:

$$\left. \begin{aligned} \frac{dS}{dt} &= \delta N + \mu R - \beta(t) \frac{SI}{N} - \delta S, & \frac{dE}{dt} &= \beta(t) \frac{SI}{N} - \omega E - \delta E, \\ \frac{dI}{dt} &= \omega E - \gamma I - \delta I & \text{and } \frac{dR}{dt} &= \gamma I - \mu R - \delta R. \end{aligned} \right\} \quad (2.10)$$

Here we have assumed a constant effective population size N and have a time-varying infection rate (which is necessary to reproduce the regular seasonality seen in real data [13]) that we assume takes a sinusoidal form $\beta(t) = \beta_0(1 + A \sin(\alpha t))$. Because such external forcing in transmission is typically believed to arise from school terms [14], we take $A = 1/3$ to be close to existing empirical estimates of the impact of school closures on disease spreading [15,16], and α can be set to $2\pi \text{yr}^{-1}$. The demographic rate δ is standardly set to $1/70 \text{yr}^{-1}$. From the results of [17], we have that $R_0 = \beta_0/\gamma$; we vary this and the rate of waning immunity μ within ranges suggested by Simmons *et al.* [13], and then run the model (2.10) to determine its long-term behaviour for different fitted values of the parameters (ω, γ) .

A norovirus vaccine is likely to be available in the future [18], and we model the impact of a vaccination policy starting at time u and with effective coverage v (defined as the product of coverage and efficacy) by modification of the demographic term for S and addition of a vaccinated V compartment:

$$\left. \begin{aligned} \frac{dS}{dt} &= \delta(1 - v\Theta(t-u))N + \mu R - \beta(t) \frac{SI}{N} - \delta S \\ \text{and } \frac{dV}{dt} &= \delta v\Theta(t-u)N - V. \end{aligned} \right\} \quad (2.11)$$

Here Θ is the step function, leaving us with a set of time-inhomogeneous ordinary differential equations.

2.2.4. Ebola

Ebola is both much less common than influenza and norovirus and much more dangerous. This means that the modelling framework for it is less established—although it is typically assumed to follow an SEIR-type framework [19,20]—and also that challenge studies cannot be performed. Instead, existing cases are recruited and their viral loads are monitored. Our modelling approach is based on the results of such studies [21]. This is shown diagrammatically in figure 1 and is described by equations

$$\left. \begin{aligned} \frac{dJ_1}{dt} &= -\theta_1 J_1, & \frac{dJ_2}{dt} &= \theta_1 J_1 - \theta_1 J_2, \\ \frac{dI_1}{dt} &= \theta_2 \theta_1 J_2 - \theta_3 I_1, & \frac{dI_2}{dt} &= \theta_3 I_1 - \theta_3 I_2, \\ \frac{dH_1}{dt} &= (1 - \theta_2) \theta_1 J_2 - \theta_4 H_1, & \frac{dH_2}{dt} &= \theta_4 H_1 - \theta_4 H_2, \\ \frac{dR}{dt} &= \theta_3 I_2 & \text{and } \frac{dD}{dt} &= \theta_4 H_2. \end{aligned} \right\} \quad (2.12)$$

In this study design, the latent E states are not observed and so the initial condition is $J_1(0) = 1$ with all other quantities initially 0. Parameter interpretations and priors are given in table 1.

The transmission rates for low- and high-viraemic pathways are, respectively,

$$\left. \begin{aligned} A_l &= \beta(\theta_5(J_1 + J_2) + \theta_6(I_1 + I_2)) \\ \text{and } A_h &= \beta(\theta_5(J_1 + J_2) + \theta_7(H_1 + H_2)). \end{aligned} \right\} \quad (2.13)$$

2.3. Shedding model and data

Our aim is to extract parameter estimates for compartmental epidemic models from challenge studies in which human volunteers are infected with a pathogen, or observational studies based on

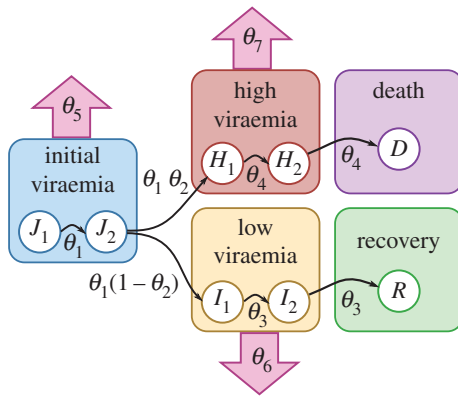


Figure 1. Ebola model compartmental structure and role of parameters. Compartments are shown as circles; flows between compartments are shown as thin black arrows labelled with parameters; and infectiousness is indicated by outward facing block arrows labelled with parameters.

existing cases, and the level of live virus they produce (or ‘shed’) is measured over time. The concentration of live virus is quantified as a ‘titre’, which is essentially an estimate of the concentration of live virus. The relationship between this quantity and transmissibility is complex, but generally agreed to be sub-linear [22,23]. We assume here that at each time point t the log titre y is proportional to the expected intensity of infection $\Lambda(t)$ plus additive Gaussian noise representing experimental error and other factors such as individual-level variability, leading to likelihood functions based on products of normal distribution probability mass functions. The details are, however, different for the four scenarios we consider and so we define our models separately below.

2.3.1. The SIR model

For the SIR model, we assume one observation y of shedding with standard deviation σ at time t leading to likelihood function

$$L(y|\tau, \gamma) = \mathcal{N}(y|\tau e^{-\gamma}; \sigma). \quad (2.14)$$

We will use this as a toy model to demonstrate our methodological approach, using numerical values $y = 1$, $\sigma = 0.02$ and $t = 1$. Note that here and throughout, we write the probability density function for the normal distribution with mean μ and standard deviation σ evaluated at x as

$$\mathcal{N}(x|\mu; \sigma) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (2.15)$$

2.3.2. Influenza

For influenza, we use the meta-analysis data of Carrat *et al.* [24] for viral titre in the nasal passages of individuals infected with influenza A H1N1, which is shown in figure 2a. Here, observations are made at regularly spaced times belonging to a set \mathcal{T} . Under our general modelling assumptions, the likelihood of observing a set of mean log-titres (y_t) among participants, with associated standard deviations (σ_t), given the model parameters, is

$$L(\mathbf{y}, \boldsymbol{\sigma}|\boldsymbol{\theta}) = \prod_{t \in \mathcal{T}} \mathcal{N}(y_t; (\tau/\beta)\Lambda(t|\boldsymbol{\theta}), \sigma_t), \quad (2.16)$$

where \mathcal{N} is the normal distribution probability density function intended to capture various sources of experimental error and individual-level variability as would be expected due to the central limit theorem. Here, we assume that the variance at each time point is measured, as given in [24]. Because the infection rate β depends *inter alia* on the rate of contact between individuals, which cannot be estimated from shedding data, we rescale the

Table 1. Parameters of the Ebola model, their interpretation and prior distribution.

parameter	interpretation	prior
θ_1	2/(mean time in initial viraemic state)	Exp(0.01)
θ_2	proportion entering high-viraemic state	Uniform([0,1])
θ_3	2/(mean time in low-viraemic state)	Exp(0.01)
θ_4	2/(mean time in high-viraemic state)	Exp(0.01)
θ_5	scaling parameter for initial viraemic state	Exp(0.1)
θ_6	scaling parameter for low-viraemic state	Exp(0.1)
θ_7	scaling parameter for high-viraemic state	Exp(0.1)

infectiousness Λ using the scaling parameter τ meaning that our parameters for estimation are $\boldsymbol{\theta} = (\tau, \omega, \gamma)$.

2.3.3. Norovirus

For norovirus, we use data from the study of Atmar *et al.* [25], where individuals were infected under controlled conditions and observations of viral titre in faeces made irregularly at times belonging to a set \mathcal{T} . These data are shown in figure 3a and since they do not aggregate in the same way as the influenza data the likelihood function is

$$L(\mathbf{y}|\boldsymbol{\theta}, \sigma) = \prod_{t \in \mathcal{T}} \mathcal{N}\left(y_t; \left(\frac{\tau}{\beta}\right)\Lambda(t|\boldsymbol{\theta}), \sigma\right), \quad (2.17)$$

so σ is here an additional parameter that must be estimated, and we have rescaled the infectiousness as before using τ . This makes the norovirus parameter space dimension 4, in contrast to dimension 3 for influenza, with $\boldsymbol{\theta} = (\tau, \omega, \gamma, \sigma)$.

2.3.4. Ebola

For Ebola, we use the data of Ksiazek *et al.* [21] on viral titre in the blood of hospitalized Ebola cases, which are stratified into low- and high-viraemic disease pathways as shown in figure 1, together with the model described by equations (2.12). This leads to a likelihood function that takes the form of a product of low and high trajectories, each of which is similar to the influenza likelihood:

$$L(\mathbf{y}, \boldsymbol{\sigma}|\boldsymbol{\theta}) = \prod_{t \in \mathcal{T}_l} \mathcal{N}\left(y_{l,t}; \left(\frac{1}{\beta}\right)\Lambda_l(t|\boldsymbol{\theta}), \sigma_{l,t}\right) \times \prod_{t \in \mathcal{T}_h} \mathcal{N}\left(y_{h,t}; \left(\frac{1}{\beta}\right)\Lambda_h(t|\boldsymbol{\theta}), \sigma_{h,t}\right), \quad (2.18)$$

where we use a natural subscripting of ‘l’ for low and ‘h’ for high, and the parameters for estimation are $(\theta_1, \dots, \theta_7)$ as shown in table 1.

2.4. Statistical framework

2.4.1. The Bayesian approach to identifiability

It is long-established that fitting of a sum of exponentials to data is potentially troublesome; in particular, Acton [26]

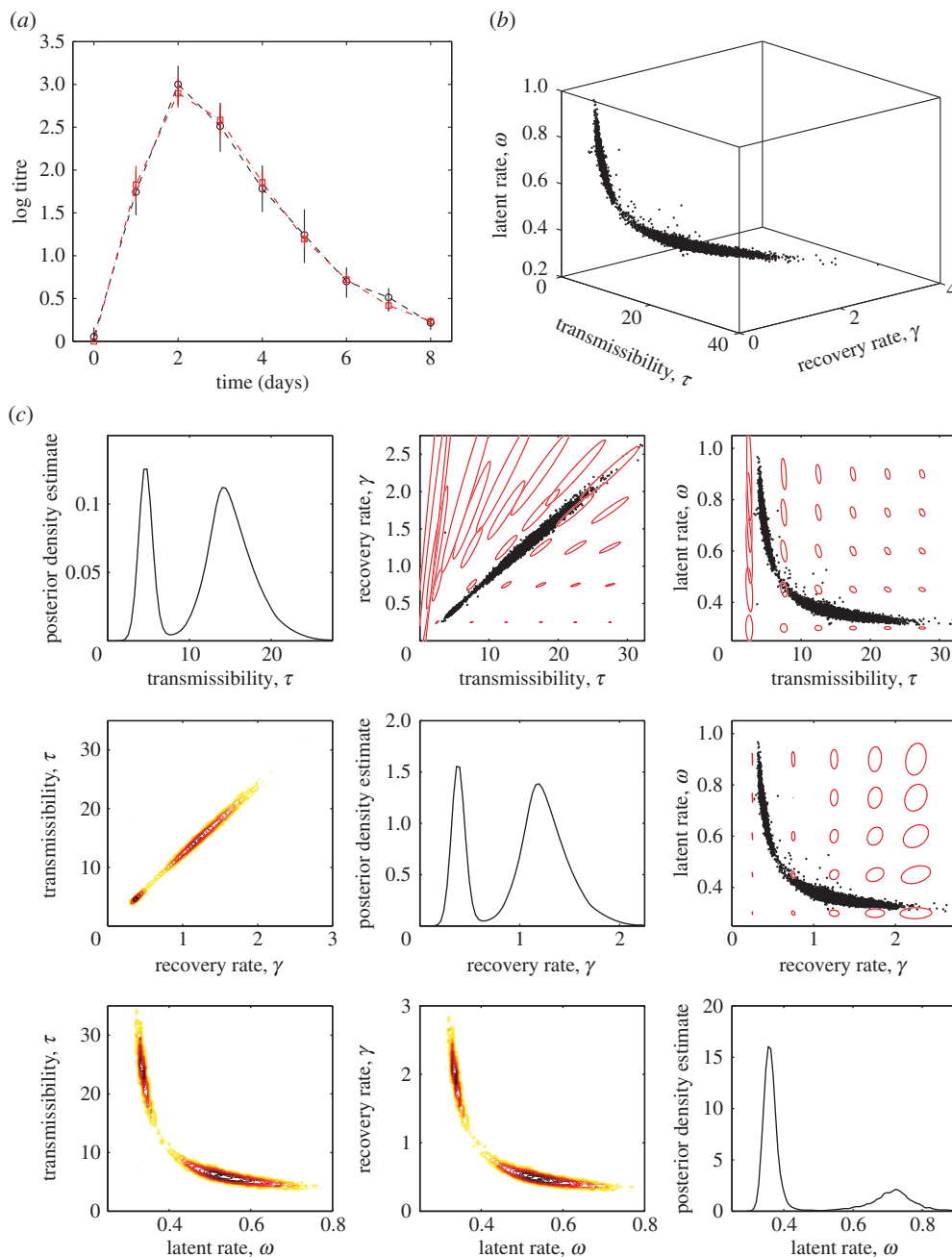


Figure 2. Analysis of influenza shedding data. (a) Data and empirical confidence intervals (black) against model mean and 95% credible interval (red). (b) Posterior samples obtained using WHLMC. (c) Diagonal plots: marginal posterior density estimates for each parameter. Bottom left plots: marginal log posterior for pairs of parameters—first-order gradients will be perpendicular to contours. Top right plots: samples (black) with visualizations of the metric—based on expectations of second-order derivatives as detailed in the main text—as red curves, centred on evenly spaced points, whose radius is inversely proportional to the infinitesimal distance at that angle as defined in (2.28).

considers fitting the model $y = Ae^{-at} + Be^{-bt}$ to (y, t) pairs and notes that ‘there are many combinations of (a, b, A, B) that will fit most exact data quite well indeed [...] and when experimental noise is thrown into the pot, the entire operation becomes hopeless’.

Our compartmental models are more complex than sums of exponentials, but exhibit the same lack of a clear mode in the likelihood function. While there are various methods to address this issue in other applications (e.g. [27]), another response is (informally speaking) to consider all parameter combinations that fit well, and to investigate the epidemiological consequences of this uncertainty in parameter values.

More formally, we work in a Bayesian framework, meaning that we attempt to calculate the posterior density p over parameters θ from the likelihood function L and the prior function

f using Bayes’ rule

$$p(\theta|\mathbf{y}) = \frac{L(\mathbf{y}|\theta)f(\theta)}{\int L(\mathbf{y}|\vartheta)f(\vartheta)d\vartheta}. \quad (2.19)$$

Given fixed data \mathbf{y} , the measure $p(\theta|\mathbf{y})d\theta$ is higher in more credible regions of parameter space, and can be multi-modal and/or with many combinations of parameters having the same level of posterior support.

Here, we attempt to use priors that are broadly speaking uninformative—either uniform or improper if there is sufficient data, or low-rate exponential if there is less data. It is important to note, however, that use of strongly informative priors is another method for restoration of identifiability, in the sense of an approximately multivariate normal posterior distribution that is concentrated in the region of a unique mode.

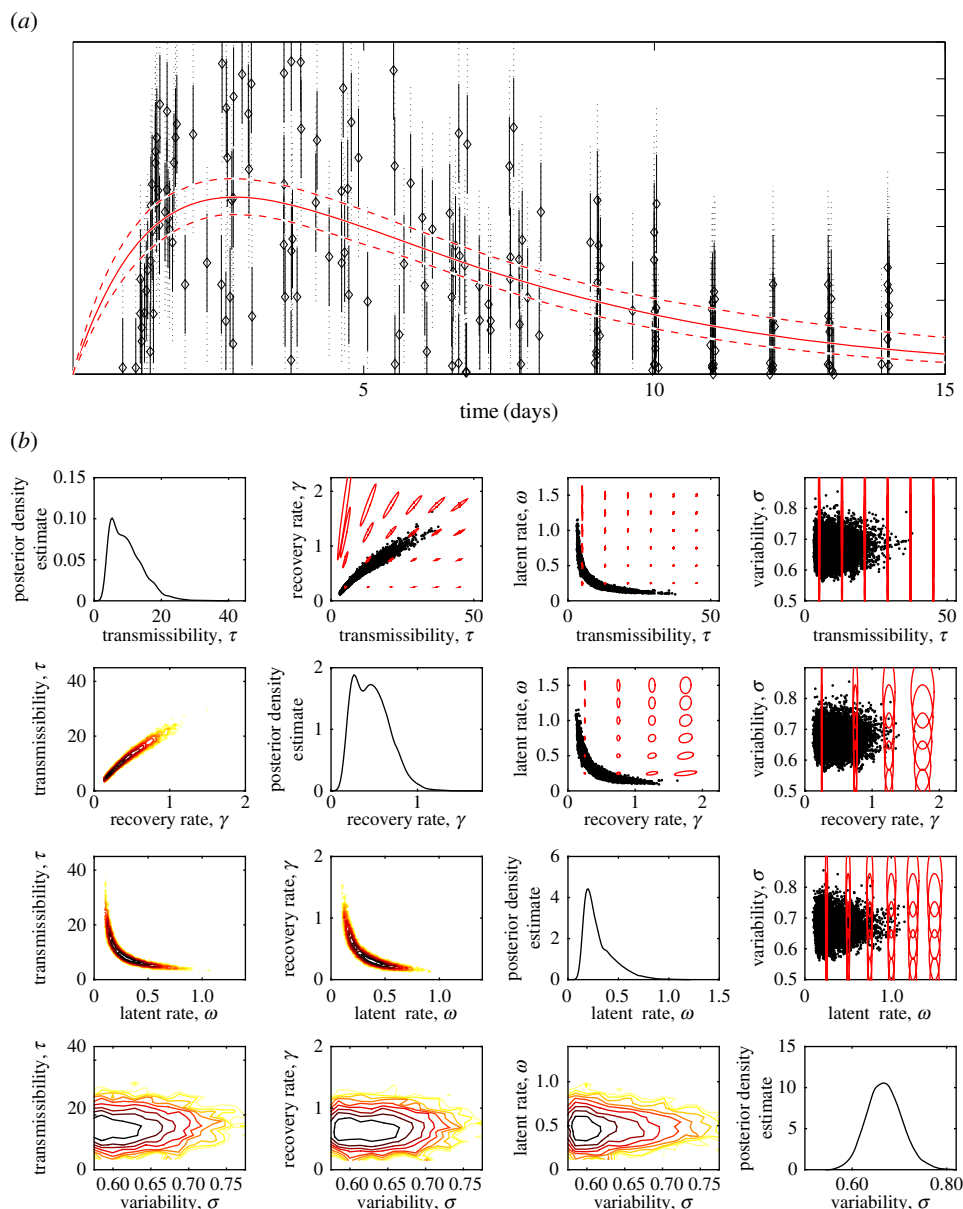


Figure 3. Analysis of norovirus shedding data. (a) Data (black circles) with 67% (solid black line) and 95% (dotted black line) against model mean and 95% credible interval (red). (b) Diagonal plots: marginal posterior density estimates for each parameter. Bottom left plots: marginal log posterior for pairs of parameters—first-order gradients will be perpendicular to contours. Top right plots: samples (black) with visualizations of the metric—based on the empirical Fisher information as detailed in the main text—as red curves, centred on evenly spaced points, whose radius is inversely proportional to the infinitesimal distance at that angle as defined in (2.28).

2.4.2. Markov chain Monte Carlo

Typically, the integral in the denominator of (2.19) is not tractable so we adopt the popular methodology of defining a Markov chain on parameter space whose stationary distribution has probability density function p , i.e. Markov chain Monte Carlo (MCMC) [28,29], in particular the Metropolis–Hastings algorithm [30,31]. This is a discrete-time Markov chain in which a change of state from θ^n to θ^* is proposed with probability $q(\theta^*|\theta^n)$ and the proposal is accepted with probability

$$\alpha = 1 \wedge \frac{L(\mathbf{y}|\theta^*)f(\theta^*)q(\theta^n|\theta^*)}{L(\mathbf{y}|\theta^n)f(\theta^n)q(\theta^*|\theta^n)}. \quad (2.20)$$

We shall now outline five popular approaches to MCMC, three that do not make use of derivatives and two that do, with all being in some sense a special case of (2.20).

2.4.3. Derivative-free Markov chain Monte Carlo algorithms

(i) *Independence sampling.* In independence sampling, there is no dependence on the current state for the proposal distribution.

A popular choice is simply to sample from the prior so that

$$q(\theta'|\theta) = f(\theta'). \quad (2.21)$$

Intuitively, such an approach is expected to work well when the posterior is ‘close’ to the prior.

(ii) *Random walk.* In a random walk approach, the current state of the Markov chain is only used to inform the mean of the proposal distribution. The most popular choice is the multivariate normal

$$q(\theta'|\theta) = \mathcal{N}(\theta'|\theta, \Sigma), \quad (2.22)$$

where the constant matrix Σ is often adaptively tuned to optimize algorithmic performance [32].

(iii) *Gibbs.* If it is possible to sample from the marginal posterior for a parameter θ_a then we can propose with density

$$q(\theta'|\theta) = p(\theta'_a|\theta_1, \dots, \theta_{a-1}, \theta_{a+1}, \dots, \theta_n, \mathbf{y}). \quad (2.23)$$

From (2.19) and (2.20), we then see that the acceptance probability for such a proposal is 1. If the marginal posteriors for

all parameters are known, then pure Gibbs sampling can be undertaken and involves cycling through proposals (2.23) for all a .

2.4.4. Derivative-based Markov chain Monte Carlo

In the field of numerical optimization, methods such as gradient descent that make use of the first-order derivatives of the function to be optimized are popular. Significant care must be taken when extending these to stochastic algorithms such as MCMC, but there are two popular methods that make use of the first-order derivatives of $\mathcal{L} := \ln(Lf)$. We use the following notation:

$$\partial_a \mathcal{L} := \frac{\partial \mathcal{L}}{\partial \theta^a} \quad \text{and} \quad \partial \mathcal{L} = (\partial_a \mathcal{L}); \quad (2.24)$$

note that throughout we write (x_i) for the vector \mathbf{x} with i th element x_i and (M_{ij}) for the matrix \mathbf{M} with (i, j) th element M_{ij} . There are then two main families of derivative-based MCMC algorithms that we consider.

(i) *MALA*. The first algorithm family starts with the Langevin equation

$$d\boldsymbol{\theta} = \frac{1}{2} \partial \mathcal{L} dt + d\mathbf{W}. \quad (2.25)$$

This stochastic differential equation model has a stationary distribution equal to the posterior distribution as defined in (2.19), and the Metropolis-adjusted Langevin algorithm (MALA) uses the Euler approximation to (2.25):

$$\Delta \boldsymbol{\theta} \approx \frac{1}{2} \partial \mathcal{L} \boldsymbol{\epsilon} + \boldsymbol{\epsilon}^{1/2} \mathbf{U} \sim \mathcal{N}\left(\left(\frac{\boldsymbol{\epsilon}}{2}\right) \partial \mathcal{L}, \boldsymbol{\epsilon} \mathbf{I}\right), \quad (2.26)$$

where \mathbf{U} is a vector of independent standard normal random variables and \mathbf{I} is the identity matrix. The approximation (2.26) can then be used as a proposal within the Metropolis–Hastings algorithm [33].

(ii) *HMC*. The second family of algorithms starts from the following system of ordinary differential equations that are a special case of Hamiltonian dynamics:

$$\frac{d\boldsymbol{\theta}}{dt} = \mathbf{v} \quad \text{and} \quad \frac{d\mathbf{v}}{dt} = \partial \mathcal{L}. \quad (2.27)$$

The randomness in the proposals arises as a result of the starting value of a vector of auxiliary variables \mathbf{v} , by default chosen as a vector of standard normal random variables. MCMC algorithms based on Hamilton's equations (2.27) are called hybrid [34] or Hamiltonian [35] Monte Carlo (HMC).

One important thing to note about these algorithms is that they use likelihood derivatives to improve acceptance rates, but since they include a Metropolis–Hastings step the derivative calculations could be approximate.

2.4.5. Geometric concepts in Markov chain Monte Carlo

While both MALA and HMC remove some of the inefficiencies of random walks through use of local gradients, they are not particularly efficient for the curved 'boomerang'-shaped posteriors that we see for the shedding models and data defined above. To address this issue, recent work has made progress through use of concepts from differential geometry [36].

In general, we have n real-valued parameters, where the support for the posterior distribution is a set $P \subseteq \mathbb{R}^n$. A general vector of parameter values is $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in P$, and informally speaking a metric is defined as a smooth symmetric map $s: P \times P \rightarrow [0, \infty)$ such that $s(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is the 'distance' between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. In practice, we will only need to define this over small local distances, which requires a metric matrix $\mathbf{G}(\boldsymbol{\theta}) = (G_{ab})$; explicitly, the infinitesimal distance between $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + d\boldsymbol{\theta}$ is

$$ds(\boldsymbol{\theta}, d\boldsymbol{\theta}) = \left(\sum_{a,b} G_{ab} d\theta_a d\theta_b \right)^{1/2}. \quad (2.28)$$

Here \mathbf{G} can be any smooth matrix function and conceptually speaking this means that even the most complex posterior can be efficiently sampled with a choice of metric that brings all high-density regions of parameter space sufficiently 'close' to each other.

Throughout this work, we will visualize the impact of the metric in the plane for two parameters of a general model using ellipses, which are defined as follows. Let $\mathbf{G}(\theta_1, \dots, \theta_n)$ be the metric matrix and consider the first two parameters (θ_1, θ_2) without loss of generality. Now let

$$(\mathbf{G}(\theta_1, \theta_2, \hat{\theta}_3, \dots, \hat{\theta}_n))_{ab} = (G_{ab}), \quad a, b \in \{1, 2\}, \quad (2.29)$$

where $\hat{\theta}_i$ is a point estimate (we choose the posterior median) for the parameter θ_i . Consider the ellipse defined by the following equation for polar coordinates distance r (from (θ_1, θ_2) in the plane) and angle α :

$$r(\alpha) = (G_{11} \cos^2(\alpha) + G_{12} \cos(\alpha) \sin(\alpha) + G_{21} \sin(\alpha) \cos(\alpha) + G_{22} \sin^2(\alpha))^{-1/2}. \quad (2.30)$$

Plotting several such ellipses in the plane allows us to visualize the impact of the metric in the following sense: points on each ellipse are all the same 'distance' from the centre as each other under the assumption of a locally constant metric.

While it is not simple to optimize the metric for a particular model, a generally well-motivated choice is the Fisher–Rao metric as suggested by Girolami & Calderhead [36]

$$G_{ab}(\boldsymbol{\theta}) = \mathbb{E}[\partial_a \mathcal{L} \partial_b \mathcal{L}], \quad (2.31)$$

where the expectation is taken over data. Benefits of this metric include that it ensures the matrix \mathbf{G} will be positive definite, and hence that the inverse matrix \mathbf{G}^{-1} will exist and be positive definite. Calculations of the Fisher–Rao metric for the models under consideration are given in the electronic supplementary material, showing that it is also available in a closed form for our models. We note that other metrics are sometimes preferable, as discussed by, for example, Betancourt [37]; however in our case the Fisher–Rao metric proved to be adequate.

We used two different geometric algorithms, chosen based on features of the posterior.

(i) *SMMALA*. The simplified manifold Metropolis-adjusted Langevin algorithm (SMMALA) was introduced in [36] and shown to be competitive in terms of computational effort in several applied contexts by Calderhead and co-workers [38,39]. In this approach, the proposal distribution is

$$q(\boldsymbol{\theta}' | \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{\theta}'; \boldsymbol{\theta} + \left(\frac{\boldsymbol{\epsilon}}{2}\right) \mathbf{G}^{-1} \partial \mathcal{L}, \boldsymbol{\epsilon} \mathbf{G}^{-1}\right), \quad (2.32)$$

with standard MALA recovered if we set $\mathbf{G} = \mathbf{I}$. Note that the inverse of the metric matrix is used to ensure that the expected distance (as defined in (2.28)) of a move is directionally invariant. We used SMMALA to sample from the norovirus and Ebola posteriors, which each had one mode but were strongly correlated with variable local correlation structure.

(ii) *WHLMC*. The idea behind wormhole Lagrangian Monte Carlo (WHLMC) is that for a multi-modal posterior, a metric can be defined that dramatically reduces the distance between modes, and a modified form of the dynamics (2.27) can exploit this proximity. Full details of the algorithm are highly technical and are given in the papers that first introduced it [40,41], as well as in our electronic supplementary material.

3. Results

3.1. Selection of Markov chain Monte Carlo algorithm

Taking the simple likelihood function (2.14) together with prior distribution uniform ($[0, 5] \times [0, 3]$) we were able to

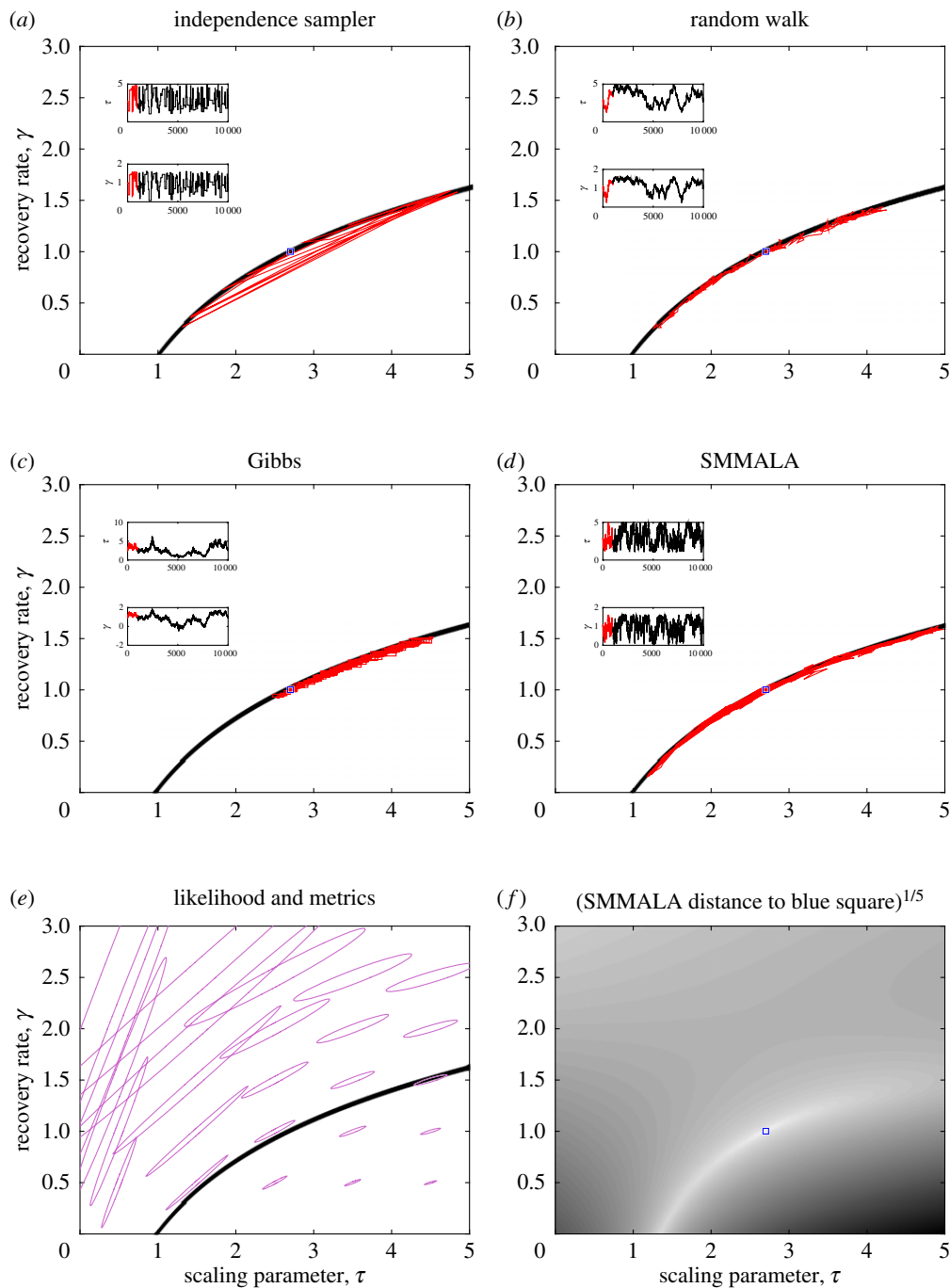


Figure 4. Results for the SIR model with recovery rate γ and scaling parameter τ (the constant of proportionality for the observed level of shedding at a given expected infectiousness). (a–d) MCMC trajectories for 1000 likelihood evaluations starting from the blue square are shown as red lines, with trace plots for a further 9000 steps shown in insets for different algorithms. (e) Visualizes the impact of the metric through the use of ellipses as detailed in §2.4.5. (f) Visualizes the impact of the metric through shading with shading intensity dependent on the distance to the starting point of the MCMC run.

run the full set of MCMC algorithms discussed above to assess their efficiency. Figure 4 shows the results of running these algorithms.

In terms of the derivative-free algorithms, the low-dimensional nature of the problem means that the independence sampler does relatively well. Both of the random walk and Gibbs samplers are not able to move efficiently through the region of high posterior density due to its narrow, curved structure. By contrast, SMMALA is able to adjust to variations in local posterior structure and as such provides a series of samples from the posterior that are much more independent of each other than other approaches. This is explained by figure 4e,f that visualize the impact of the geometry in this algorithm.

Figures 2, 3 and 5 also show that the Fisher–Rao metric and associated geometry generally correctly resolves the difficulties associated with our highly correlated posterior distributions for influenza, norovirus and Ebola, allowing accurate quantification of uncertainty in epidemiological rates. The question then becomes under what circumstances the additional computational effort of implementation of these algorithms is warranted, which has no simple answer here as in other areas of computational statistics; however, we note the following points.

First, standard measures of performance such as minimum effective sample size per CPU second [36] often overstate the effectiveness of inaccurate algorithms for our models. Figure 4b shows that over iterations 2–3000,

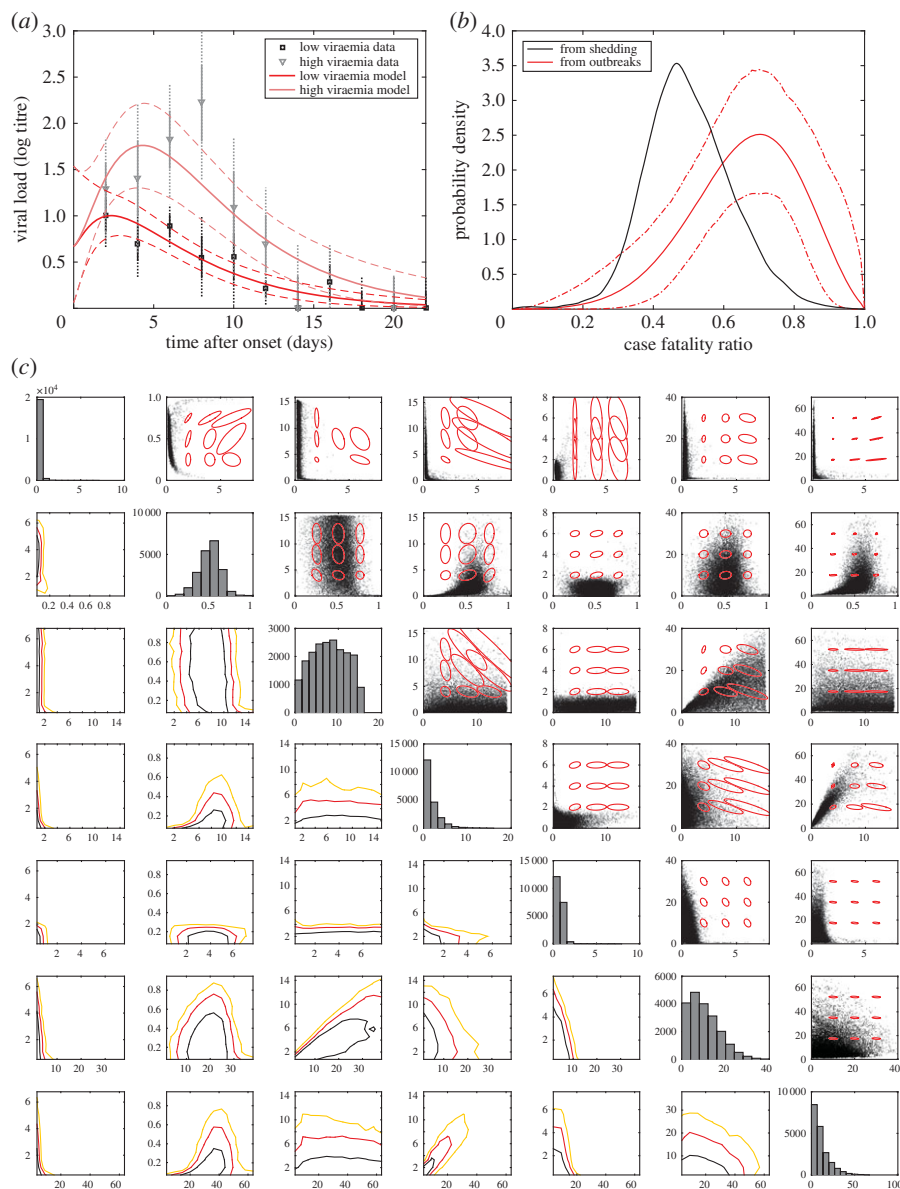


Figure 5. Ebola results. (a) Point estimate and posterior predictive interval for shedding over time against data for different viraemia levels. (b) Case fatality ratio from shedding data against that obtained for outbreaks [42]. (c) Posterior as for influenza and norovirus, with the a th row/column corresponding to parameter θ_a .

random walk sampling appears to be well-behaved and would yield a high ESS despite only being in a small sub-area of the region of parameter space from which we would like to sample.

Secondly, if the posterior density is concentrated in a nonlinear region, derivative-free methods such as Gibbs, independence sampling and random walk will have a general tendency to get ‘stuck’ in sub-areas. This will not be problematic if there are sufficient computational resources available to perform significant thinning—i.e. removal of MCMC samples to reduce correlations between those that remain.

Thirdly, computational resources for these algorithms will almost certainly become overstretched in any of the following three limits: (i) As the dimensionality of parameter space becomes larger, for example, in our Ebola model. (ii) In the presence of multi-modality as in our influenza model. (iii) For extreme cases of unidentifiability, for example, the $\sigma \rightarrow 0$ limit of our SIR model.

Finally, the geometric methods for MCMC that we present and employ here are designed to be particularly well suited to complex nonlinear relationships between

parameters where the derivatives of the log-likelihood and log-prior are available in an analytically closed form, which is the case for our models. Despite this we note that there are many other sophisticated approaches to computationally intensive Bayesian inference [29] that could be of use due to their generality.

3.2. Influenza parameters and antiviral treatment

Figure 2 shows the results for our influenza model given rate-0.1 exponential priors on each parameter (chosen not to influence the posterior significantly but to ensure that the small number of data points does not become problematic). This shows that the credible ranges of individual parameters are close to typical values in the literature—Baguelin *et al.* [8], for example, consider scenarios with $\omega \in [0.5, 10]$ and $\gamma \in [0.5, 2.5]$.

More importantly, however, the bimodal and highly correlated nature of the posterior distribution means that for some models of antiviral action it is not possible to make firm predictions based on parameter values from challenge studies.

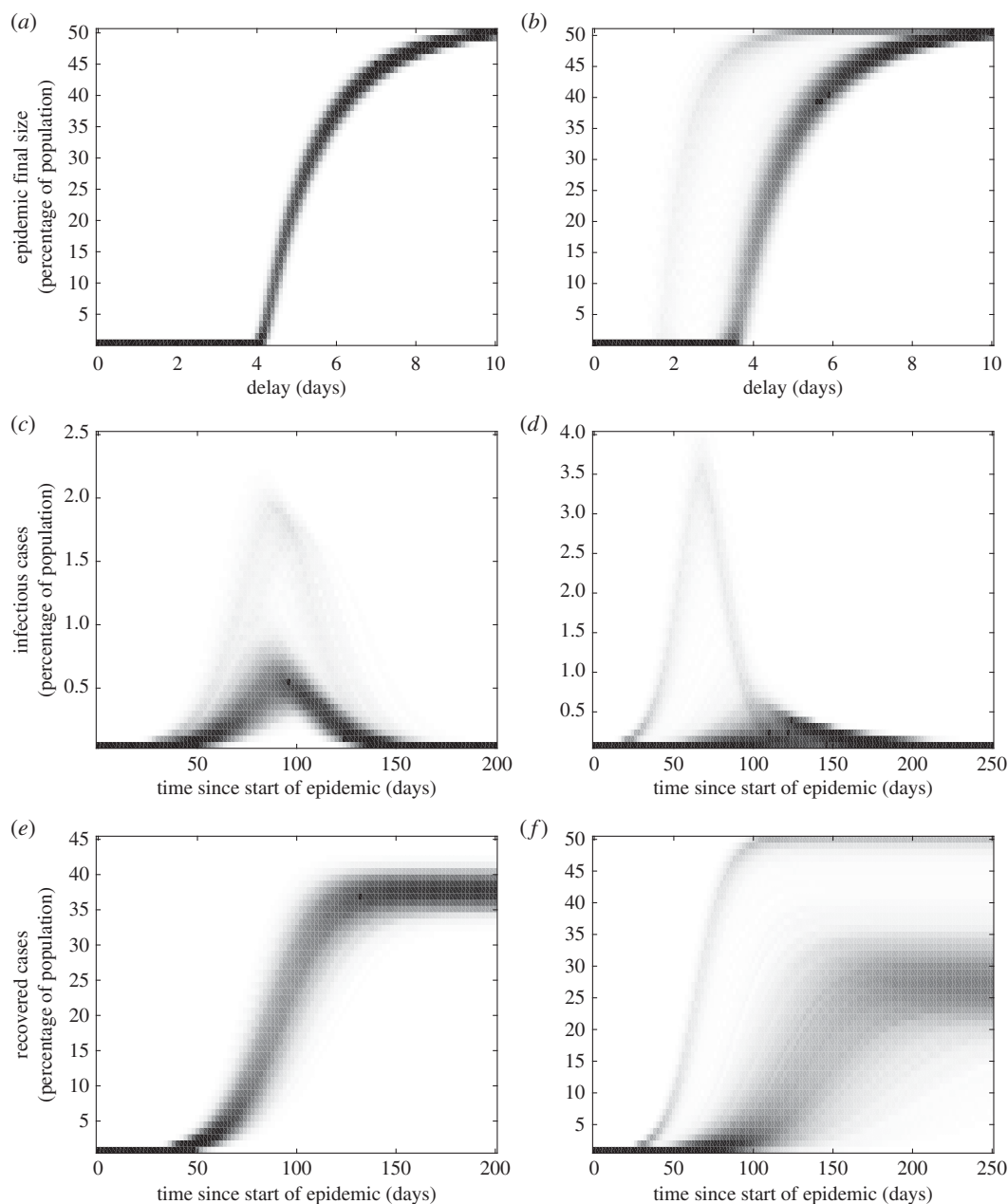


Figure 6. Interventions administered after a constant delay against influenza given parameter uncertainty. (*a,c,e*) Assuming that the intervention always stops transmission. (*b,d,f*) Assuming that the intervention only stops transmission if the individual has not yet become infectious. (*a,b*) Final sizes for a range of intervention delays. Plots of infection (*c,d*) and recovery (*e,f*) time series are shown for delay $d = 6$ (*c,e*) and $d = 4.5$ (*d,f*). Intensity of shading is proportional to probability^{0.1} to render features visible in regions with large uncertainty.

Figure 6*a,c,e* shows the results of a quarantine-like intervention that is always effective after a delay, where the relationship between delay in antiviral administration and epidemic final size at constant R_0 is predictable to within a few percentage points, although there is much greater uncertainty in the peak prevalence. Figure 6*b,d,f* shows an antiviral-like intervention that is only effective if administered during the latent period, meaning the absolute uncertainty in final size can be almost 50% and the relative uncertainty in peak prevalence can amount to a factor of four or more.

3.3. Norovirus parameters and seasonality

Figure 3 shows the results for our norovirus model; given the large amount of data we use improper priors and see that the credible ranges of individual norovirus parameters are also close to typical values in the literature; e.g. Simmons *et al.* [13] take $\omega = 1$ and $\gamma = 0.5$.

Figure 7 shows that the impact of this uncertainty (for other parameter values as given above) is mainly seen in the height (with peak prevalence differing by a factor of 3 or more) and timing within the year of seasonal epidemics. For the chaotic/irregular scenario (figure 7*b*) however, the overall epidemic dynamics are subject to significant uncertainty. Fortunately, conditioned on knowing that epidemics are regular and annual and with a particular peak, the broad impact of a vaccine policy can be predicted as shown in figure 7*e*.

3.4. Ebola parameters and case fatality ratio

Figure 5 shows the results obtained for our Ebola model. The point estimates and 95% CIs of mean infections period come out at 5.3[3.3, 7.4] days for low-viraemia cases and 6.8[5.1, 10.1] days for high-viraemia cases, which are reasonable values [20]. Figure 5*a* shows that the model produces shedding output that is consistent with the data, and figure 5*b*

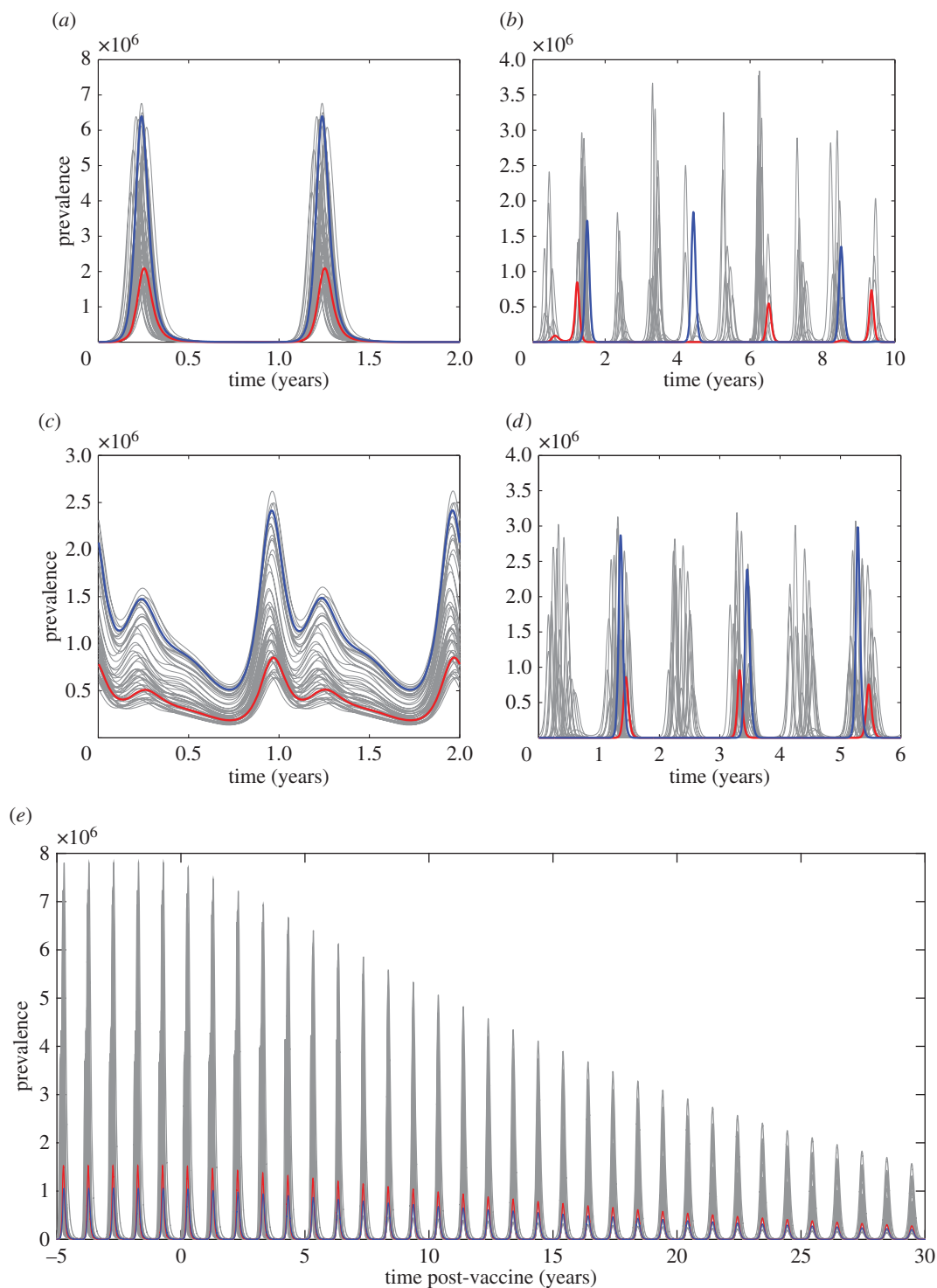


Figure 7. Implications of uncertainty in norovirus shedding for long-term temporal behaviour. Deterministic ODE trajectories are shown for 50 different posterior samples with two randomly chosen ones highlighted in red and blue. (a) $R_0 = 1.8$, $T_L = 6$ months shows regular annual oscillations. (b) $R_0 = 1.8$, $T_L = 60$ months shows irregular/chaotic behaviour. (c) $R_0 = 4$, $T_L = 6$ months shows annual oscillation but not well-defined epidemics. (d) $R_0 = 4$, $T_L = 60$ months shows bi-annual oscillation but with variable peak heights. (e) Time series for parameters as in plot (a) given vaccination at birth with 90% efficacy.

compares the posterior for the case fatality ratio obtained from shedding data to the one obtained from known outcomes of previous outbreaks [42], again suggesting that the model outcome is reasonable but that uncertainties are very large.

4. Discussion

In summary, we have shown that it is possible to use modern Bayesian MCMC methods, based on derivatives of

the log-likelihood and information geometry, to make a full uncertainty quantification of epidemiological parameters fitted to human viral shedding data. We have performed our analysis for two of the most prevalent pathogens: influenza and norovirus, as well as for Ebola, a highly virulent zoonotic disease.

Shedding data allow disease ‘natural history’ parameters to be fitted; these usually need to be combined with population-level measurements such as the basic reproductive ratio R_0 to specify policy-relevant models fully. Our results show that the epidemiological consequences of uncertainty

in natural history parameters can often be highly significant since these are important for interventions such as reducing transmission through quarantine or medication, as well as prediction of long-term disease behaviour and clinical outcomes. Natural history parameters also strongly affect other aspects of infectious disease epidemiology such as outbreak reconstruction and we would expect similarly strong effects in these contexts.

To make progress, we have had to base our analysis on simplifying assumptions that we would hope can be relaxed in future work as the field develops. One example is that the simple likelihood functions (2.18) and (2.17) assume independence that could be extended to include more general functional relationships, which would be particularly important if the methodology were extended to diseases such as human immunodeficiency virus where there are very different time scales involved in passing between compartments [43]. In such an example, one might wish to consider more general models, for example, ones in which the progression between the latent, infectious and removed classes is governed by more general distributions than those we have considered here. Provided the Laplace transformations of

the probability density functions for these distributions are available, then expressions for Λ and its derivatives with respect to the parameters can be obtained via the convolution theorem, although this can result in a computationally intensive likelihood function. Alternatively, it might be possible to approximate the derivatives since inaccuracies in any such approximation will lead to algorithmic inefficiency rather than bias.

An additional assumption we have made is that the parameters we are not fitting (for example, the basic reproductive ratio R_0) are fixed. This is particularly important to relax if multiple data sources are to be used in a principled way in infectious disease modelling for public health [2]. In particular, the measurements at the population level required to estimate R_0 are likely to carry their own uncertainty, which can be combined with our uncertainty quantification for disease natural history parameters as the next step towards systematic evidence synthesis for infectious diseases.

Competing interests. We declare we have no competing interests.

Funding. Work funded by the UK Engineering and Physical Sciences Research Council (refs EP/J016934/2 and EP/L001950/1).

References

1. Heesterbeek H *et al.* 2015 Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, aaa4339. (doi:10.1126/science.aaa4339)
2. Angelis DD, Presanis AM, Birrell PJ, Tomba GS, House T. 2015 Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics* **10**, 83–87. (doi:10.1016/j.epidem.2014.09.004)
3. Keeling MJ, Rohani P. 2007 *Modeling infectious diseases in humans and animals*. Princeton, NJ: Princeton University Press.
4. Diekmann O, Heesterbeek H, Britton T. 2013 *Mathematical tools for understanding infectious diseases dynamics*. Princeton, NJ: Princeton University Press.
5. Pellis L, Spencer SEF, House T. 2015 Real-time growth rate for general stochastic SIR epidemics on unclustered networks. *Math. Biosci.* **265**, 65–81. (doi:10.1016/j.mbs.2015.04.006)
6. Fraser C, Riley S, Anderson RM, Ferguson NM. 2004 Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151. (doi:10.1073/pnas.0307506101)
7. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA. 2009 Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect. Dis.* **9**, 291–300. (doi:10.1016/S1473-3099(09)70069-6)
8. Baguelin M, Hoek AJV, Jit M, Flasche S, White PJ, Edmunds WJ. 2010 Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. *Vaccine* **28**, 2370–2384. (doi:10.1016/j.vaccine.2010.01.002)
9. Gani R, Hughes H, Fleming D, Griffin T, Medlock J, Leach S. 2005 Antiviral therapy and hospitalization estimates during possible influenza pandemic. *Emerg. Infect. Dis.* **11**, 1355–1362. (doi:10.3201/eid1109.041344)
10. Muthuri SG *et al.* 2014 Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A(H1N1pdm09) virus infection: a meta-analysis of individual participant data. *Lancet Respir. Med.* **2**, 395–404. (doi:10.1016/S2213-2600(14)70041-4)
11. Hollingsworth TD, Klinkenberg D, Heesterbeek H, Anderson RM. 2011 Mitigation strategies for pandemic influenza A: balancing conflicting policy objectives. *PLoS Comput. Biol.* **7**, e1001076. (doi:10.1371/journal.pcbi.1001076)
12. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, Nicoll A. 2009 Closure of schools during an influenza pandemic. *Lancet Infect. Dis.* **9**, 473–481. (doi:10.1016/S1473-3099(09)70176-8)
13. Simmons K, Gambhir M, Leon J, Lopman B. 2013 Duration of immunity to norovirus gastroenteritis. *Emerg. Infect. Dis.* **19**, 1260–1267. (doi:10.3201/eid1908.130472)
14. Earn DJD, Rohani P, Bolker BM, Grenfell BT. 2000 A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670. (doi:10.1126/science.287.5453.667)
15. Hens N, Ayele G, Goeyvaerts N, Aerts M, Mossong J, Edmunds J, Beutels P. 2009 Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC Infect. Dis.* **9**, 187. (doi:10.1186/1471-2334-9-187)
16. Eames KTD, Tilston NL, Edmunds WJ. 2011 The impact of school holidays on the social mixing patterns of school children. *Epidemics* **3**, 103–108. (doi:10.1016/j.epidem.2011.03.003)
17. Bacaër N, Ouifki R. 2007 Growth rate and basic reproduction number for population models with a simple periodic factor. *Math. Biosci.* **210**, 647–658. (doi:10.1016/j.mbs.2007.07.005)
18. Vinjé J. 2010 A norovirus vaccine on the horizon? *J. Infect. Dis.* **202**, 1623–1625. (doi:10.1086/657088)
19. Camacho A *et al.* 2015 Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. *PLoS Curr. Outbreaks* **7**. (doi:10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2)
20. WHO Ebola Response Team. 2016 Ebola virus disease among male and female persons in West Africa. *New Engl. J. Med.* **374**, 96–98. (doi:10.1056/NEJMc1510305)
21. Ksiazek TG *et al.* 1999 Clinical virology of Ebola hemorrhagic fever (EHF): virus, virus antigen, and IgG and IgM antibody findings among EHF patients in Kikwit, Democratic Republic of the Congo, 1995. *J. Infect. Dis.* **179**(Suppl. 1), S177–S187. (doi:10.1086/514321)
22. Pitzer VE, Leung GM, Lipsitch M. 2007 Estimating variability in the transmission of severe acute respiratory syndrome to household contacts in Hong Kong, China. *Am. J. Epidemiol.* **166**, 355–363. (doi:10.1093/aje/kwm082)
23. Charleston B *et al.* 2011 Relationship between clinical signs and transmission of an infectious disease and the implications for control. *Science* **332**, 726–729. (doi:10.1126/science.1199884)
24. Carrat F, Vergu E, Ferguson NM, Lemaître M, Cauchemez S, Leach S, Valleron AJ. 2008 Time lines of infection and disease in human influenza: a review

- of volunteer challenge studies. *Am. J. Epidemiol.* **167**, 775–785. (doi:10.1093/aje/kwm375)
25. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, Graham DY. 2008 Norwalk virus shedding after experimental human infection. *Emerg. Infect. Dis.* **14**, 1553–1557. (doi:10.3201/eid1410.080117)
 26. Acton FS. 1970 *Numerical methods that work*. New York, NY: Harper & Row.
 27. Pereyra V, Scherer G. 2010 *Exponential data fitting and its applications*. Sharjah, United Arab Emirates: Bentham eBooks.
 28. Gilks WR, Richardson S, Spiegelhalter DJ. 1995 *Markov chain Monte Carlo in practice*. London, UK: Chapman and Hall.
 29. Brooks S, Gelman A, Jones GL, Meng X-L (eds). 2011 *Handbook of Markov chain Monte Carlo*. Boca Raton, FL: CRC Press.
 30. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953 Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092. (doi:10.1063/1.1699114)
 31. Hastings WK. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109. (doi:10.1093/biomet/57.1.97)
 32. Roberts GO, Rosenthal JS. 2009 Examples of adaptive MCMC. *J. Comput. Graphical Stat.* **18**, 349–367. (doi:10.1198/jcgs.2009.06134)
 33. Roberts GO, Tweedie RL. 1996 Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**, 341–363. (doi:10.2307/3318418)
 34. Duane S, Kennedy A, Pendleton BJ, Roweth D. 1987 Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216–222. (doi:10.1016/0370-2693(87)91197-X)
 35. Neal RM. 2011 MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo* (eds S Brooks, A Gelman, GL Jones, X-L Meng), pp. 113–162. Boca Raton, FL: CRC Press.
 36. Girolami M, Calderhead B. 2011 Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. B* **73**, 123–214. (doi:10.1111/j.1467-9868.2010.00765.x)
 37. Betancourt M. 2013 A general metric for Riemannian manifold Hamiltonian Monte Carlo. In *Geometric science of information* (eds F Nielsen, F Barbaresco). Lecture Notes in Computer Science, vol. 8085, pp. 327–334. Berlin, Germany: Springer. (doi:10.1007/978-3-642-40020-9_35)
 38. Calderhead B, Girolami M. 2011 Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. *Interface Focus* **1**, 821–835. (doi:10.1098/rsfs.2011.0051)
 39. Kramer A, Calderhead B, Radde N. 2014 Hamiltonian Monte Carlo methods for efficient parameter estimation in steady state dynamical systems. *BMC Bioinform.* **15**, 253. (doi:10.1186/1471-2105-15-253)
 40. Lan S, Stathopoulos V, Shahbaba B, Girolami M. 2014 Markov chain Monte Carlo from Lagrangian dynamics. *J. Comput. Graph. Stat.* **24**, 357–378. (doi:10.1080/10618600.2014.902764)
 41. Lan S, Streets J, Shahbaba B. 2014 Wormhole Hamiltonian Monte Carlo. In *Proc. 28th AAAI Conf. on Artificial Intelligence, Quebec City, Canada, 27–31 July 2014*, pp. 1953–1959.
 42. House T. 2014 Epidemiological dynamics of Ebola outbreaks. *eLife* **3**, e03908. (doi:10.7554/eLife.03908)
 43. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. 2007 Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc. Natl Acad. Sci. USA* **104**, 17 441–17 446. (doi:10.1073/pnas.0708559104)