



Research article

Bayesian inference in epidemics: linear noise analysis

Samuel Bronstein¹, Stefan Engblom^{2,*} and Robin Marin²

¹ Department of Mathematics and Applications, ENS Paris, 75005 Paris, France

² Division of Scientific Computing, Department of Information Technology, Uppsala University, SE-751 05 Uppsala, Sweden

* Correspondence: Email: stefane@it.uu.se; Tel: +46184712754; Fax: +4618511925.

Abstract: This paper offers a qualitative insight into the convergence of Bayesian parameter inference in a setup which mimics the modeling of the spread of a disease with associated disease measurements. Specifically, we are interested in the Bayesian model's convergence with increasing amounts of data under measurement limitations. Depending on how weakly informative the disease measurements are, we offer a kind of 'best case' as well as a 'worst case' analysis where, in the former case, we assume that the prevalence is directly accessible, while in the latter that only a binary signal corresponding to a prevalence detection threshold is available. Both cases are studied under an assumed so-called linear noise approximation as to the true dynamics. Numerical experiments test the sharpness of our results when confronted with more realistic situations for which analytical results are unavailable.

Keywords: parameter estimation; Bayesian modeling; stochastic epidemiological models; network model; Ornstein-Uhlenbeck process

1. Introduction

Computational models in epidemics are commonly relied upon to estimate the disease spread at fairly large spatial- and temporal scales, often referred to as *scenario generation*. With the increasing volumes and improved resolution of data from, e.g., mobile apps and disease testing time series from hospitals and nursing homes, predictive data-driven models formed from first principles are within reach. The accuracy of such models is ultimately limited by the specifics of the available disease surveillance data. In this paper we attempt to gain a qualitative understanding of how Bayesian inference of epidemiological parameters may be expected to perform and what the limiting factors are.

Epidemiological models are typically formed by postulating laws for the flow of individuals between different *compartments* in a large population and have been studied in this form for a long time. When connected with data in the form of observations, the associated inference problem is also a

fairly mature field, see, e.g., [1–3]. With the increasing qualities and quantities of data, the various incarnations of data-driven modeling allow for substantially higher modeling resolution compared to traditional macroscopic approaches [4–8]. For example, individual-level contact tracing has been used to study disease spread models at various population sizes [9–13]. Data-driven models have allowed epidemic and endemic conditions to be investigated at a level of detail not previously possible [14–16].

The identification of the epidemiological parameters from data falls under the scope of problems formally studied in *System Identification* [17]. However, a ‘system’ viewpoint of epidemiological modeling is not yet standard, and identification of parameters is rather more often approached through calibration of residuals [5, 16, 18], sometimes also blending in aspects of Bayesian arguments. Fully Bayesian approaches are rarer, albeit with some exceptions [8, 19], typically due to the technical difficulties with formulating suitable (pseudo-)likelihoods and the slow convergence associated with the conditioning of the problem. Although Bayesian inference is notably well-posed thanks to its use of prior distributions, the posterior distribution itself is often a computationally ill-conditioned object whenever strong parameter correlations are present, e.g., resulting from nearly singular maps from parameters to observables. These conditions, together with the societal importance of this modeling domain, make it relevant to reason around the limits of fully Bayesian techniques.

Fundamental questions concerning Bayesian convergence in general settings, including infinite-dimensional ones, have been treated [20, 21], and have also been revisited with specific tools and applications in mind, e.g., Gaussian processes and other machine learning algorithms [22, 23]. “Brittleness” [24, 25], or high sensitivity to small perturbations have been proposed to be a problematic phenomenon under certain conditions. For applications, this points to the importance of striking a balance between the granularity of the model and the information content and level of detail of the available data.

With the specific aim of reaching qualitative conclusions for Bayesian modeling in epidemiology, we will connect some usual modeling approaches in infectious disease spread with a basic linear stochastic differential equation. We analyze the pre-asymptotic Bayesian posterior in the high-quality data regime as well as the estimate’s convergence under weakly informative disease measurements. This is motivated by current trends in disease spread monitoring through for example sewage water analysis and symptoms data collection using smartphones [26, 27].

The rest of the paper is organized as follows. In §2 we suggest the Ornstein-Uhlenbeck process as a meta-model of epidemiological models, covering various Susceptible-Infected-Recovered (SIR)-type models locally in time, during endemic as well as under epidemic conditions. In §3 we briefly analyze the posterior convergence when accurate full state measurements are available. The main results are found in §4 where we more fully develop an analysis for the case of poorly informative data. We offer some examples of relevance in §5 and a summarizing discussion is found in §6.

2. The Ornstein-Uhlenbeck process

We connect in this section the Ornstein-Uhlenbeck process with some basic epidemiological models via linearization and quasi steady-state arguments. We also briefly discuss, by means of a backward analysis, how continuous and discrete time in this setting can be connected in probability law under a certain parameter map.

2.1. Disease spread modeling

Epidemiological modeling typically involves ordinary differential equations (ODEs), e.g., the SIR-model which is often used to model the annual flu [1, 28],

$$\left. \begin{array}{l} S'(t) = -(\beta\Sigma^{-1})S(t)I(t) \\ I'(t) = (\beta\Sigma^{-1})S(t)I(t) - \gamma I(t) \\ R'(t) = \gamma I(t) \end{array} \right\} \quad (2.1)$$

in terms of Susceptible, Infected, and Recovered individuals, respectively, and for a total population size $\Sigma \equiv S + I + R$. The model parameters β and γ define the transmission and recovery rates, and the basic reproduction number is then given by $R_0 = \beta/\gamma$, that is, the expected number of new infections resulting from a single index case [1].

Simplifying the SIR-model by removing the R -compartment one obtains the SIS-model, where the recovered state has been removed and hence effectively identified with the susceptible state,

$$\left. \begin{array}{l} S'(t) = -(\beta\Sigma^{-1})S(t)I(t) + \gamma I(t) \\ I'(t) = (\beta\Sigma^{-1})S(t)I(t) - \gamma I(t) \end{array} \right\} \quad (2.2)$$

and with the same R_0 as the SIR-model. It is sometimes useful to add an environmental compartment expressing the *infectious pressure* φ , i.e., the amount of infectious substance per unit of space. This defines the SIS_E -model [29], where the E stands for the Environmental compartment,

$$\left. \begin{array}{l} S'(t) = -\beta S(t)\varphi(t) + \gamma I(t) \\ I'(t) = \beta S(t)\varphi(t) - \gamma I(t) \\ \varphi'(t) = \Sigma^{-1}I(t) - \rho\varphi(t) \end{array} \right\} \quad (2.3)$$

where the indicated governing equation for φ is just a basic example. The SIS_E -model is convenient to adapt to spread over a network and to detection situations involving sampling the environment [19]. Due to the indirect transmission the basic reproduction number now scales as a square root, with $R_0 = \sqrt{\beta/(\gamma\rho)}$.

The use of ODEs can be justified for epidemiological models in sufficiently large populations. In smaller populations, e.g., networks of small communities, stochastic variants are necessary to properly capture the underlying dynamics of the spread [3, 30]. For example, a stochastic differential equation (SDE)-version of (2.2) was analyzed in [31] which essentially replaces βdt by a Brownian diffused version $\beta dt + \eta dB(t)$. This has the effect of lowering the basic reproduction number to $R_0 = (\beta - \eta^2/2)/\gamma$. A first-principle stochastic approach is to rather express the dynamics as a continuous-time discrete state Markov chain (CTMC). The SIS-model above is then defined via discrete transitions with exponentially distributed waiting times,

$$\left. \begin{array}{ccc} S + I & \xrightarrow{\beta\Sigma^{-1}} & 2I \\ I & \xrightarrow{\gamma} & S \end{array} \right\} \quad (2.4)$$

meaning, e.g., that one infectious individual may infect one susceptible individual, and such that β and γ are now understood as rate parameters in the driving Poissonian processes. One can show that now

$R_0 = (1 - \Sigma^{-1}/2) \times \beta/\gamma$ such that consistency with the previous SDE formulation would require that $\eta^2 = \Sigma^{-1}\beta$ and hence follows from the typical Poissonian population-dependent noise scaling $\eta \propto \Sigma^{-1/2}$.

Let $P(t)$ be some given measure of the intensity of the disease, such as the absolute or relative disease prevalences, $I(t)$ or $\Sigma^{-1}I(t)$, respectively. Another alternative could be the infectious pressure $\varphi(t)$, which also informs on the current disease intensity. Gathering data from the disease spread means collecting information about $P(t)$, typically in the form of a time-series, $(F(P(t_i)))_i$, say, for some measurement operator F . Consider *endemic* conditions first, that is, P is considered stationary. As an ansatz, suppose

$$P(t_i) \sim \mathbf{P}[P = p] = \rho_\infty(p) \propto \exp(-2/\sigma^2 E(p)), \quad (2.5)$$

for some *epidemiological potential* E . Under endemic conditions we can expect this to be a single well potential, say,

$$E(p) \propto (p - \alpha)^2 + \text{constant}, \quad (2.6)$$

or at least can locally be approximated by this form. This is consistent with the *Ornstein-Uhlenbeck (OU-) process* [32],

$$X(0) = X_0, \quad (2.7)$$

$$dX(t) = k - \mu X(t) dt + \sigma dW(t), \quad (2.8)$$

where $X(t) \in \mathbf{R}$ and where the parameters of the model are $\theta = [k, \mu, \sigma] \in \mathbf{R}_+^3$ and in (2.6), $\alpha = k/\mu$.

As a concrete example, we may approximate the continuous-time Markov chain (2.4) by linearizing the rates around the non-trivial stationary state $I_\infty = \Sigma(1 - \gamma/\beta)$ of (2.2). Similarly, the noise term σ can be determined by inspecting the total variance of the Poissonian rates in (2.4) around this equilibrium. This yields a *linear noise OU*-approximation to the state variable I of the SIS-model (2.4) with parameters

$$\left. \begin{array}{lll} k &= \Sigma(\beta - \gamma)^2/\beta &= \gamma\Sigma(R_0 - 1)^2/R_0 \\ \mu &= \beta - \gamma &= \gamma(R_0 - 1) \\ \sigma^2 &= 2\Sigma(\beta - \gamma)^2/\beta &= 2\gamma\Sigma(R_0 - 1)/R_0 \end{array} \right\} \quad (2.9)$$

where we use the approximation $R_0 = (1 - \Sigma^{-1}/2) \times \beta/\gamma \approx \beta/\gamma$. Hence, *with this specific interpretation of the OU-process*,

$$R_0 = \left(1 - \Sigma^{-1}k/\mu\right)^{-1} = \left(1 - \Sigma^{-1}\alpha\right)^{-1}. \quad (2.10)$$

The quality of this approximation is exemplified in Figure 1; the effect of linearizing around the stationary state can be seen as a slightly too fast transient compared to the Markov chain.

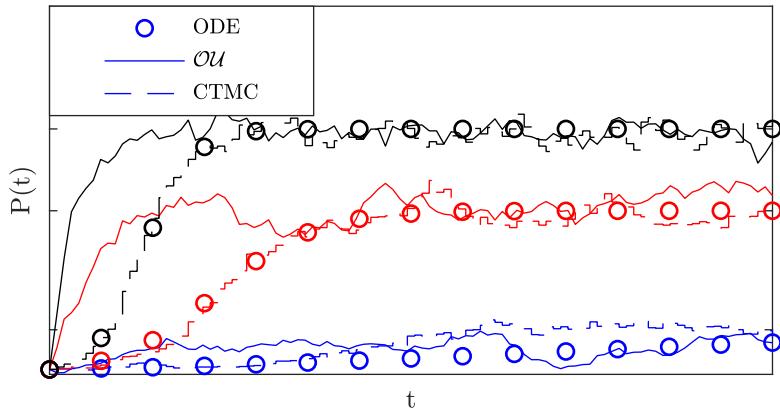


Figure 1. The approximation (2.9) exemplified for $R_0 = [1.1, 1.5, 2]$ (bottom and up), with $\Sigma = 1000$ and $\gamma = 1$. For comparison the ODE-, the \mathcal{OU} -, and the continuous-time Markov chain (CTMC) interpretation are shown.

During *epidemic* conditions we are rather sampling a transient of the process and, moreover, there are also typically many kinds of feedback involved, e.g., from minor adjustments of individual level behavior, to major societal changes and governmental intervention strategies. Although this means that there is now a greater challenge in formulating a data-centric meta-model of the situation we may still consider a window of time for which the disease spread parameters are approximately constant and such that (2.7)–(2.8) remain a relevant model of the situation. A difference is then that data is gathered out of equilibrium (hence away from α) and presumably also under more noisy conditions with larger values of σ . More general epidemiological potentials could be considered through the SDE in gradient formulation,

$$dP(t) = -\nabla E(P_t) dt + \sigma dW(t), \quad (2.11)$$

for which the stationary measure is still given by (2.5). For general SDE models, a change of variables allows us to locally consider an SDE with constant noise [33] and, in turn, any time-homogeneous SDE with constant noise,

$$dX(t) = f(X_t)dt + \sigma dW(t), \quad (2.12)$$

is of course readily linearized into an \mathcal{OU} -process.

2.2. Exact sampling

The \mathcal{OU} -process (2.7)–(2.8) is a Gaussian continuous process with mean and covariance given by

$$\mathbb{E}[X_t] = \frac{k}{\mu}(1 - e^{-\mu t}) + e^{-\mu t}X_0, \quad (2.13)$$

$$\text{Cov}[X_t, X_s] = \frac{\sigma^2}{2\mu}(e^{-\mu|t-s|} - e^{-\mu(t+s)}), \quad (2.14)$$

such that the stationary distribution is $X_\infty \sim \mathcal{N}(k/\mu, \sigma^2/(2\mu))$.

Numerical simulation procedures are typically based on Euler-type discretization methods. Assume for convenience a fixed numerical time step Δt and compute

$$X_{n+1} = X_n + (k - \mu X_n) \Delta t + \sigma \Delta W_n, \quad (2.15)$$

with ΔW_n being i.i.d. normally distributed numbers of zero mean and variance Δt . The numerical trajectory (X_n) is then an approximation to the $O\mathcal{U}$ -process $(X(t_n))$ at discrete times $t_n = n\Delta t$ for $n = 0, 1, \dots$ and is also a Gaussian vector with mean and covariance

$$\mathbb{E}[X_n] = \frac{k}{\mu} + (1 - \mu \Delta t)^n (X_0 - \frac{k}{\mu}), \quad (2.16)$$

$$\text{Cov}(X_n, X_{n+p}) = \frac{\sigma^2 \Delta t}{1 - (1 - \mu \Delta t)^2} (1 - (1 - \mu \Delta t)^{2n}) (1 - \mu \Delta t)^p. \quad (2.17)$$

Actually, (2.15) forms an $AR(1)$ -sequence [34]. Comparing (2.13)–(2.14) with (2.16)–(2.17) we readily find the following useful result.

Proposition 2.1 (*Exact $O\mathcal{U}$ -samples and backward analysis*). *The discrete process (X_n) given by the explicit Euler method (2.15) follows the law of the discrete samples of an exact $O\mathcal{U}$ -process with perturbed parameters $(k_{\Delta t}, \mu_{\Delta t}, \sigma_{\Delta t})$, where*

$$\left. \begin{array}{l} 1 - \mu \Delta t = \exp(-\Delta t \mu_{\Delta t}) \\ \frac{k}{\mu} = \frac{k_{\Delta t}}{\mu_{\Delta t}} \\ \frac{\sigma^2}{2\mu - \mu^2 \Delta t} = \frac{\sigma_{\Delta t}^2}{2\mu_{\Delta t}} \end{array} \right\} \quad (2.18)$$

This system of equations can be solved explicitly provided $\Delta t < \mu^{-1}$,

$$\left. \begin{array}{l} k_{\Delta t} = -\frac{k \log(1 - \mu \Delta t)}{\mu \Delta t} \\ \mu_{\Delta t} = -\frac{\log(1 - \mu \Delta t)}{\Delta t} \\ \sigma_{\Delta t}^2 = -2\sigma^2 \frac{\log(1 - \mu \Delta t)}{2\mu \Delta t - \mu^2 \Delta t^2} \end{array} \right\} \quad (2.19)$$

These perturbed parameters are all first order perturbations in Δt of the exact parameters.

The inverse of (2.19) is that

$$\left. \begin{array}{l} k = \mu \frac{k_{\Delta t}}{\mu_{\Delta t}} \\ \mu = \frac{1 - \exp(-\Delta t \mu_{\Delta t})}{\Delta t} \\ \sigma^2 = \sigma_{\Delta t}^2 \frac{2 - \Delta t \mu}{2\mu_{\Delta t}} \mu \end{array} \right\} \quad (2.20)$$

It follows that, for *given* parameters $[k_{\Delta t}, \mu_{\Delta t}, \sigma_{\Delta t}]$, a forward Euler simulation using the new set of parameters $[k, \mu, \sigma]$ defined by (2.20) will produce a sample trajectory obeying the law of the $O\mathcal{U}$ -process with the given parameters exactly.

3. Bayesian filtering with full information

We first briefly consider in this section the behavior of the Bayesian posterior as a function of the prior and of data in the sense of full state measurements. Hence we suppose that sampled process

data $D_N = (d_i)_{i=0}^N$ is available at some fixed time step h , that is, $d_i = X(t_i)$ with $t_i = ih$ and $X(\cdot)$ an $O\mathcal{U}$ -process with X_0 for simplicity considered a known sure number.

It will be convenient to consider the following reparametrization of the $O\mathcal{U}$ -process

$$\left. \begin{array}{l} \alpha = \frac{k}{\mu} \in \mathbf{R} \\ \beta = e^{-\mu h} \in (0, 1) \\ \gamma = \frac{2\mu}{\sigma^2} \cdot \frac{1}{1-e^{-2\mu h}} \in \mathbf{R}_+ \end{array} \right\} \quad (3.1)$$

and we put $u \equiv [\alpha, \beta, \gamma]$ for brevity, also defining $u_0 \equiv [\alpha_0, \beta_0, \gamma_0]$, i.e., the true parameters generating the data.

Under this reparametrization we readily find from Proposition 2.1 that the *exact* discrete process can be written in the $AR(1)$ -sequence form

$$X_{i+1} = \beta X_i + \alpha(1 - \beta) + \gamma^{-1/2} \xi_i, \quad (3.2)$$

where the $\xi_i \sim \mathcal{N}(0, 1)$ are independent. Writing $\delta := \alpha(1 - \beta)$, explicit least squares estimators for the parameters can be found by solving

$$\min_{\delta, \beta} \|A \times [\delta, \beta]^T - b\|_2^2 =: \min_{\delta, \beta} \left\| \begin{bmatrix} \mathbf{1} & [d_i]_{i=0}^{N-1} \end{bmatrix} \times [\delta, \beta]^T - [d_{i+1}]_{i=0}^{N-1} \right\|_2^2, \quad (3.3)$$

where the brackets over the data form column vectors. The residual of this solution implies the corresponding estimator for γ ,

$$\hat{\gamma}^{-1} = (N - 2)^{-1} \|A \times [\hat{\delta}, \hat{\beta}]^T - b\|_2^2. \quad (3.4)$$

By the Gaussian character of the $O\mathcal{U}$ -process these estimators coincide with the maximum likelihood estimators.

The Bayesian convergence to the true parameters as $N \rightarrow \infty$ can be characterized by either the Bernstein von Mises theorem (BvM), for h fixed, or by contrast functions convergence for a fixed time window $t \in [0, T]$. BvM states that the Bayesian posterior converges to a normal distribution centered at the maximum likelihood estimate with the inverse Fisher information matrix (FIM) as covariance [35–37]. Similarly, the contrast functions approach considers the convergence of the approximation of discretized observations, approaching the same normal distribution for N large enough [38–40]. In the present case the FIM can be determined explicitly using the log-likelihood $\log \mathcal{L}(X_{i+1}|X_i; u)$ induced by (3.2) and the definition

$$\Sigma \equiv -\mathbb{E} \left[\frac{\partial^2}{\partial u^2} \log \mathcal{L} \right] (u_0). \quad (3.5)$$

The result is that

$$\Sigma = \text{Diag} \left((1 - \beta_0)^2 \gamma_0, 1/(1 - \beta_0^2), 1/(2\gamma_0^2) \right). \quad (3.6)$$

Proposition 3.1 (*BvM Theorem for (3.2)*). *Provided that the prior has a continuous and positive density in an open neighborhood of (α, β, γ) , as $N \rightarrow \infty$ we have that the ML-estimators (3.3)–(3.4) converge to the true values as*

$$\sqrt{N}(\hat{u} - u_0) \xrightarrow{d} \mathcal{N}(0, \Sigma^{-1}), \quad (3.7)$$

(convergence in the sense of distribution) with Σ defined by (3.6). Alternatively, (3.7) dictates the asymptotic convergence of the mean of the Bayesian posterior.

To summarize, the asymptotic variances of all parameters are independent of α_0 . The noise term γ_0 mainly affects the convergence of α_0 (and γ_0 in an absolute sense). Finally, an increase of β_0 implies a faster convergence towards β_0 at the cost of a slower convergence towards α_0 . Under the interpretation of an SIS-model (2.9)–(2.10), the parameter α_0 is in one-to-one correspondence with the basic reproduction number R_0 , hence its central interest here.

The asymptotic nature of both the BvM Theorem and the contrast function convergence is a poor match in epidemiological situations where data is often scarce and poorly informative at the time scale over which the parameters can be considered static. This motivates our interest in also the pre-asymptotic regime of the Bayesian posterior. We therefore consider prior densities of the specific form

$$\pi(u) \propto \gamma^r \exp\left(-\frac{\gamma}{2}P(\alpha, \beta)\right), \quad (3.8)$$

where, for integrability, P is to be a polynomial of degree 2 in α and nonnegative for $\beta \in (0, 1)$. This choice has the convenient property that the posterior measure after N observations is

$$\Pi^{(N)}(u) \propto \gamma^{\frac{N}{2}+r} \exp\left(-\frac{\gamma}{2}(Q_N + P)(\alpha, \beta)\right), \quad (3.9)$$

where

$$Q_N(\alpha, \beta) \equiv \sum_{i=0}^{N-1} (d_{i+1} - \beta d_i - \alpha(1 - \beta))^2, \quad (3.10)$$

and where $d_i = X(t_i)$ is the observation at time t_i .

Remark. To include also the case of flat priors, while avoiding technicalities for nonintegrable densities, we note that, since the value of a constant prior has no influence on the posterior, we can still determine a posterior after a single initial observation, and then use this posterior as a prior for the rest of the observations.

The following result examines the convergence of the Bayesian posterior.

Theorem 3.2 (Convergence of log-likelihood). *Consider the polynomial q_N :*

$$q_N(\alpha, \beta) \equiv N^{-1} Q_N(\alpha, \beta) = N^{-1} \sum_{i=0}^{N-1} (d_{i+1} - \beta d_i - \alpha(1 - \beta))^2. \quad (3.11)$$

Then as $N \rightarrow \infty$ we have the almost everywhere uniform convergence on every compact of q_N towards the function f ,

$$f(\alpha, \beta) = (1 - \beta)^2 \left[\frac{1}{\gamma_0(1 - \beta_0^2)} + (\alpha_0 - \alpha)^2 \right] + \frac{2\beta}{\gamma_0(1 + \beta_0)}. \quad (3.12)$$

Proof. For a general point (α, β) , we rely on the ergodic theory of Markov chains [41, p. 472] to get that $N^{-1} \sum g(d_i, d_{i+1})$ converges almost surely towards $\mathbb{E}[g(d_i, d_{i+1})]$ for g integrable against the stationary measure. Using (2.13)–(2.14) this implies the limits:

$$N^{-1} \sum_{i=0}^{N-1} d_{i+1} - \beta d_i \longrightarrow \alpha_0(1 - \beta), \quad (3.13)$$

$$N^{-1} \sum_{i=0}^{N-1} (d_{i+1} - \beta d_i)^2 \longrightarrow (\alpha_0(1 - \beta))^2 + \frac{1 + \beta^2}{\gamma_0(1 - \beta_0^2)} - \frac{2\beta\beta_0}{\gamma_0(1 - \beta_0^2)}. \quad (3.14)$$

Combined, we obtain the claimed limit in a pointwise sense. As each (q_N) is a polynomial in (α, β) of bounded degree, pointwise convergence is equivalent to convergence of the coefficients, and hence implies the convergence on all compact sets.

An obvious extension is to consider measurements polluted by noise, say, $d_i = X_i + \eta_i$, for i.i.d. $\eta_i \sim \mathcal{N}(0, \eta)$ and some variance η . The posterior so obtained is readily computed via a recursive Kalman filter but does not have a simple analytic form. To first order in η , however, (3.8)–(3.9) still holds provided γ is replaced with $[1/\gamma + \eta(1 + \beta^2)]^{-1}$ and Q_N in (3.10) is replaced with

$$Q'_N \equiv \sum_{i=0}^{N-1} [d_{i+1} - \beta d_i - \alpha(1 - \beta) + \eta\beta\gamma(d_i - \beta d_{i-1} - \alpha(1 - \beta))]^2, \quad (3.15)$$

where the right $\eta\beta\gamma$ -term is skipped when $i = 0$. Intuitively, the first order effect of noise in data is to broaden the posterior with the variance of this noise. We next proceed to investigate more severe truncations of the measurements from the epidemiological process.

4. Bayesian filtering of surveillance data

In the previous section we considered full process data to be available without any extrinsic noise. This models the best possible Bayesian setup but is also clearly unrealistic in epidemics. As a model of a more challenging situation we consider in this section the recorded data to be some (possibly stochastic) function of the *OU*-process $X(\cdot)$, which itself is considered a latent variable. We are initially concerned with *binary data* of the form $D_N = (d_i)_{i=0}^N$ where $d_i = Y_i = \mathbf{1}_{X(t_i) \geq c}$, again over a uniform grid in time $t_i = ih$, and for a known *filter* cut-off value c (see also the related setup in [42]). That is, the epidemiological interpretation is that the data is considered to be time-discrete information about whether or not the prevalence $X(\cdot)$ of the population is above or below a certain known threshold c . With the prevalence process hidden one is forced to estimate it simultaneously with any parameter estimates. Using the *AR(1)*-form (3.2) we have that the stationary measure for the $(p + 1)$ steps (X_i, \dots, X_{i+p}) is Gaussian $\mathcal{N}(\alpha, \Sigma)$ with Σ given by (for $p \geq 0$)

$$\Sigma = \Sigma(\beta, \gamma) = \frac{\gamma^{-1}}{1 - \beta^2} \begin{bmatrix} 1 & \beta & \beta^2 & \cdots & \beta^p \\ \beta & 1 & \beta & \cdots & \beta^{p-1} \\ \beta^2 & \ddots & \ddots & \ddots & \cdots \\ \vdots & \ddots & \beta & 1 & \beta \\ \beta^p & \cdots & \beta^2 & \beta & 1 \end{bmatrix}. \quad (4.1)$$

In order to be able to conduct an analysis where information is obtained only from the observable $Y(\cdot)$, we take (4.1) as a motivation for the following:

Assumption 4.1 (*p-step Markovian stationary assumption*). We say that we work under the stationary assumption whenever we assume that the law of the latent variable $Y(\cdot)$ can be directly inferred from the stationary p -step law (4.1).

We stress that it is known that this type of clipped Gaussian processes are not p -step Markov for any p [43]. Assumption 4.1 rather serve as an approximation where we approximately model $Y(\cdot)$ as if it was p -step Markov with law deduced from (4.1). The pseudo-likelihood for the filtered variable Y then becomes

$$\mathcal{L}_N(u) = \prod_{i=0}^{N-p} \varphi_u(Y_i, \dots, Y_{i+p}) \quad (4.2)$$

where $\varphi_u(e)$ denotes the probability for a Gaussian stationary filtered process with parameters u to be $e \in \{0, 1\}^{p+1}$. We also define the pseudo-potential

$$\Phi_N^u \equiv \log \mathcal{L}_N(u) = \sum_{i=0}^{N-p} \log \varphi_u(Y_i, \dots, Y_{i+p}). \quad (4.3)$$

We show below in §4.1 that the stationary assumption allows for converging posterior estimates for the parameter β , but leaves any prior density unchanged over a certain curve in (α, γ) . In §4.2 we briefly consider the filter cutoff value c (the sensitivity) to be uncertain and straightforwardly show that any prior density on c is unaffected by data. We next replace the sharp deterministic filter by a stochastic filter implementing a sigmoidal response function and sharpen our results in this more general setting in §4.3. Finally, in §4.4 we consider slightly more informative measurements consisting of a finite discrete response and we show that this resolves the singularity issues associated with purely binary measurements.

4.1. Binary measurements

We first show that under the p -fold stationary assumption one can estimate the correlation term β rather well, but increasing the gap between the filter threshold c and the mean α has an effect on the likelihood which is indistinguishable in law from increasing the noise γ .

We start with a technical lemma.

Lemma 4.1 (*Equality in law*). *For u in any admissible set of parameters, let p_u be the law of (Y_0, \dots, Y_p) under Assumption 4.1 corresponding to $\mathcal{N}(\alpha_u, \Sigma_u)$. Then $p_u = p_w$ if and only if $\beta_u = \beta_w$ and $\sqrt{\gamma_u}(c - \alpha_u) = \sqrt{\gamma_w}(c - \alpha_w)$.*

Proof. (\implies) Suppose $p_u = p_w$. Consider (X_0^u, \dots, X_p^u) a Gaussian vector with law $\mathcal{N}(\alpha_u, \Sigma_u)$. As $\mathbf{P}(X_0^u \leq c) = \sum_{e \in \{0,1\}^p} \varphi(0, e)$, we deduce that $\mathbf{P}(X_0^u \leq c) = \mathbf{P}(X_0^w \leq c)$. This can be written as an equality between standard cumulative distribution functions of Gaussians, since $\sqrt{\gamma_u(1 - \beta_u^2)}(X_0^u - \alpha_u)$ and $\sqrt{\gamma_w(1 - \beta_w^2)}(X_0^w - \alpha_w)$ are standard Gaussians. This translates into:

$$\sqrt{\gamma_u(1 - \beta_u^2)}(c - \alpha_u) = \sqrt{\gamma_w(1 - \beta_w^2)}(c - \alpha_w) \quad (4.4)$$

We also have $\mathbf{P}(X_0^u, X_1^u \leq c) = \mathbf{P}(X_0^w, X_1^w \leq c)$, and this implies, with $A = \sqrt{\gamma_u(1 - \beta_u^2)}(c - \alpha_u)$,

$$\begin{aligned} \frac{1}{\sqrt{1 - \beta_u^2}} \int_{(-\infty, A]^2} \exp\left(-\frac{x^2 + y^2 - 2\beta_u xy}{\sqrt{1 - \beta_u^2}}\right) dx dy \\ = \frac{1}{\sqrt{1 - \beta_w^2}} \int_{(-\infty, A]^2} \exp\left(-\frac{x^2 + y^2 - 2\beta_w xy}{\sqrt{1 - \beta_w^2}}\right) dx dy. \end{aligned} \quad (4.5)$$

This last expression shows that the map $\beta_u \mapsto \mathbf{P}(X_0^u, X_1^u \leq c)$ is locally analytic. As a corollary of Slepian's lemma [44], this map is also increasing and hence strictly increasing. It follows that the last equality implies $\beta_u = \beta_w$.

(\Leftarrow) The law p_u is uniquely determined by its 2^{p+1} values. All these values are of the type

$$I(Q; u) = \int_{Q(A)} (\det M(\beta_u))^{-(p+1)/2} \times \exp(-xM(\beta_u)x^T) dx, \quad (4.6)$$

where $Q(A)$ is a product of $p + 1$ intervals, each being either $(-\infty, A]$ or $[A, +\infty)$. And so, if $\beta_u = \beta_w$ and $\sqrt{\gamma_u}(c - \alpha_u) = \sqrt{\gamma_w}(c - \alpha_w)$, the laws p_u and p_w are indeed equal.

Theorem 4.2 (Non-identifiability). *Assume $p \geq 1$. For any integrable prior π , non-zero on \mathcal{E} , and for f bounded, as $N \rightarrow \infty$,*

$$\int f \mathcal{L}_N d\pi \rightarrow \int_{\mathcal{E}} f d\tilde{\pi}, \quad (4.7)$$

where $\mathcal{E} \equiv \{u; \sqrt{\gamma}(c - \alpha) = \sqrt{\gamma_0}(c - \alpha_0), \beta = \beta_0\}$ and $\tilde{\pi}$ is the prior restricted to the set \mathcal{E} .

Proof. At first, thanks to Lemma 4.1, the law $\varphi(Y_1, \dots, Y_{i+p})$ is not characterized by $u_0 = [\alpha_0, \beta_0, \gamma_0]$, but rather by $[\sqrt{\gamma_0}(c - \alpha_0), \beta_0]$. As the Y_i 's are binary, the quantity $\varphi(Y_1, \dots, Y_{i+p})$ can only take 2^{p+1} values, which we denote by φ_e for $e \in \{0, 1\}^{p+1}$. This means that the log-likelihood under Assumption 4.1 can be written as a finite sum:

$$\log \mathcal{L}_N = \sum_{e \in \{0, 1\}^{p+1}} N(e) \log \varphi_e. \quad (4.8)$$

where

$$N(e) = \#\{i \in \{0, N-p\}; (Y_i, \dots, Y_{i+p}) = e\} \quad (4.9)$$

Now, as the discrete \mathcal{OU} -process is an $AR(1)$ -sequence [34], we may infer the convergence of $N(e)/N$ towards a stationary value:

$$\frac{\log \mathcal{L}_N}{N} \rightarrow \sum_{e \in \{0, 1\}^{p+1}} \varphi_e^{u_0} \log \varphi_e^u, \quad (4.10)$$

where φ_e^u denotes the law of a stationary filtered process with parameter u . We note that

$$\lim_{N \rightarrow \infty} \frac{\log \mathcal{L}_N}{N} = \sum_{e \in \{0, 1\}^{p+1}} \varphi_e^{u_0} \log \frac{\varphi_e^u}{\varphi_e^{u_0}} + \sum_{e \in \{0, 1\}^{p+1}} \varphi_e^{u_0} \log \varphi_e^{u_0}. \quad (4.11)$$

Up to a constant this limit is the negative of the Kullback-Leibler divergence between the laws φ^u and φ^{u_0} , which vanishes if and only if the two distributions are equal. We already know that the laws

of φ^u and φ^{u_0} are equal if and only if $\beta = \beta_0$ and $\sqrt{\gamma}(c - \alpha) = \sqrt{\gamma_0}(c - \alpha_0)$, i.e., if $u \in \mathcal{E}$. Consider the associated relation:

$$u = (\alpha, \beta, \gamma) \sim u' = (\alpha', \beta', \gamma') \text{ iff } \beta = \beta' \text{ and } \sqrt{\gamma}(c - \alpha) = \sqrt{\gamma'}(c - \alpha'). \quad (4.12)$$

It is straightforward to show that this relation is reflexive, symmetric, and transitive, thus forming an equivalence relation. Thanks to Lemma 4.1, we know that the map $u \mapsto \mathcal{L}_N(u)$ factorizes through the equivalence relation to a map $w \mapsto \tilde{\mathcal{L}}_N(w)$ where w ranges over the different equivalence classes of u .

Lemma 4.1 also allows us to state that the map $\tilde{\mathcal{L}}_N$ is injective. Consider w_0 in the equivalence class of u_0 . Then for any neighborhood W of w_0 , there is a $\delta > 0$ such that for N large enough, $w \notin W$ implies

$$\frac{\log \tilde{\mathcal{L}}_N(w_0)}{N} \geq \frac{\log \tilde{\mathcal{L}}_N(w)}{N} + \delta, \quad (4.13)$$

which means that the posterior measure of W converges towards 1. Since the factorized map $\tilde{\mathcal{L}}_N \rightarrow \delta_{w_0}$, the theorem follows.

The same argument implies that, if we consider the case $p = 0$, the posterior will be even more degenerate. For $p = 0$ one gets the convergence

$$\int f \mathcal{L}_N d\pi \rightarrow \int_{\mathcal{E}'} f d\tilde{\pi}, \quad (4.14)$$

where now $\tilde{\pi}$ is the prior restricted to \mathcal{E}' :

$$\mathcal{E}' \equiv \{u; \sqrt{\gamma_u(1 - \beta_u^2)}(c - \alpha_u) = \sqrt{\gamma_0(1 - \beta_0^2)}(c - \alpha_0)\}. \quad (4.15)$$

This case is relevant whenever we consider large gaps of time in between the measurements such that they can be considered practically independent.

At this point, let us remind that the quantities $\mathcal{L}_N(u)$ and $\Phi_N^p(u)$ are defined in Eqs (4.2) and (4.3).

Theorem 4.3 (*Rate of convergence of pseudo-potential*). *Under the p -Markovian stationary Assumption 4.1 there exists a mapping f_p and a constant $\sigma_p > 0$ such that, as $N \rightarrow \infty$, we have the convergence in law:*

$$\sqrt{N} \left(\frac{\Phi_N^p(u)}{N} - f_p(u) \right) \rightarrow \mathcal{N}(0, \sigma_p). \quad (4.16)$$

Proof. This is a direct consequence of the Central Limit Theorem, with $f_p(u)$ being the mean value of $\Phi_N^p(u)$ and σ_p a nonnegative constant.

4.2. Propagation of filter uncertainty

A relevant variation of the theme is to consider the parameter c (the “test sensitivity”) an uncertain parameter. Mathematically, this means considering the likelihood \mathcal{L}_N a function of $(\alpha, \beta, \gamma, c)$, and priors and posteriors depending also on c . However, the following result shows that this setup will not produce any more information about the parameters.

Theorem 4.4 (*Translation of filter uncertainty*). Consider a prior of the form $\mu(\alpha)\pi(c - \alpha, \beta, \gamma)$ and the pseudo-likelihood in §4.1. Then the resulting posterior measure will be of the form $\mu(\alpha)\Pi(c - \alpha, \beta, \gamma)$, where, for fixed c , $\Pi(c - \cdot, \cdot, \cdot)$ is the pseudo-posterior from the prior $\pi(c - \cdot, \cdot, \cdot)$.

Proof. This result is immediate once one realizes that the pseudo-likelihood has the property that, for any $t \in \mathbf{R}$,

$$\mathcal{L}_N(\alpha, \beta, \gamma, c) = \mathcal{L}_N(\alpha + t, \beta, \gamma, c + t). \quad (4.17)$$

In other words, the only information we can infer on the parameters (c, α) is the gap $c - \alpha$. Any uncertainty of c can be understood as an uncertainty on α , and vice-versa.

4.3. Non-perfect binary measurements

Rather than a sharp cut-off value c , most environmental sampling methods obey some kind of *sensitivity response*, e.g., of sigmoid character, with a quick rise in the detection probability as one progresses through some threshold region. Examples here could include sampling and subsequent analysis of sewage water or animal droppings, but this would also be a relevant model in the case of statistical regression estimates using data obtained via self-reporting smartphones applications.

A general ansatz to capturing this situation is to consider

$$Y_i \sim \mathcal{B}(s(X_i)), \quad (4.18)$$

where s is a map from \mathbf{R} to $[0, 1]$ and \mathcal{B} denotes the Bernoulli law. Typically, the map s is sigmoidal with a gradient around the threshold c which depends on the sensitivity and specificity of the test. We naturally ask that the efficiency of the filter does not depend on its previous use, i.e., that $(Y_i|X_i)_i$ is an independent family. To construct such an object, one may consider an i.i.d. family of uniformly distributed variables (ξ_i) on $(0, 1)$, independent from (X_i) , and set

$$Y_i = \mathbf{1}_{\xi_i \leq s(X_i)}. \quad (4.19)$$

This object clearly fulfills all properties mentioned and is general enough to capture also quite specific situations.

We will now establish our result in this more general setting. Consider a bounded and continuous map $f : \mathbf{R}^{p+1} \rightarrow \mathbf{R}$. We observe the $O\mathcal{U}$ -process $X(t)$ through the map f , so the observations are $Y_i = f(X_i, \dots, X_{i+p})$ where as before $X_i = X(t_i)$. To define a pseudo-likelihood, we reason as if the observed data can be regarded as stationary. We thus denote by f_u the law of a *stationary* $O\mathcal{U}$ -process with parameters u filtered via the measurement map f and we remind ourselves that the latent process $X(t)$ is not necessarily stationary.

We consider a likelihood of the kind

$$\mathcal{L}_N \propto \exp \left(\sum_{i=0}^{N-p} \log f_u(Y_i) \right), \quad (4.20)$$

normalized to mass one, and we denote the potential by

$$g_N(u) := N^{-1} \sum_{i=0}^{N-p} \log f_u(Y_i). \quad (4.21)$$

This setup simply corresponds to observing the data portion (X_i, \dots, X_{i+p}) via the filter f and, assuming stationarity, building a pseudo-likelihood. Unfortunately this does not directly include the intended case of a sigmoid filter, as this would rather involve measurements $h(\xi_i, X_i, \dots, \xi_{i+p}, X_{i+p})$ with (ξ_i) an i.i.d. sequence independent from (X_i) and uniform on $(0, 1)$. However, once the proposed case is examined, one can get to the latter case by exchanging the order of integration using the Fubini property, i.e., studying the behavior of $h(t_0, X_i, \dots, t_p, X_{i+p})$ and then integrate over (t_0, \dots, t_p) . For example, consider the case $p = 1$. Then the intended map h would be

$$h(t_0, x_0, t_1, x_1) = (\mathbf{1}_{t_0 \leq s(x_0)}, \mathbf{1}_{t_1 \leq s(x_1)}) \quad (4.22)$$

For a fixed u , f_u can take at most 4 values, e.g.,

$$f_u(1, 1) = \mathbb{E}[s(Z_0)s(Z_1)], \quad (4.23)$$

where (Z_0, Z_1) is a stationary $O\mathcal{U}$ -process of parameter u , and similarly for $f_u(0, 1)$, $f_u(1, 0)$, and $f_u(0, 0)$. As this technique would render the proof lengthy, we decided to put it aside.

In order to state our result, we need to specify some minimal set of regularity conditions. Convergence of the potential is required as well as definiteness in the Kullback-Leibler divergence (D_{KL}). Additionally, we shall also require a separation condition in the large data limit.

Assumption 4.2 (Regularity). We assume the following specifics:

- 1) The potential g_N converges uniformly on the compacts of u .
- 2) $D_{KL}(f_{u_0}, f_u) = 0 \implies u = u_0$.
- 3) There is a $\delta > 0$ and a compact neighborhood K of u_0 such that, for N large enough and for $u \in K^C$, $g_N(u_0) > g_N(u) + \delta$.

Theorem 4.5 (Weak convergence). *Under Assumption 4.2, we have the weak convergence*

$$\mathcal{L}_N \rightarrow \delta_{u_0}. \quad (4.24)$$

This theorem can be adapted to cover other situations:

- If $D_{KL}(f_{u_0}, f_u) = 0$ does not imply $u = u_0$, one can try to factorize the map through an equivalence relation (as in the proof of Theorem 4.2) to get the convergence towards the indicative function of the set $\{u; D_{KL}(f_{u_0}, f_u) = 0\}$.
- If we only want to consider parameters within a subset U of the set of parameters, one can always consider the assumptions restricted to the set U , and then one would have the convergence only for priors with support in U .

Proof. We divide the proof in two steps as follows.

Step 1 Let us call g the limit of (g_N) , which by assumption is not random. The same argument as in the proof of Theorem 3.2, the ergodicity of the process (X_t) , allows us to state that the limit function must be $\mathbb{E}[\log f_u]$ (up to a constant, the limit is actually $-D_{KL}(f_{u_0}, f_u)$). As adding a constant is equivalent to multiplying all posteriors by a nonnegative constant, we may assume the limit function to be equal to $u \mapsto -D_{KL}(f_{u_0}, f_u)$.

Step 2 As \mathcal{L}_N is proportional to $\exp(g_N)$, the assumptions allow us to deduce that the mass of any neighborhood of u_0 converges towards 1. This implies the weak convergence of the posteriors towards the Dirac δ_{u_0} .

4.4. Non-binary measurements

As a final variation on the theme we now show how Theorem 4.5 can be applied in such a way as to overcome the issues with the non-identifiability of Theorem 4.2. The idea is that the test with one filter c evidently at best gives us a curve containing the parameter u_0 , namely the one satisfying $\beta = \beta_0$, and $\sqrt{\gamma}(c - \alpha) = \sqrt{\gamma_0}(c - \alpha_0)$. Suppose instead that we have two kinds of tests, one with cut-off c_1 and one with cut-off c_2 . Since the intersection of the two curves implied by the respective filters c_1 and c_2 is exactly the point u_0 , it is natural to assume that two filters are enough to get the convergence of the posterior towards δ_{u_0} . We show that this is indeed the case.

Theorem 4.6 (Trinary filter). *Let $c_1 < c_2$ and consider the filtered values Y_i to be*

$$Y_i = \mathbf{1}_{X_i > c_1} + \mathbf{1}_{X_i > c_2}, \quad (4.25)$$

and the associated pseudo-likelihood

$$\mathcal{L}_N = \prod_{i=0}^{N-p} \varphi(Z_i, \dots, Z_{i+p}), \quad (4.26)$$

where φ is the multivariate cumulative distribution function of a Gaussian $\mathcal{N}(\alpha, \Sigma)$ (4.1). Then the sequence of pseudo-likelihoods converges weakly:

$$\mathcal{L}_N \rightarrow \delta_{u_0}. \quad (4.27)$$

Proof. It is sufficient to check that the sequence (Y_i) verifies Assumption 4.2. Considering $g_N := N^{-1} \log \mathcal{L}_N$, one has

$$g_N = \sum_{e \in \{0,1,2\}^p} \frac{N_e}{N} \log p_u(e), \quad (4.28)$$

where N_e is the number of times (Y_i, \dots, Y_{i+p}) matches e , and p_u is the cumulative distribution function of the underlying Gaussian process $\mathcal{N}(\alpha, \Sigma)$. From this we get directly conditions (4.2) and (4.2) of Assumption 4.2. For condition (4.2), if the Kullback-Leibler divergence is zero, then we must have $p_u(e) = p_{u_0}(e)$ for all e . This implies the equalities

$$\left. \begin{aligned} \beta &= \beta_0 \\ \sqrt{\gamma}(c_1 - \alpha) &= \sqrt{\gamma_0}(c_1 - \alpha_0) \\ \sqrt{\gamma}(c_2 - \alpha) &= \sqrt{\gamma_0}(c_2 - \alpha_0) \end{aligned} \right\} \quad (4.29)$$

As $c_1 \neq c_2$ we have $u = u_0$, and so the assumptions are fulfilled.

5. Illustrations

We devote this section to some illustrations of selected results from §§3 and 4. We shall do this in the intended epidemiological setting and thus no longer assume the $O\mathcal{U}$ -process, but rather the SIS- and SIS_E -models from §2 in the form of continuous-time Markov chains over a discrete state-space. In §5.1 we investigate the precision of the predicted posterior uncertainty under full state measurements, while in §5.2 we offer a demonstration of the singular behavior under filtered measurements. Finally, in §5.3 we highlight the use of synthetic data when approaching more realistic problems defined over a network.

The software for the numerical experiments is available for download via the corresponding author's web-page*.

5.1. Asymptotic uncertainty

We first consider the Bayesian uncertainty under accurate measurements and take the continuous-time Markov chain version of the SIS-model (2.4) as an example. Using the $SIS \leftrightarrow O\mathcal{U}$ approximate interpretation (2.9) we have from (3.1) the relations

$$\left. \begin{array}{l} \alpha_{OU} = \Sigma(1 - R_0^{-1}) \\ \beta_{OU} = \exp(-\gamma h(R_0 - 1)) \\ \gamma_{OU} = \Sigma^{-1} R_0 \cdot [1 - \beta_{OU}^2]^{-1} \end{array} \right\} \quad \left. \begin{array}{l} R_0 = 1 + \alpha_{OU}(1 - \beta_{OU}^2)\gamma_{OU} \\ \Sigma = \alpha_{OU} + [(1 - \beta_{OU}^2)\gamma_{OU}]^{-1} \\ \gamma = -\log \beta_{OU} \cdot [h(R_0 - 1)]^{-1} \end{array} \right\} \quad (5.1)$$

where $\{R_0, \Sigma, \gamma\}$ are the SIS-model parameters.

We generate synthetic data from the Markov chain as illustrated in Figure 1 for $\Sigma = 1000$ and with ranges of values $\gamma \in [0.1, 1]$ and $R_0 \in (1, 3.5]$. We let $I(0) = 0.01 \times \Sigma$ and sample exact values of $I(t_i)$ for $t_i = ih$, $i = 1, \dots, N$, and $h \equiv 1$. This corresponds to 100 perfect samples in a closed population at a rate equivalent to between one to one tenth the disease period unit ($= 1/\gamma$).

We evaluate the posterior over a grid of values in the (R_0, γ) -plane by simply normalizing the likelihood for the Markov chain given the synthetic data discussed previously. The likelihood of the Markov chain is formally obtained by solving the associated master equation which, however, is inconvenient to do except for small populations Σ . A more general approach is via a local linear Gaussian approximation and a Kalman filter. Put $I_0 = I(0)$ and define

$$\begin{aligned} I_{k+1} &= \left(1 + \Delta t \beta \Sigma^{-1} (\Sigma - I_k) - \Delta t \gamma\right) I_k + w_k, \\ w_k &\sim \mathcal{N}(0, [\Delta t \beta \Sigma^{-1} (\Sigma - I_k) + \Delta t \gamma] I_k), \end{aligned} \quad (5.2)$$

that is, this is the forward Euler discretization of the Langevin equations approximating the Markov chain. The Kalman filter associated with (5.2) computes a likelihood for each data point, albeit for a perturbed model. The relative error in the Langevin approximation generally scales with the inverse of the population size Σ , and can be expected to be rather small in the present context (see [45, Ch. 11.3]). Further, Proposition 2.1 suggests analyzing the Euler discretization via backward analysis as a parameter perturbation, but unfortunately this is not generalizable to non-additive noise [46]. For a resolved discretization, however, $\gamma \Delta t \ll 1$, weak first order convergence can be expected under broad conditions. We use the constant Kalman resolution $\Delta t = h/4$ and next focus on the estimation error.

*Refer to the BISDE-code at <https://user.it.uu.se/~stefane/freeware.html>

We have already evaluated the asymptotic covariance matrix under accurate data in (3.7). Using the linear uncertainty transformation $Q' \approx JQJ^T$, where J is the Jacobian of the parameter map (5.1), and where Q is the (diagonal) covariance matrix in (3.7), we can estimate the posterior variance

$$\text{Var}(R_0) \approx \frac{R_0^3}{\Sigma N} \times \frac{1 + \beta_{OU}}{1 - \beta_{OU}}. \quad (5.3)$$

A similar formula can be worked out for the variance of γ as well, although a bit more involved. For small enough h , the denominator $1 - \beta_{OU} \sim \gamma h(R_0 - 1)$ in (5.3), and so the *relative* uncertainty in any consistent estimator of R_0 can be expected to depend weakly on R_0 itself. This effect is seen for $R_0 > 1$ in Figure 2 (*top*), where it can also be seen that the approximation (5.3) derived from the OU -approximation is somewhat optimistic. Similarly, we find that the relative uncertainty of R_0 goes down with increasing values of γ , or, which by (5.1) is the same thing, with decreasing correlation β_{OU} (cf. Figure 2, *bottom*).

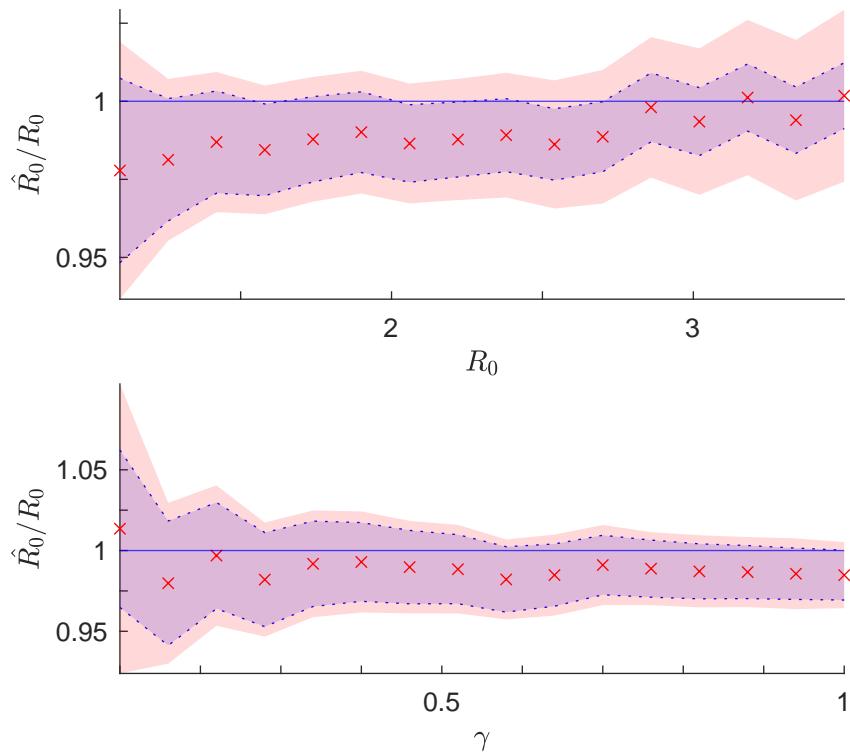


Figure 2. Marginal posterior uncertainty (± 2 SD) for the SIS-model and a range of parameters. *Top*: with $\gamma \equiv 1$ fixed, *bottom*: with $R_0 \equiv 1.5$ fixed. *Crosses*: MMSE-estimators \hat{R}_0 (i.e., posterior means), *dotted*: estimated uncertainty according to (5.3), *red*: posterior width (± 2 SD).

5.2. Limits of convergence

The SIS-model investigated previously was dependent on two parameters only and hence the singularity detected in §4 is not likely to be limiting any convergence. While various model modifications naturally lead to additional independent parameters, e.g., an extra transition $S \rightarrow I$ modeling external

infectious events, the most immediate modification is to simply consider the population size Σ uncertain. For instance, this is a possible setup for inference relying on sewage water analysis, where the data is binary according to whether the infectious substance is above or below some known threshold value c , but the sewage uptake area is populated with an unknown number of individuals.

We remark that this is a considerably challenging task, and although we are able to demonstrate the sharpness of our negative results from §4 in this setting, the fact that this problem can at all be approached is quite remarkable.

As ground truth we use the same parameters as in the previous section, but with $\gamma = 1/10$ (time $^{-1}$) and $R_0 = 1.5$ fixed, and we need to sample more, $N = 1000$ points equispaced with $h = 1$. The data is then taken to be the filtered sequence $Y_i = \mathbf{1}_{X_i \geq c}$ with $c = 0.9 \times I_\infty = 0.9 \times \Sigma(1 - R_0^{-1})$.

The measurement map is strongly nonlinear and so the Kalman filter needs to be extended in some way. We took an immediate approach by simply discretizing the state variable I into $M = 100$ Gaussian particles, distributed according to the percentiles of the stationary measure for a given proposed set of parameters. Each particle is evolved in time $[0, h]$ using steps of size Δt according to the Kalman filter (5.2), after which the prior distribution is formed by aggregating the probability mass in the vicinity of each particle. This yields the likelihood for a single data point after which the posterior distribution for the state is obtained by setting selected particles' mass to zero (according to the data point) and rescaling appropriately.

From Theorem 4.2 we have the singular curve $\sqrt{\gamma_{OU}}(c - \alpha_{OU}) = \text{constant}$, which gets transferred via the map (5.1) into a surface in (R_0, Σ, γ) -space. After arbitrarily fixing γ we thus obtain a curve in the (R_0, Σ) -plane. Since the SIS-to-OU map (2.9) is an approximation and, moreover, the likelihood is approximated via a Kalman filter, this analytical curve can be expected to be a perturbation of the observed numerical posterior level curves. As shown in Figure 3 the match is quite remarkable.

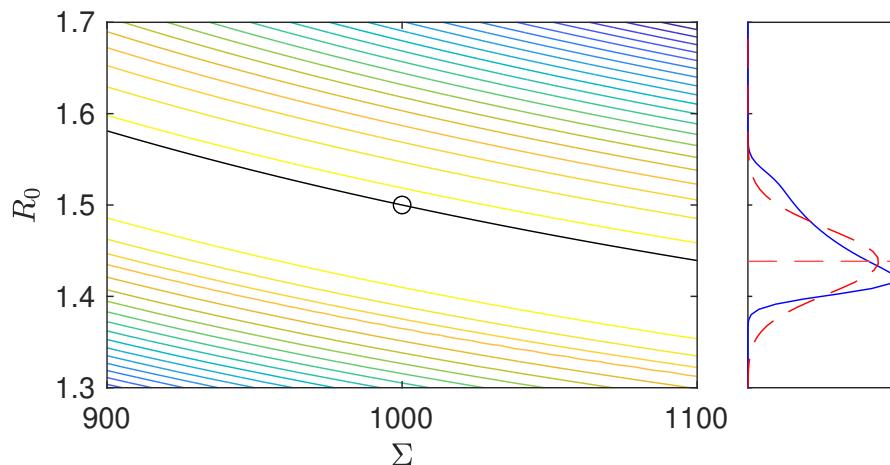


Figure 3. *Left:* (log-)posterior for the SIS-model under filtered data and conditioned on the true value $\gamma = 1/10$ (time $^{-1}$). Also indicated is the singular curve as predicted by theory passing through the true parameter generating the data (circle). *Right:* marginal distribution for R_0 together with a normal fit (dashed).

5.3. Network epidemics

Epidemic models on networks can give rise to phenomena not observed in single-node systems, e.g., the *rescue effect* [1], where the infection is “rescued” from extinction through the network structure. Here we consider both the SIS- and the SIS_E-model, respectively, cf. (2.2)–(2.3). We assume these models at each node in a network implicitly defined by pre-recorded movements of individuals between the nodes. As a concrete example this would be an appropriate model for estimating disease parameters in a monitoring program for bovine animals using cheap, but low-informative, tests collected on a weekly basis for a subsample of *sentinel* nodes.

The nodal model is replicated across 1600 nodes, populated with 196,168 individuals, and the nodes are connected using 466,692 prescribed movements of individuals over four years, see Figure 4. The system is not well stirred on the aggregate level, but events occur frequently; the average # of events per sample node and day = 0.20 [0.19, 0.21] with 50% credible interval (CrI). This particular network was constructed by anonymizing a set of recorded cattle movements and can be accessed through the publicly available R-package SimInf [19, 47]. We extract model measurements from the same 100 randomly pre-selected sentinel nodes every 7th day for a total of 4 years. Each measurement is the outcome of a binary test: if the prevalence ($P = \Sigma^{-1}I$) in the node is above a threshold value ($c = 30\%$ or 4% for the SIS- or the SIS_E-model, respectively), where all nodes are seeded at 10% or 2% initially. In Figure 5, we illustrate the sampling output in time for the SIS_E-model.

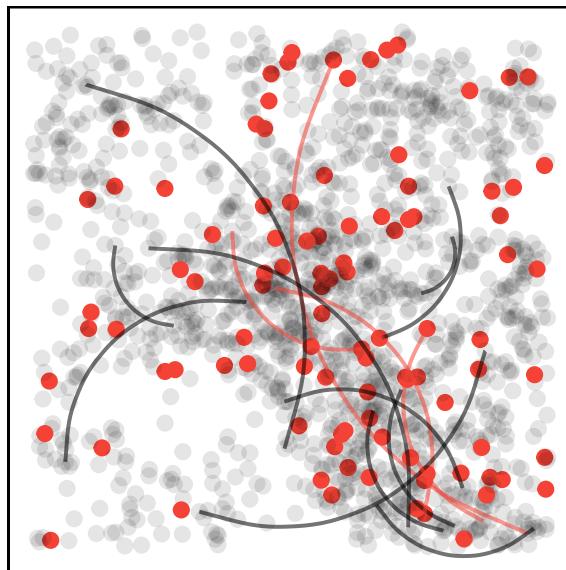


Figure 4. Illustration of the transport network; the red points are the sentinel nodes, and the grey points are the latent ones. Red/black lines are transport events into a sentinel- or latent node, respectively.

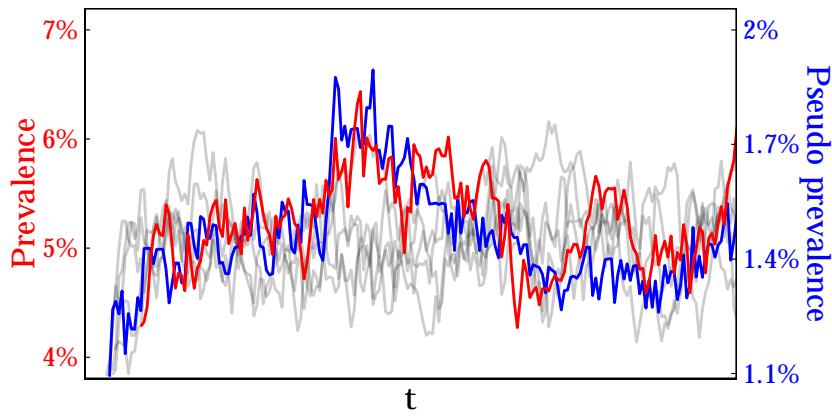


Figure 5. The population-weighted average prevalence (red) is unobservable, but the pseudo prevalence (blue) is obtained from weighting together multiple binary measurements. The least squares $O\mathcal{U}$ -fit for the pseudo prevalence is used as summary statistics (a few samples in grey are shown).

A challenging aspect of many data-driven inference problems, e.g., including network dynamics is that the likelihood function is intractable and must be estimated through repeated simulations. Bayesian inference in this setting is termed Likelihood-free inference (LFI), or Approximate Bayesian Computations (ABC); see [48, 49] for reviews. In this example, we consider the Sequential Monte Carlo (SMC) adaptation of ABC (SMC-ABC) implemented in SimInf and described in [50].

Our SMC-ABC implementation determines proposal rejections per generation n using the normalized Euclidean kernel $K_\varepsilon(x, y) = \sqrt{\sum_i((x_i - y_i)/x_i)^2} < \varepsilon_n$ for statistics of the simulation proposal y and observation x and with a series of decreasing tolerances ε_n . The statistics are computed as follows. Each simulation generates a time series of *pseudo prevalences*, i.e., a population-weighted sum of positive samples. We interpret the time series as an $O\mathcal{U}$ -process and select the least square estimates (3.3)–(3.4) as *indirect* summary statistics [51]. A word in favor of this particular choice of statistics for other ABC implementations, e.g., synthetic likelihoods [52], is that least square estimates are asymptotically normally distributed under broad assumptions [53]. Notably, we get away with using statistics with one more or equal dimension as the parameter set, suggesting that this characterization is indeed very fitting.

For the inference we use a single initial simulation with the true parameters $(\beta, \gamma, R_0)_{SIS} = (0.16, 0.1, = 1.6)$ and $(\beta, \gamma, \rho, R_0)_{SIS_E} = (0.054, 0.1, 0.44, = 1.108)$, respectively, and we infer all parameters simultaneously. For priors, since we have no likelihood and thus cannot easily produce a strictly non-informative prior, we take uniform distributions over quite large intervals in parameter space: β and $\gamma \sim \mathcal{U}(0, 1)$ in both cases and $\rho \sim \mathcal{U}(0.4, 0.5)$. We use the decreasing ABC tolerances $\varepsilon_n = 100 \exp(-0.25(n - 1))$, for $n = 1 \dots 15$ and 1000 SMC particles.

We found that ρ requires a tighter prior than the others for the computations to complete in a reasonable time: the SIS_E-model is considerably more challenging but is also more realistic, particularly so in our setting on a network where rather large prevalences are required for the SIS-model to not simply die out.

The results for the R_0 -marginals are displayed in Figure 6, where we also investigate the concentration effect of data through the relative change in quartile coefficient of dispersion (QCD); a small

concentration factor indicates an accurately identifiable parameter. Although the SIS_E-model is clearly more challenging, R_0 is still well reconstructed for both models.

Binary data implies identifiability when given multiple observations at the same time, e.g., over a small collection of nodes in a network rather than on a single node, as indicated in a qualitative sense by Theorem 4.6. Additionally, R_0 is identifiable with quite high accuracy even when the dependent parameters covaries. The example is prototypical of using synthetic data to evaluate the feasibility of an intended setup. Since the posteriors are robust with respect to capturing the synthetic truth, we have good reasons to also have some faith in the design when approaching real data.

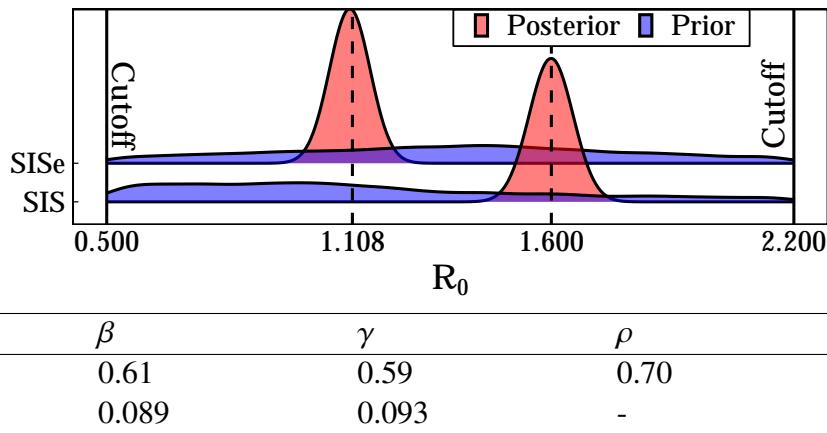


Figure 6. Top: posterior and prior distributions for R_0 with the true values indicated. Bottom: recorded QCD concentration factors for all the parameters (see text for details).

6. Conclusions

Throughout this work we have employed the Ornstein-Uhlenbeck process as a meta-model of more involved epidemiological models. We indicated in §3 a convergence analysis of the Bayesian posterior under direct process observations. Since this is an unrealistic setup in most epidemiological applications, one can think of these results as *best possible*.

We next took the opposite standpoint and considered data to be severely filtered such that, literally, each data point contributed only a single bit of information. For instance, this could be a model of pooled data obtained through environmental sampling and subsequent analysis. To obtain a closed framework we added a fairly general stationary p -step Markov assumption and worked out conditions on the data to obtain a non-singular inverse problem.

With increasing compute power and improving possibilities for gathering data, fully Bayesian first-principle epidemiological models can be realized. As a minimum standard, we propose, such methods should be preceded by a proof of self-consistency: data generated from the model itself and chosen “nearby” the actual data should allow for accurate parameter identifiability. Under this basic standard, Bayesian epidemiological modeling with both short-term prediction and generation of forecasting scenarios can be included as an integrated part of the public health’s methodological arsenal.

Acknowledgments

This work was financially supported by the ENS Paris (S. Bronstein) and in part by the Swedish Research Council Formas (S. Engblom) and the Swedish Innovation Agency Vinnova (S. Engblom, R. Marin). Several helpful comments and kind suggestions on earlier versions of the manuscript were offered by friends and colleagues. These helped to improve the presentation considerably.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. M. J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, 2011. <https://doi.org/10.1086/591197>
2. T. McKinley, A. R. Cook, R. Deardon, Inference in epidemic models without likelihoods, *Int. J. Biostat.*, **5** (2009). <https://doi.org/10.2202/1557-4679.1171>
3. H. Andersson, T. Britton, *Stochastic Epidemic Models and Their Statistical Analysis*, Springer Science & Business Media, **151** (2012). <https://doi.org/10.1007/978-1-4612-1158-7>
4. S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, et al., Modelling disease outbreaks in realistic urban social networks, *Nature*, **429** (2004), 180–184. <https://doi.org/10.1038/nature02541>
5. N. M. Ferguson, D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, et al., Strategies for containing an emerging influenza pandemic in southeast Asia, *Nature*, **437** (2005), 209–214. <https://doi.org/10.1038/nature04017>
6. D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, A. Vespignani, Multiscale mobility networks and the spatial spreading of infectious diseases, *Proc. Natl. Acad. Sci. USA*, **106** (2009), 21484–21489. <https://doi.org/10.1073/pnas.0906910106>
7. S. Merler, M. Ajelli, A. Pugliese, N. M. Ferguson, Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: implications for real-time modelling, *PLoS Comput. Biol.*, **7** (2011), e1002205. <https://doi.org/10.1371/journal.pcbi.1002205>
8. E. Brooks-Pollock, G. O. Roberts, M. J. Keeling, A dynamic model of bovine tuberculosis spread and control in Great Britain, *Nature*, **511** (2014), 228–231. <https://doi.org/10.1038/nature13529>
9. J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, et al., Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees, *BMC Med.*, **9** (2011), 87. <https://doi.org/10.1186/1741-7015-9-87>
10. P. Bajardi, A. Barrat, L. Savini, V. Colizza, Optimizing surveillance for livestock disease spreading through animal movements, *J. R. Soc. Interface*, **9** (2012), 2814–2825. <https://doi.org/10.1186/1741-7015-9-87>

11. M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, J. H. Jones, A high-resolution human contact network for infectious disease transmission, *Proc. Natl. Acad. Sci. USA*, **107** (2010), 22020–22025. <https://doi.org/10.1073/pnas.1009094108>
12. T. Obadia, R. Silhol, L. Opatowski, L. Temime, J. Legrand, A. C. M. Thiébaut, et al., Detailed contact data and the dissemination of *Staphylococcus aureus* in hospitals, *PLoS Comput. Biol.*, **11** (2015), e1004170. <https://doi.org/10.1371/journal.pcbi.1004170>
13. D. J. Toth, M. Leecaster, W. B. Pettey, A. V. Gundlapalli, H. Gao, J. J. Rainey, et al., The role of heterogeneity in contact timing and duration in network models of influenza spread in schools, *J. R. Soc. Interface*, **12** (2015), 20150279. <https://doi.org/10.1098/rsif.2015.0279>
14. Q. Zhang, K. Sun, M. Chinazzi, A. P. y Piontti, N. E. Dean, D. P. Rojas, et al., Spread of Zika virus in the Americas, *Proc. Natl. Acad. Sci. USA*, **114** (2017), E4334–E4343. <https://doi.org/10.1073/pnas.1620161114>
15. Q. H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, A. Vespignani, Measurability of the epidemic reproduction number in data-driven contact networks, *Proc. Natl. Acad. Sci. USA*, **115** (2018), 12680–12685. <https://doi.org/10.1073/pnas.1811115115>
16. S. Widgren, S. Engblom, U. Emanuelson, A. Lindberg, Spatio-temporal modelling of verotoxigenic *Escherichia coli* O157 in cattle in Sweden: exploring options for control, *Vet. Res.*, **49** (2018). <https://doi.org/10.1186/s13567-018-0574-2>
17. T. Söderström, P. Stoica, *System Identification*, Prentice-Hall International, 1989.
18. G. Fournié, A. Waret-Szkuta, A. Camacho, L. M. Yigezu, D. U. Pfeiffer, F. Roger, A dynamic model of transmission and elimination of peste des petits ruminants in Ethiopia, *Proc. Natl. Acad. Sci. USA*, **115** (2018), 8454–8459. <https://doi.org/10.1073/pnas.1711646115>
19. S. Engblom, R. Eriksson, S. Widgren, Bayesian epidemiological modeling over high-resolution network data, *Epidemics*, **32** (2020), 100399. <https://doi.org/10.1016/j.epidem.2020.100399>
20. X. Shen, L. Wasserman, Rates of convergence of posterior distributions, *Ann. Stat.*, **29** (2001), 687–714. <https://doi.org/10.1214/aos/1009210686>
21. A. M. Stuart, Inverse problems: a Bayesian perspective, *Acta Numer.*, **19** (2010), 451–559. <https://doi.org/10.1017/S0962492910000061>
22. A. Stuart, A. Teckentrup, Posterior consistency for Gaussian process approximations of Bayesian posterior distributions, *Math. Comput.*, **87** (2018), 721–753. <https://doi.org/10.1090/mcom/3244>
23. J. Latz, On the well-posedness of Bayesian inverse problems, *SIAM/ASA J. Uncert. Quant.*, **8** (2020), 451–482. <https://doi.org/10.1137/19M1247176>
24. H. Owhadi, C. Scovel, T. Sullivan, On the brittleness of Bayesian inference, *SIAM Rev.*, **57** (2015), 566–582. <https://doi.org/10.1137/130938633>
25. B. Sprungk, On the local Lipschitz stability of Bayesian inverse problems, *Inverse Probl.*, **36** (2020), 055015. <https://doi.org/10.1088/1361-6420/ab6f43>

26. A. Galani, R. Aalizadeh, M. Kostakis, A. Markou, N. Alygizakis, T. Lytras, et al., SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions, *Sci. Total Environ.*, **804** (2022), 150151. <https://doi.org/10.1016/j.scitotenv.2021.150151>

27. B. Kennedy, H. Fitipaldi, U. Hammar, M. Maziarz, N. Tsereteli, N. Oskolkov, et al., App-based COVID-19 syndromic surveillance and prediction of hospital admissions in COVID Symptom Study Sweden, *Nat. Commun.*, **13** (2022), 2110. <https://doi.org/10.1038/s41467-022-29608-7>

28. W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. Lond. A*, **115** (1927), 700–721. <https://doi.org/10.1098/rspa.1927.0118>

29. S. Engblom, S. Widgren, Data-driven computational disease spread modeling: from measurement to parametrization and control, in *Disease Modeling and Public Health: Part A, Handbook of Statistics, Chapter 11* (eds. C. R. Rao, A. S. Rao, S. Payne), Elsevier, Amsterdam, **36** (2017), 305–328. <https://doi.org/10.1016/bs.host.2017.05.005>

30. T. Britton, Stochastic epidemic models: a survey, *Math. Biosci.*, **225** (2010), 24–35. <https://doi.org/10.1016/j.mbs.2010.01.006>

31. A. Gray, D. Greenhalgh, L. Hu, X. Mao, J. Pan, A stochastic differential equation SIS epidemic model, *SIAM J. Appl. Math.*, **71** (2011), 876–902. <https://doi.org/10.1137/10081856X>

32. G. E. Uhlenbeck, L. S. Ornstein, On the theory of the Brownian motion, *Phys. Rev.*, **36** (1930), 823. <https://doi.org/10.1103/PhysRev.36.823>

33. I. Shoji, Approximation of continuous time stochastic processes by a local linearization method, *Math. Comput.*, **67** (1998), 287–298. <https://doi.org/10.1090/S0025-5718-98-00888-6>

34. A. P. Ghosh, W. Qin, A. Roitershtein, Discrete-time Ornstein-Uhlenbeck process in a stationary dynamic environment, *J. Interdiscip. Math.*, **19** (2016), 1–35. <https://doi.org/10.1080/09720502.2013.857921>

35. I. V. Basawa, B. P. Rao, Chapter 10 - Bayesian Inference for Stochastic Processes, in *Statistical Inference for Stochastic Processes* (eds. I. V. Basawa, B. P. Rao), Academic Press, London, **1980** (1980), 255–293. <https://doi.org/10.1016/B978-0-12-080250-0.50017-8>

36. M. Mishra, B. Prakash Rao, Rate of convergence in the Bernstein-Von Mises theorem for a class of diffusion processes, *Stochastics*, **22** (1987), 59–75. <https://doi.org/10.1080/17442508708833467>

37. J. Bishwal, Rates of convergence of the posterior distributions and the Bayes estimations in the Ornstein-Uhlenbeck process, *Random Oper. Stoch. Equ.*, **8** (2000), 51–70. <https://doi.org/10.1515/rose.2000.8.1.51>

38. D. Florens-Zmirou, Approximate discrete-time schemes for statistics of diffusion processes, *Statistics: J. Theor. Appl. Stat.*, **20** (1989), 547–557. <https://doi.org/10.1080/02331888908802205>

39. V. Genon-Catalot, Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, **21** (1990), 99–116. <https://doi.org/10.1080/02331889008802231>

40. M. Kessler, Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.*, **24** (1997), 211–229. <https://doi.org/10.1111/1467-9469.00059>

41. K. B. Athreya, S. N. Lahiri, *Measure Theory and Probability Theory*, Springer Science & Business Media, 2006. <https://doi.org/10.1007/978-0-387-35434-7>

42. E. A. Stoltzenberg, N. L. Hjort, Models and inference for on-off data via clipped Ornstein–Uhlenbeck processes, *Scand. J. Stat.*, **48** (2021), 908–929. <https://doi.org/10.1111/sjos.12472>

43. E. Slud, Clipped Gaussian processes are never M-step Markov, *J. Multivariate Anal.*, **29** (1989), 1–14. [https://doi.org/10.1016/0047-259X\(89\)90072-9](https://doi.org/10.1016/0047-259X(89)90072-9)

44. K. Joag-Dev, M. D. Perlman, L. D. Pitt, Association of normal random variables and Slepian’s inequality, *Ann. Probab.*, **11** (1983), 451–455. <https://doi.org/10.1214/aop/1176993610>

45. S. N. Ethier, T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1986. <https://doi.org/10.1002/9780470316658>

46. T. Shardlow, Modified equations for stochastic differential equations, *BIT Numer. Math.*, **46** (2006), 111–125. <https://doi.org/10.1007/s10543-005-0041-0>

47. S. Widgren, P. Bauer, R. Eriksson, S. Engblom, SimInf: An R package for data-driven stochastic disease spread simulations, *J. Stat. Software*, **91** (2019), 1–42. <https://doi.org/10.18637/jss.v091.i12>

48. J. M. Marin, P. Pudlo, C. P. Robert, R. J. Ryder, Approximate Bayesian computational methods, *Stat. Comput.*, **22** (2012), 1167–1180. <https://doi.org/10.1007/s11222-011-9288-2>

49. S. A. Sisson, Y. Fan, M. Beaumont, *Handbook of Approximate Bayesian Computation*, CRC Press, 2018. <https://doi.org/10.1201/9781315117195>

50. T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M. P. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, *J. R. Soc. Interface*, **6** (2009), 187–202. <https://doi.org/10.1098/rsif.2008.0172>

51. C. C. Drovandi, A. N. Pettitt, M. J. Faddy, Approximate Bayesian computation using indirect inference, *J. R. Stat. Soc.: Ser. C (Appl. Stat.)*, **60** (2011), 317–337. <https://doi.org/10.1111/j.1467-9876.2010.00747.x>

52. S. N. Wood, Statistical inference for noisy nonlinear ecological dynamic systems, *Nature*, **466** (2010), 1102. <https://doi.org/10.1038/nature09319>

53. W. K. Newey, D. McFadden, Chapter 36 Large sample estimation and hypothesis testing, *Handb. Econom.*, **4** (1994), 2111–2245. [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4)

