Plug-and-play Bayesian inference for compartmental models in PLOM

Joseph Dureau, Sébastien Ballesteros

July 7, 2013

1 Introduction

2 Compartmental models

2.1 Definition

Compartmental models are a general framework used to represent the state of a countable population (of humans, animals, molecules, etc) and its evolution. At a given time t, the population is described by the number of individuals in each of n possible states: the ensemble of individuals in a same state defines what is termed as a compartment. Each individual belongs to one and only one compartment. Individuals within a same compartment are considered indistinguishable. We will consider in this document that n is known, fixed, and finite.

Each compartment can correspond to very diverse characterisations, depending on the context. They can be used to track the status of an individual with regards to a given disease in a human or animal population (susceptible or infected, for example), their age or their geographical location. Additionnally, compartments can be used to track the number of specimens of different animal species in an ecosystem, as in the Lotka-Volterra predator-prey model. They can also be used in physics and chemistry to determine molecule types, their electronic charge or radioactive state, for example. Less classical illustrations of the use of compartmental models include tracking the spread of rumors among a population, or the propagation of economic difficulties among countries following a financial crisis.

We note $z_t^{(i)}$ the size of compartment i ($1 \leq i \leq c$) at time t, and $z_t = [z_t^{(1)}, .., z_t^{(c)}]$. A model is defined by a (finite) number m of transformations of the system called reactions (the ensemble of all indexes is noted $\mathcal{R}$). These reactions correspond to one or several individuals passing from one compartment to another, or arriving or leaving the total population. In any case, each reaction k is characterised by its effect on the structure of the population corresponding to a vector $l^{(k)} \in \mathbb{Z}^c$, and its intensity of occurrence. In the remainder of this document, we will make the classic assumption that the probability of occurrence of each reaction is proportional to the number of individuals in a given compartment. The latter can be specified through a mapping $\chi : \mathcal{R} \to [1; c]$ such that the transition rate of reaction k can be written $r_t^{(k)}(z_t, \theta) z_t^{\chi(k)}$. This assumption implies the density dependance of transition rates, i.e. the transition rates of the model where the state variable has been normalised ($\dot{z}_t = z_t/N$) can be simply written as $r_t^{(k)}(z_t, \theta) \dot{z}_t^{\chi(k)}$.

We allow for these rates to depend on time, in order to reflect the influence of potential forcing of external factors on the system, and to depend on a finite set of constant quantities gathered in a parameter vector θ. In the remaining of this document, we will define compartmental models using the following formalism:

Reaction	**Effect**	**Rate**
reaction 1	$z_t \to z_t + l^{(1)}$	$r^{(1)}(z_t, \theta)$
...	...	...
reaction k	$z_t \to z_t + l^{(k)}$	$r^{(k)}(z_t, \theta)$
...	...	...
reaction m	$z_t \to z_t + l^{(m)}$	$r^{(m)}(z_t, \theta)$

Formally, this framework leads to the definition of a Markovian jump process, which dynamic can be expressed in the following way:

Markovian jump process compartmental model

$$\begin{aligned} P(z_{t+dt} = z_t + l^{(k)} | z_t) &= r_t^{(k)}(z_t, \theta) z_t^{\chi(k)} dt + o(dt) \qquad \text{for any } k \in \mathcal{R} \\ P(z_{t+dt} = z_t | z_t) &= \big(1 - \sum_{k \in \mathcal{R}} r_t^{(k)}(z_t, \theta) z_t^{\chi(k)} dt\big) + o(dt) \end{aligned} \tag{1}$$

Under some regularity conditions detailed in Ethier and Kurtz (1986), Fuchs (2013) or Guy et al. (2013), and due to the density-dependance of transition rates, the dynamic of the system converges to a deterministic behaviour as the population size tends to infinity. For finite populations, the additional stochastic behaviour is termed *demographic* stochasticity.

2.2 Environmental stochasticity

Extensions of the compartmental modeling framework introduced in the previous section have been proposed by the authors of Breto et al. (2009), to account for additional sources of uncertainty related to fluctuations in extrinsic determinants of the epidemic. This uncertainty is reflected through additional sources of stochasticity, termed *environmental* stochasticity. The approach suggested in Breto et al. (2009) is to consider stochastic transition rates $\tilde{r}_t^{(k)}$ for a subset $\mathcal{R}^e$, under the following constraints for all t:

$$\begin{aligned} \mathbb{E}\big(\tilde{r}_t^{(k)}\big) &\sim r_t^{(k)} \\ \tilde{r}_t^{(k)} &\geq 0 \end{aligned}$$

However, uncertain variations of extrinsic factors cannot always be modeled through high-frequency independent fluctuations. The evolution of climate, for example, has been shown to exhibit complex seasonal and inter-annual variations that influence epidemic dynamics Viboud et al. (2004). Following the work of Cazelles and Chau (1997) and Cori et al. (2009), the authors of Dureau et al. (2013) have proposed a general inferential framework for time-varying parameters, that is extended in the present document. Under this approach, parameters are modeled through stochastic differential equations or extensions thereof. The state vector is extended with additional components $x_t^{\theta_t}$ which dynamic is determined by the following equation:

$$dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + L^{\theta_t} dB_t^{Q^{\theta_t}} \tag{2}$$

Note that some constraints as positivity or boundedness generally need to be preserved when allowing parameters to vary over time, which is achieved by defining x^{θ_t} as the respectively the log or logit transformation of the quantity of interest.

2.3 Examples

2.4 Tractable approximations of compartmental models

2.4.1 Ordinary differential equations

The simplest and most stringent approximation of compartmental models are ordinary differential equations (ode's):

$$\frac{dz_t}{dt} = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} \tag{3}$$

Under this formalism, the number of individuals in each compartment takes continues values, and varies continuously (and in a differentiable manner) over time. More specifically, all kind of demographic or environmental stochasticity are neglected, leading the state of the system to evolve deterministically. From a practical perspective, the use of ordinary differential equations drastically simplifies the process of Bayesian inference, mainly due to the deterministic one-to-one mapping between trajectories $z_{0:T}$ and parameters θ.

This formalism can be legitimately used for large populations and when all significant environmental factors have been explicitly incorporated in the deterministic skeleton of the model. However, in alternative cases results should be treated with caution, and the use of alternative formalisms accounting for demographic or environmental stochasticity may be required.

2.4.2 Stochastic differential equations

Stochastic differential equations (sde's) are a natural extension of ode's, wherein state variables still take continuous values over time, and evolve continuously over time, but trajectories of the system are no longer deterministic and differentiable due to the introduction of a driving Brownian motion reflecting the stochasticity of the system. To introduce this formalism, we rely on the notations used by the author of Särkkä (2006) that will be helpfull to handle and represent different and independent sources of stochasticity:

$$dx_t = \mu_t(x_t, \theta)dt + L dB_t^{Q_t} \tag{4}$$

In this equation, μ_t is referred to as the drift, L as the dispersion matrix, and Q_t as the diffusion matrix of the driving Browian motion. In particular, in our models the state variable x_t is built from the concatenation of z_t and $x_t^{\theta_t}$ respectively corresponding to the variables describing the structure of the population and to the variables monitoring the

evolution of diffusing parameters over time. We can reformulate 4, utilising the notations that have been introduced earlier in this document:

$$
\begin{aligned}
dz_t &= \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt + L dB_t^{Q_t} \\
dx_t^{\theta_t} &= \mu^{\theta_t}(x_t^{\theta_t}, \theta) dt + L^{\theta_t} dB_t^{Q^{\theta_t}}
\end{aligned}
$$

Note that the deterministic skeleton of the population variables' dynamic correspond to the ode model introduced in the previous section. The dispersion matrix Q_t is a square matrix of size $n_{Q_t} \times n_{Q_t}$, and L is a rectangular matrix of size $c \times n_{Q_t}$. Let us illustrate the use of these objects by introducing how demographic stochasticity can be incorporated in the model, based on the diffusion approximation, and further how the white noise environmental stochasticity can be incorporated.

Diffusion approximation of the demographic stochasticity

In order to provide an SDE approximation of the demographic stochasticity, we rely on theoretical results on state-dependent Markov jump processes presented in Ethier and Kurtz (1986). The adaptation of these results on compartmental epidemic models models has been illustrated in Fuchs (2013). Extensions of these results in non-homogeneous settings are provided in Guy et al. (2013).

The diffusion approximation builds up on the definition of jump process models through their master equation:

$$
\frac{\partial}{\partial t} P(z_t) = \sum_{k \in \mathcal{R}} r^{(k)} \tilde{z}_{k,t}^{\chi(k)} P(z_t - l^{(k)}) - \sum_{k \in \mathcal{R}} r^{(k)}(z_t, \theta) z_t^{\chi(k)} P(z_t) \tag{5}
$$

Where $\tilde{z}_{k,t} = z_t - l^{(k)}$. The first term corresponds to the probability for the state vector of evolving into z_t, and the second corresponds to the probability of leaving the state z_t. In a SIR model setting, the master equation becomes:

$$
\begin{aligned}
\frac{\partial}{\partial t} P(S_t, I_t, R_t) = & \beta \frac{(S_t + 1)}{N} (I_t - 1) P(S_t + 1, I_t - 1, R_t) \\
& + \gamma (I_t + 1) P(S_t, I_t + 1, R_t - 1) \\
& - \beta \frac{S_t}{N} I_t P(S_t, I_t, R_t) \\
& - \gamma I_t P(S_t, I_t, R_t)
\end{aligned} \tag{6}
$$

This equation can be written in terms of normalised quantities, with $\varepsilon = 1/N$:

$$
\begin{aligned}
\frac{\partial}{\partial t} P(s_t, i_t, r_t) = & \frac{1}{\varepsilon} \beta (s_t + \varepsilon)(i_t - \varepsilon) P(s_t + \varepsilon, i_t - \varepsilon, , r_t) \\
& + \frac{1}{\varepsilon} \gamma (i_t + \varepsilon) P(s_t, i_t + \varepsilon, r_t - \varepsilon) \\
& - \frac{1}{\varepsilon} \beta s_t i_t P(s_t, i_t, r_t) \\
& - \frac{1}{\varepsilon} \gamma i_t P(s_t, i_t, r_t)
\end{aligned} \tag{7}
$$

The diffusion approximation relies on the limit of this expression when $\varepsilon \to 0$ while N is kept constant. The author of Fuchs (2013) shows that in this case, the former master equation converges to the following partial differential equation:

$$\begin{aligned}\frac{\partial}{\partial t}P(s_t,i_t,r_t) = &\frac{\partial}{\partial s}\beta s_t i_t P(s_t,i_t,r_t) - \frac{\partial}{\partial i}(\beta s_t i_t - \gamma i_t)P(s_t,i_t,r_t)\\ &+ \frac{1}{2}\frac{\partial^2}{\partial s^2}\frac{1}{N}\beta s_t i_t P(s_t,i_t,r_t)\\ &- \frac{1}{2}\frac{\partial^2}{\partial i^2}\frac{1}{N}(\beta s_t i_t - \gamma i_t)P(s_t,i_t,r_t)\\ &- \frac{\partial^2}{\partial s \partial i}\frac{1}{N}\beta s_t i_t P(s_t,i_t,r_t),\end{aligned} \tag{8}$$

which is equivalent to

$$\frac{\partial}{\partial t}P(s_t,i_t,r_t) = -\frac{\partial}{\partial x}[\dot{A}(s_t,i_t,r_t)P(s_t,i_t,r_t)] + \frac{1}{2}\frac{\partial}{\partial x}\frac{\partial}{\partial x}[\dot{\Sigma}(s_t,i_t,r_t)P(s_t,i_t,r_t)] \tag{9}$$

Where

$$\dot{A}(s_t,i_t,r_t) = \begin{pmatrix} -\beta s_t i_t \\ \beta s_t i_t - \gamma i_t \\ \gamma i_t \end{pmatrix} \tag{10}$$

and

$$\dot{\Sigma}(s_t,i_t,r_t) = \frac{1}{N}\begin{pmatrix} \beta s_t i_t & -\beta s_t i_t & 0 \\ -\beta s_t i_t & \beta s_t i_t + \gamma i_t & -\gamma i_t \\ 0 & -\gamma i_t & \gamma i_t \end{pmatrix} \tag{11}$$

Following Kloeden and Platen (1999), 9 is a Fokker-Planck equation corresponding to a diffusion process that is a solution of

$$d\dot{z}_t = \dot{A}(\dot{z}_t)dt + LdB_t^{\dot{Q}_t^d} \tag{12}$$

Here, we follow the formalism of Särkkä (2006) where $dB_t^{\dot{Q}_t^d}$ is a Brownian motion with diffusion matrix $\dot{Q}_t^d$ and L is a stoichiometric dispersion matrix such that $L\dot{Q}_t^d L = \dot{\Sigma}$:

$$\dot{Q}^d(s_t,i_t) = \frac{1}{N}\begin{pmatrix} \beta s_t i_t & 0 \\ 0 & \gamma i_t \end{pmatrix} \quad and \quad L = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \tag{13}$$

Equation 12 can be transposed in the natural scale of $z_t = [S_t, I_t, R_t]^T$, with $A = N\dot{A}$ and $Q^d = N^2\dot{Q}^d$:

$$dz_t = A(z_t)dt + LdB_t^{Q^d} \tag{14}$$

This result can be generalised based on the density-dependance property of rates $(r^{(k)}z_t^{\chi(k)})_{1\leq k\leq n}$. Formal proofs for the general case of density-dependent jump processes can be found in Ethier and Kurtz (1986). The authors demonstrate that the dynamic of a

density-dependent Markov jump process can be approximated with equation 14 with dBt being a multivariate Brownian motion with diffusion matrix $Q^d = diag\{r^{(k)} z_t^{\chi(k)}, k \in \mathcal{R}\}$, and L being the rectangular stoichiometric matrix which columns are the stoichiometric vectors $l^{(k)}$ with $k \in \mathcal{R}$. Additionally, the drift component $A(t)$ is determined by:

$$A(z_t) = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} \tag{15}$$

Lastly, the resulting expression for Σ is the following:

$$\Sigma(z_t) = LQ^d L' = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} l^{(k)\prime} \tag{16}$$

Diffusion approximation of the environmental stochasticty

This section focuses on environmental stochasticity. In this perspective, we consider an infinite population leading to a deterministic behaviour in the absence of environmental stochasticity or time-varying parameters following a diffusion:

$$dz_t = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt \tag{17}$$

In the case of the SIR model:

$$\begin{cases} dS_t = -\beta S_t \dfrac{I_t}{N} dt \\ dI_t = (\beta S_t \dfrac{I_t}{N} - \gamma I_t) dt \\ dR_t = \gamma I_t dt \end{cases} \tag{18}$$

The framework proposed in Breto et al. (2009) introduces environmental stochasticity by replacing deterministic time increments dt by stationnary and nonnegative increments $d\Gamma_t$ with mean dt and variance $\sigma^2 dt$. Here, if environmental noise is put over the infection reaction:

$$\begin{cases} dS_t = -\beta S_t \dfrac{I_t}{N} d\Gamma_t \\ dI_t = \beta S_t \dfrac{I_t}{N} d\Gamma_t - \gamma I_t dt \\ dR_t = \gamma I_t dt \end{cases} \tag{19}$$

We propose to derive a Gaussian formulation of epidemic models with environmental stochasticity by approximating $d\Gamma_t$ as $dt + \sigma dB_t$, i.e. the Gamma-distributed increments are replaced with a deterministic drift and a Brownian motion term with corresponding mean and variance. Thus, the model can be written as a stochastic differential equation:

$$\begin{cases} dS_t = -\beta S_t \dfrac{I_t}{N} dt - \sigma \beta S_t \dfrac{I_t}{N} dB_t^{(1)} \\ dI_t = (\beta S_t \dfrac{I_t}{N} d\Gamma_t - \gamma I_t) dt + \sigma \beta S_t \dfrac{I_t}{N} dB_t^{(1)} \\ dR_t = \gamma I_t dt \end{cases} \tag{20}$$

In the general case, independent environmental noise can be enforced upon any subset $\mathcal{R}^e \in \mathcal{R}$ of all reactions:

$$dz_t = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt + L^e dB_t^{Q^e} \tag{21}$$

L^e is the $c \times Card(\mathcal{R}^e)$ stoichiometric matrix which columns are the stoichiometric vectors $l^{(k)}$ with $k \in \mathcal{R}^e$. In addition, if all white noises are independent $dB_t^{Q^e}$ is a Brownian motion with diffusion matrix $Q^e = diag\big\{\big(\sigma^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)}\big)^2, k \in \mathcal{R}^e\big\}$ containing the variance of the different environmental noises imposed upon the system.

In addition, it may be useful to enforce correlation between white noises affecting different reactions. In a multi-strain epidemic model, for example, if white noise is meant to capture climatic variability its impact may be the same on all transmission reactions. The latter can be achieved by the introduction of a second level of hierarchy accounting for grouping among noisy reactions. The latter can be determined through a mapping function $\varphi : \mathcal{R}^e \to [1 : n_g]$ so that $\varphi^{-1}(p)$ corresponds to the indexes of a group of correlated reactions for each $p \in [1 : n_g]$. More details can be found in the following paragraph.

Diffusion approximation of compartmental models in the general case

From the previous results, a diffusion approximation of compartmental models in the general case is provided by the following SDE:

$$\begin{cases} dz_t = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt + L dB_t^Q \\ dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta) dt + L^{\theta_t} dB_t^{Q^{\theta_t}} \end{cases} \tag{22}$$

The matrices L and Q are constructed by concatenating the dispersion and diffusion matrices of the different sources of independent noises:

$$L = \begin{pmatrix} L^d & L^e \end{pmatrix} \quad and \quad Q = \begin{pmatrix} Q^d & 0 \\ 0 & Q^e \end{pmatrix} \tag{23}$$

With $Q^d = diag\{r^{(k)}(z_t, \theta) z_t^{\chi(k)},\ k \in \mathcal{R}\}$ and $L^d = [l^{(1)}, .., l^{(c)}]$ accounting for the demographic stochasticity on one hand. White noise environmental counterparts are defined in the following way $L^e = [l^{(k)}]_{k \in \mathcal{R}^e}$ is the concatenation of the stoichiometic vectors for noisy reactions. φ is the mapping function defined over $\mathcal{R}^e$ that attributes an equal index in $[1; n_g]$ to reactions upon which correlated environmental noise is enforced. From this function, a rectangular $card(\mathcal{R}^e) \times n_g$ dispersion matrix L^g can be constructed in which the column of group p is filled with $r(z_t, \theta) z_t^{\chi(k)}$ on rows corresponding to reactions such that $\varphi(k) = p$, and zero's everywhere else. With $Q^g = diag\big\{\big(\sigma^{(p)}\big)^2, p \in [1 : n_g]\big\}$, Q^e can be computed as $Q^e = L^g Q^g L^{g\prime}$. Naturally, this method for constructing correlated noise terms hold for uncorrelated noises.

2.4.3 Poisson process with stochastic rates

The continuous approximation of the number of individuals contained in each compartment, and its evolution, may be questionable when populations as stake are not large

enough and more specifically when the size of at least one compartment becomes small. Such situations typically correspond to the extinction of diseases or species in epidemic or ecological models. The Markovian jump process introduced earlier accounts for the discrete nature of the size of each compartment, and the discontinuities induced by the occurrence of each reaction. Nevertheless, due to the density-dependence of transformation rates the frequency of reactions increases infinitely as $N \to \infty$. Hence, the reference Markov jump process formalism quickly becomes intractable for other than small populations. The authors of Breto et al. (2009) have proposed an approximation of the Markov jump process based on a multinomial approximation of the number of reactions occurring over a short period of time dt. Here, we reformulate the solution proposed in Breto et al. (2009) and extend it to the general framework for compartmental models proposed in PLOM.

The Poisson process model determines the probability that each reaction k ($k \in \mathcal{R}$) respectively occurred n_k times over a given period dt. If all sources of environmental stochasticity are neglected:

$$p(n_1, \ldots, n_m | z_t, \theta) = \prod_{i=1}^{c} \left\{ M_i \left(1 - \sum_{\chi(k)=i} p_k \right)^{\overline{n}_i} \prod_{\chi(k)=i} (p_k)^{n_k} \right\} + o(dt)$$

Using the following notations:

$$p_k = p_k \left(r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt \right) = \left(1 - \exp \left\{ - \sum_{\chi(k')=i} r^{(k')}(z_t, \theta) z_t^{\chi(k')} dt \right\} \right) \frac{r^{(k)}(z_t, \theta)}{\sum_{\chi(k')=i} r^{(k')}(z_t, \theta)}$$

$$\overline{n}_i = z_t^{(i)} - \sum_{\chi(k)=i} n_k$$

$$M_i = \begin{pmatrix} z_t^{(i)} \\ \{n_k\}_{\chi(k)=i} \, \overline{n}_i \end{pmatrix} \qquad \text{(multinomial coefficient)}$$

In addition, white noise can be introduced on reaction k ($k \in \mathcal{R}^e$) by replacing time increments dt by random increments $d\Gamma_k$ with Gamma distribution, mean dt and standard deviation $\sigma^{(k)}\sqrt{dt}$:

$$p_k = p_k \left(r^{(k)} z_t^{\chi(k)} d\Gamma_k \right) = \left(1 - \exp \left\{ - \sum_{\chi(k')=i} r^{(k')}(z_t, \theta) z_t^{\chi(k')} d\Gamma_{k'} \right\} \right) \frac{r^{(k)} d\Gamma_k}{\sum_{\chi(k')=i} r^{(k')} d\Gamma_{k'}}$$

Lastly, time-varying parameters can be introduced in a similar manner as under previous formalisms:

$$dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta) dt + L^{\theta_t} dB_t^{Q^{\theta_t}}$$

3 Library of inference methods

In this section we will consider the slightly more general class of state space models evolving in continuous time, with discrete observations. The state of the system at time

t is note x_t. We will abusively note x_i the value of x_t at time t_i, hence $x_{0:n}$ denotes a trajectory of the system between t_0 and t_n. The prediction density $p(x_{i+1}|x_i, \theta)$ is generally untractable, which means that the probability of a path cannot be computed. However, we consider that it is possible to simulate trajectories from this prediction density. In addition, an observation model $p(y_i|x_i, \theta) = f(h(x_i); y_i, \theta)$ needs to be defined, to determine what is the probability of observing y_i conditionnally on the set of parameters θ and its proxy $h(x_i)$ built from the state of the system x_i.

3.1 Inference for state space models

State space models can be seen as a hypothesised probabilistic relation between the trajectories $x_{0:n}$ of a system and constant related quantities grouped in a parameter vector θ. This relation determines a joint probability density $p(x_{0:n}, \theta)$. From a Bayesian perspective, the knowledge or the uncertainty over the components of θ are enforced through the *a priori* density $p(\theta)$. For a given parameter vector θ, the likely trajectories of the system are reflected by the density $p(x_{0:n}|\theta)$. The primary objective of inference with state space models is the estimation of the posterior density $p(\theta|y_{1:n})$, and of the marginal density $p(x_{0:n}|y_{1:n})$. Additionally, model choice indicators can play a key role in disentangling between different hypothesised state space models (Spiegelhalter et al., 2002).

As suggested by the now classic motto "all models are wrong, but some are useful" (Box and Draper, 1987), models will only ever be a rough approximation of a complex reality. Yet, the latter does not prevent from following a scientific inductive approach to derive conclusions from the confrontations of models to data. By reconstructing the trajectory $x_{0:n}$ of the partially observed system or learning about uncertain components of θ, experience suggests that this process is likely to revise our understanding of infectious diseases (King et al., 2008). As in any other context, the validity of inference results shall be critically examined at least from a three-fold perspective. First, the uncertainties associated with the data collection should be reflected in the observation model. Then, the limitations of the model itself should be acknowledged and questioned, while considering the practical feasibility of proposing extensions to palliate the imperfections of the model. A minimal condition requires the output of the model to be able to fit the available observations of mechanisms they are meant to reproduce (Gelman and Shalizi, 2012). At last, the information derived regarding $x_{0:n}$ and θ, reflected by the discrepancies between their marginal prior and posterior densities, should not be considered as hard truth but rather as plausible and testable hypothesis (Popper, 2002).

An additional dimension arises when working with state space models, that requires specific attention. Although the joint posterior density $p(x_{0:n}, \theta|y_{1:n})$ can be computed pointwise up to a multiplicative constant hrough the Bayes rule (for a given trajectory $x_{0:n}$ and parameter θ), there is generally no direct way of deriving tractable formulas for the quantities of interest, i.e. $p(\theta|y_{1:n})$ and $p(x_{0:n}|y_{1:n})$. For sufficiently small-dimensional problems, efficient solutions for routine inference are offered by Gibbs MCMC samplers as the ones implemented in the Bugs library (Lunn et al., 2000). In its current version, the library of inference methods implemented in PLOM is constructed around the particle Marginal Metropolis Hastings algorithm (pMMH), which is one of the two versions of the particle Markov Chain Monte Carlo algorithm (pMCMC) (Andrieu et al., 2010). Along with this algorithm, PLOM provides with a series of tools to ensure and facilitate the convergence and efficient performance of the pMCMC algorithm. It

further provides with automatic diagnostic tools based on the CODA package (Plummer et al., 2006), and encourages peer validation for results published on PLOM.IO. We will introduce the different inference tools in the remainder of this section, and motivate and illustrate their combination in the following one.

3.2 Conditional state exploration: $p(x_{0:n}|y_{1:n},\theta)$ and $p(y_{1:n}|\theta)$

Sequential Monte Carlo

Sequential Monte Carlo (SMC) methods, also known as particle filters in this setting, provide an efficient solution to explore the space of trajectories of the system conditioned on a given parameter θ and available observations $y_{1:n}$. They are targeted to problems where the target density is sequentially structured, and can be decomposed as a product of terms. These terms are aggregated progressively in order to achieve a smooth transfer from a simple initial density corresponding to a single term of the product, up to the full target density. For example, for state space models the algorithm starts by approximating the prior density of initial conditions $p(x_0|\theta)$ with a swarm of samples called particles. At each iteration of the algorithm, an additional observation is accounted for, progressively increasing the dimension of the explored state. Particles are weighted according to how well they fit the new datapoint, and a resampling step is made to ensure that the exploration focuses on informative regions of the target space.

A classic version of the SMC algorithm, referred to as Systematic Importance Resampling algorithm, is presented in Algorithm 1 (Doucet and Johansen, 2009). If J is the number of particles, this algorithm can provide a sample $\tilde{x}_{0:n}$ from $\hat{p}^J_{pf}(x_{0:n}|y_{1:n})$, and an unbiased estimator $\hat{p}^J_{pf}(y_{1:n}|\theta)$ of $p(y_{1:n}|\theta)$. Under mild assumptions, the authors of Del Moral (2004) and Andrieu et al. (2010) have proved the following properties:

$$\begin{aligned} \|\hat{p}^J_{pf}(x_{0:n}|y_{1:n}) - p(x_{0:n}|y_{1:n})\| &\leq \frac{C_n}{J} \\ Var\left(\frac{\hat{p}^J_{pf}(y_{1:n}|\theta)}{p(y_{1:n}|\theta)}\right) &\leq \frac{D_n}{J} \end{aligned} \tag{24}$$

Where C_n and D_n are constants depending on the model and on the number of observations n. The distance $\|p_2 - p_1\|$ is defined as the total variation distance between the two distributions. Consequently, the particle filter is a solution to achieve asymptotically exact estimation of the marginal likelihood with precisions increasing as $O(J^{1/2})$.

Extended Kalman Filter

An approximate solution to the filtering problem for state space models is provided by the Extended Kalman Filter (EKF) algorithm (Jazwinski, 1970; Särkkä, 2006). We consider its continuous-discrete version tailored to dynamic models formulated as stochastic differential equations, with μ corresponding to the drift component of the model (which Jacobian is noted $\nabla\mu$), and diffusion and dispersion matrices being respectively noted Q and L. R_k is the variance of the observation process at time k. The EKF, described in Algorithm 2, is based on a gaussian approximation of the observation process h (which Jacobian is noted ∇h), resulting in a multivariate normal density for $p(x_t|y_{0:n})$ characterised by its mean m_t and covariance C_t. It provides with a deterministic and biased estimate $p^{EKF}(y_{1:n}|\theta)$ of the marginal likelihood.

Note that in Algorithm 2, only one observations is integrated at each time step. In the case of simultaneous observations, the same steps can be followed several time to update iteratively the mean and covariance of the state vector, observation per observation.

Algorithm 1 Sequential Monte Carlo algorithm

Set $L = 1$, $W_0^{(j)} = \frac{1}{J}$, sample $(x_0^{(j)})_{j=1,\dots,J}$ from $p(x_0|\theta)$

for $k = 0$ to $n - 1$ **do**

  for $j = 1$ to J **do**

    Sample $(x_{k:k+1}^{(j)})$ from $p(x_{k:k+1}|x_k, \theta)$

    Set $\alpha^{(j)} = h(y_{k+1}, x_{k+1}^{(j)}, \theta)$

  end for

  Set $W_{k+1}^{(j)} = \frac{\alpha^{(j)}}{\sum_{l=1}^{J} \alpha^l}$, and $L = L \times \frac{1}{J} \sum_j \alpha^{(j)}$

  Resample $(x_{0:k+1}^{(j)})_{j=1,\dots,J}$ according to $(W_{k+1}^{(j)})$,

end for

Algorithm 2 Continuous-discrete Extended Kalman Filter algorithm

Set $L = 1$ and initialise the mean state m_t and covariance C_t

for $k = 1$ to n **do**

  Integrate between t_{k-1} and t_k:

  $\frac{dm_t}{dt} = \mu(m_t, \theta)$

  $\frac{dC_t}{dt} = \nabla\mu(m_t, \theta)C_t + C_t\nabla\mu(m_t, \theta)^T + LQL'$

  Compute the prediction error $e = y_k - h(m_{t_k}, \theta)$, and the following quantities:

  $S = \nabla h(m_{t_k}, \theta)C_{t_k}\nabla h(m_{t_k}, \theta)' + R_{t_k}$

  $K = C_{t_k}\nabla h'(m_t, \theta)S^{-1}$

  Update the mean state and Covariance:

  $m_t = m_t + Ke$

  $C_t = C_t - KSK'$

  Update the likelihood $L(\theta) = L(\theta) \times \mathcal{N}(e; 0, S)$

end for

3.3 Full inference of paths and parameters

The central methodology utilised in PLOM to estimate the paths and parameters of compartmental models is the pMMH version of the pMCMC. For the sake of completeness, and for readers that may not be familiar with this methodology, we start by a brief introduction of the Monte Carlo Markov Chain machinery.

3.3.1 Introduction to the Monte Carlo Markov Chain machinery

Monte Carlo Markov Chain (MCMC) methods are used to estimate properties of probability densities in cases where analytic formulas cannot be directly derived, and samples cannot be directly generated. If we note x ($x \in \mathbb{R}^d$) the random variable of a target density $\pi(.)$, MCMC algorithms only require the ability to compute $\pi(x)$ for any x, up to a multiplicative factor. Their founding mechanism is the construction of a Markov chain that randomly explores $\mathbb{R}^d$ taking values $(x^{(1)}, x^{(2)}, .., x^{(N)})$ which will asymptotically ($N \to \infty$) mimic samples drawn from the target distribution. The chain is defined through a transition kernel K that determines the transition probability $p(.|x^{(i-1)})$. The chain converges to an invariant distribution if K is irreducible (from any state there is a positive probability to visit any other state) and aperiodic. The detailed balance condition is a sufficient but not necessary condition to ensure that the invariant distribution of the chain is the target density π:

$$\pi(x^i)K(x^{(i-1)}|x^{(i)}) = \pi(x^{(i-1)})K(x^{(i)}|x^{(i-1)}) \tag{25}$$

A critical dimension of MCMC algorithms is their efficiency in *mixing*, i.e. in generating samples that are as independent as possible. Unless $K(.|x^{(i)})$ is equal to $\pi(.)$, N samples of the MCMC trajectory will not provide the same amount of information as N independent and identically distributed (i.i.d.) samples from the target density π. This can be quantified by the *Effective Sample Size* (ESS), for example, that estimates how many truly i.i.d. samples the MCMC output is equivalent to (Geyer, 1992; Brooks and Roberts, 1998):

$$ESS(\{x^{(1)}, x^{(2)}, .., x^{(N)}\}) = \frac{N}{1 + 2\sum_{k=1}^{k^{max}} Correl(\{x^{(1)}, .., x^{(N-k)}\}, \{x^{(k)}, .., x^{(N)}\})} \tag{26}$$

This indicator, along with other diagnostic tools proposed in the CODA package (Plummer et al., 2006), play a central role on PLOM.IO to assess the validity of results obtained through MCMC exploration of the complex and high-dimensional target density $p(x_{0:n}, \theta|y_{0:n})$. Before diving into the presentation of basic and more advanced MCMC algorithms, we introduce the most classic way to define transition kernels that respect the detailed balance condition: the Metropolis-Hastings step (Metropolis et al., 1953; Hastings, 1970). At each iteration of the chain, a proposed value x^* is sampled from an importance distribution $q(.|x^{(i)})$, and accepted with probability:

$$1 \wedge \frac{\pi(x^*)q(x^{(i)}|x^*)}{\pi(x^{(i)})q(x^*|x^{(i)})} \tag{27}$$

Otherwise, x^{i+1} is set equal to x^i. The proportion of proposed samples that have been accepted determine the acceptance rate. The Metropolis Hastings step allows the use of any importance distribution q respecting the irreducibility and aperiodicity conditions,

Algorithm 3 random walk Metropolis algorithm

Initialise $x^{(0)}$

for $i = 0$ to N **do**

Sample $x^* \sim \mathcal{N}(x^{(i)}, \Sigma^q)$

Accept x^* with probability $1 \wedge \frac{\pi(x^*)}{\pi(x^{(i)})}$

end for

although other choices are also possible. It is generally observed that increasing the dimension of x decreases the acceptance probability.

The random walk Metropolis is based on a Metropolis-Hastings steps using a multivariate normal importance sampling distribution: $q(.|x^{(i)}) = \mathcal{N}(x^{(i)}, \Sigma^q)$ (see Algoritm 3). The efficiency of this algorithm on a given problem depends on the calibration of the covariance matrix Σ^q. Theoretical results have been demonstrated in the situation where the target distribution π is a multivariate normal density:

Proposition 1. *When π is a multivariate normal density, the acceptance rate that maximises the mixing efficiency of the random walk Metropolis algorithm is* 23.4% *(Roberts et al., 1997)*

Proposition 2. *When π is a multivariate normal density, optimal results are achieved by using $\Sigma^q = \frac{2.38^2}{d} \times Cov(\pi)$ (Roberts et al., 1997).*

When the target distribution is not a multivariate normal density, these results are extrapolated and followed as rules of conduct. They were used to derive adaptive versions of the random walk Metropolis algorithm, based on a decomposition of Σ^q into $\lambda\Sigma$. A first adaptive algorithm exploits the monotonicity of the acceptance rate as a function of λ. The Metropolis-Hastings ratio of a random walk Metrpopolis algorithm is $\frac{\pi(x^*)}{\pi(x)}$. Hence, if the mass of the target density is concentrated in a certain region and x is in this region (which is the case with high probability if the chain has converged), increasing the value of λ increases the risk for x^* to escape that region, leading to low values of $\pi(x^*)$ and rejection of x^*. On the contrary, excessively small values of λ will induce values of $\frac{\pi(x^*)}{\pi(x)}$ close to one and high acceptance rates. Therefore, the targeted acceptance rate can be approached by iteratively adapting λ with a cooling rate $a \in [0; 1[$:

$$\lambda_{i+1} = \lambda_i \times a^i(AccRate_i - 0.234) \tag{28}$$

A second adaptive algorithm relies on the fact that, as the chain progresses, the generated samples are meant to mimick i.i.d. samples generated from the target distribution π. Consequently, the empirical covariance matrix obtained from these samples can be used as a proxy for the optimal covariance matrix $\frac{2.38^2}{d} \times Cov(\pi)$. The resulting adaptive algorithm proposed in Roberts and Rosenthal (2009) is based on the following importance sampling distribution:

$$q(.|x^{(i)}) = \alpha\mathcal{N}\left(x^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(0)}\right) + (1-\alpha)\mathcal{N}\left(x^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(i)}\right) \tag{29}$$

With Σ_i being the empirical covariance matrix obtained from the i samples generated by the chain. The use of a mixture of normal distributions (α is generally set to 0.05) is meant to avoid convergence to local modes.

3.3.2 Particle Marginal Metropolis Hastings algorithm

Sequential Monte Carlo techniques are a natural choice to explore $p(x_{0:n}|y_{1:n}, \theta)$. In order to account for uncertainties regarding the parameter vector θ, we are aiming for the exploration of the joint posterior density $p(x_{0:n}, \theta|y_{1:n})$. As previously mentioned, classic MCMC methods fail to be efficient and robust solutions because of the high dimension of the target density. The particle MCMC algorithm offers a solution relying on the efficiency of particle filters (Andrieu et al., 2010). Algorithm 4 illustrates the principles of its particle marginal Metropolis Hastings version: the high-dimensional density exploration problem is reduced to the design of an MCMC algorithm over θ, based on the likelihood $\hat{p}_{pf}(y_{1:n}|\theta)$ estimated by a particle filter conditioned on θ. The authors of Andrieu et al. (2010) have shown that the algorithm was asymptotically exact for a given discretisation of time, as under classic assumptions when the number of iterations N^θ tends to infinity:

$$\|\hat{p}^J_{pf}(x^{(i)}_{0:n}, \theta^{(i)}|y_{1:n}) - p(x_{0:n}, \theta|y_{1:n})\| \to 0 \quad as\ i \to\ \infty \tag{30}$$

Every iteration of the MCMC algorithm implies running a particle filter to explore the range of likely paths of the system under the current value of θ and observed data $y_{1:n}$. Consequently, the pMCMC is a computationally demanding algorithm; its complexity if of the order of $O(nJN^\theta)$. The mixing efficiency of the MCMC scheme critically determines the applicability of the algorithm. In the absence of suitable techniques to efficiently estimate the marginal score $\nabla_\theta \log p(\theta|y_{1:n})$, random walk Metropolis algorithms are generally used. Even in its adaptive form, the parameterisation of its initial covariance matrix Σ^q_0 is a central issue: we will explore in the next subsection a mean to automatise this process, rendering the pMCMC algorithm plug-and-play.

Algorithm 4 Particle MCMC algorithm (particle marginal Metropolis Hastings version)

Initialise $\theta^{(0)}$.

Use the SMC algorithm to compute $\hat{p}^J_{pf}(y_{1:n}|\theta^{(0)})$ and sample $x^{(0)}_{0:n}$ from $\hat{p}^J_{pf}(x_{0:n}|y_{1:n}, \theta^{(0)})$

for $i = 1$ to N^θ **do**

Sample θ^* from $q(.|\theta^{(i)})$

Use the SMC to compute $L(\theta^*) = \hat{p}^J_{pf}(y_{1:n}|\theta^*)$ and sample $x^*_{0:n}$ from $\hat{p}^J_{pf}(x_{0:n}|y_{1:n}, \theta^*)$

Accept θ^* (and $x^*_{0:n}$) with probability $1 \wedge \frac{L(\theta^*)q(\theta^{(i)}|\theta^*)}{L(\theta^{(i)})q(\theta^*|\theta^{(i)})}$

Record $\theta^{(i+1)}$ and $x^{(i+1)}_{1:n}$

end for

An alternative solution is the SMC^2 algorithm presented in Chopin et al. (2012). It explores both the probability density of $x_{0:n}$ and θ with an SMC algorithm, starting from the initial target $p(x_0, \theta)$ and progressively incorporating the available observations. The global complexity of this algorithm is similar to the pMCMC, but its ability to automatically adapt the number particles being utilised and to progressively learn from

previous samples what could be seen as the equivalent of the covariance matrix Σ^q are promising features.

3.3.3 Efficiently initialisation and calibration of the PMMH algorithm

As we just mentioned, each iteration of the pMCMC algorithm is computationally demanding. For this reason, we want to reduce the calibration period of the pMCMC itself (also known as *burn-in* period), which can be done by preliminary pre-explorations of the target density $p(\theta|y_{1:n})$. A first classical caveat, when MCMC chains are initialised from an arbitrary position, is that likelihood follows an increasing trend before it stabilises, indicating that the chain has converged to a mode. Naturally, such non-stationarity in the generated samples induces a strong autocorrelation and provides non-exploitable information. It is then natural to rely on optimisation algorithms to accelerate this transient phase and directly launch the pMCMC chain close from a mode of the posterior density. In addition, complex target densities generally exhibit local modes in which MCMC or optimisation algorithms can be trapped, again leading to false and misleading results. The search for a global mode is a challenging problem in itself and should be done with suitable tools, that will now present. In addition, the pMCMC implementation proposed in PLOM allows for the adaptation of the sampling covariance Σ^q. This adaptation phase can be long and costly, due to a "chicken and egg" situation: adaptation of Σ^q is most needed when mixing is poor, which is also the situation where learning is the slowest. This issue will also be covered in this Section.

Searching for the global mode with the simplex, ksimplex and mif algorithms

One of the most classic algorithms to optimise a function of continuous variables in an unconstrained space is the simplex algorithm, also known as the Nelder-Mead algorithm (see Algorithm 5). To avoid the complications that arise when introducing constraints (positivity or boundedness, for example), the components of θ are transformed through log or logit functions (or extensions thereof, to allow for boundaries different than 0 and 1), before being handed to the simplex algorithm. The latter operates by constructing polygon with $d+1$ vertices, with $d = dim(\theta)$, and optimising the value of the target function at each of its vertices (in our case, $p(\theta^{(d+1)}|y)$) through reflection, expansion, contraction or reduction transformations of the polygon. The use of this algorithm in PLOM directly relies on its implementation in the GNU Scientific Library Galassi and Gough (2006). The complexity of this algorithm increases linearly with d. In addition, it is a local exploration algorithm, and although it does not strictly follow the gradient of the target density, it can easily be trapped in local modes. At last, the simplex algorithm requires the target function $p(\theta^{(d+1)}|y)$ to be computed deterministically: it cannot be directly plugged to the particle filter where the estimation of the likelihood is noisy. As a consequence, the `simplex` algorithm can only be used in PLOM on ode approximations of the system (see 2.4.1). In order to account for demographic or environmental sources of stochasticity, it is possible to estimate the likelihood with the Extended Kalman Filter from an sde approximation of the model (see 2.4.2). As this estimate can be obtained deterministically, it can be plugged into the simplex algorithm: this is the `ksimplex` function available in PLOM.

The optimisation routines based on algorithm 5 require the use of ode or sde ap-

proximations of the system. These may lead to biased estimated of the optimal set of parameters, with regards to what could be found using a psr formalism (see 2.4.3). However, they are only used to initialise the Markov Chain that subsequently explores the posterior density. As a consequence, potential discrepancies between the likelihoods induced under the different formalisms will only have serious consequences in some critical cases. In such situations, it is possible to rely on an asymptotically exact and plug-and-play solution to the frequentist problem of maximising the marginal likelihood $p(y|\theta)$ (Ionides et al., 2006; Breto et al., 2009; Ionides et al., 2011). This approach has been used for several applications in epidemiology. Although it has been developed and utilised as a purely Frequentist algorithm, the mif can also it can be used to efficiently initialise the Markov chain of the PMCMC by incorporating the prior density into the maximised function of θ as illustrated in Algorithm 6. Under the current implementation, corresponding to the algorithm described in Breto et al. (2009), careful parameterisation of the algorithm is required to achieve convergence to the mode. Further investigation is being carried out to increase the efficiency and stability of the iterated filtering algorithm (Ionides et al., 2012; Lindström, 2013), which could allow to directly optimise the posterior distribution under the optimal psr formalism, exploiting the interesting *tempering* feature of this approach. In the meantime, serialised `simplex` or `ksimplex` algorithms shall be preferred for routine maximisation of the posterior density.

Fast exploration of a proxy posterior density: the kMCMC algorithm

Even when the Markov Chain is initialised close from the global model of the posterior density, the adaptation of the sampling covariance matrix Σ^q of the pMCMC algorithm can be lengthy. We know that in the multivariate normal case, the optimal choice for Σ^q is proportional to the covariance of the target density $p(\theta|y_{1:n})$. As the Extended Kalman Filter provides an efficient way to deterministically obtain an estimate of the likelihood $p(y_{1:n}|\theta)$ under the sde formalism, it is natural to construct an algorithm analogous to the pMCMC, based on $\hat{p}_{EKF}(y_{1:n}|\theta)$, that will efficienctly provide an estimate of the covariance of the proxy posterior density (see Algorithm 7). In addition the one or two order of magnitudes gained by estimating the likelihood with the EKF instead of an SMC algorithm, the complete absence of noise on this estimate also significantly facilitates the automatic calibration of the kmcmc algorithm.

4 Iterated inference pipeline: illustrations

5 Perspectives: call for contributions

References

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

Box, G. E. and Draper, N. R. (1987). *Empirical model-building and response surfaces.* John Wiley & Sons.

Breto, C., He, D., Ionides, E., King, A., Ghosal, S., Lember, J., and van der Vaart, A. (2009). Time series analysis via mechanistic models. *Annals*, 3(1):319–348.

Brooks, S. and Roberts, G. (1998). Convergence assessment techniques for markov chain monte carlo. *Statistics and Computing*, 8(4):319–335.

Cazelles, B. and Chau, N. (1997). Using the kalman filter and dynamic models to assess the changing hiv/aids epidemic. *Mathematical biosciences*, 140(2):131–154.

Chopin, N., Jacob, P., and Papaspiliopoulos, O. (2012). SMC^2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Cori, A., Boelle, P., Thomas, G., Leung, G., and Valleron, A. (2009). Temporal variability and social heterogeneity in disease transmission: The case of sars in hong kong. *PLoS computational biology*, 5(8):e1000471.

Del Moral, P. (2004). *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Springer.

Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: fifteen years later. *Handbook of Nonlinear Filtering*, pages 656–704.

Dureau, J., Kalogeropoulos, K., and Baguelin, M. (2013). Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*.

Ethier, S. and Kurtz, T. (1986). *Markov processes: characterization and convergence*, volume 6. Wiley.

Fuchs, C. (2013). *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer.

Galassi, M. and Gough, B. (2006). Gnu scientific library: reference manual.

Gelman, A. and Shalizi, C. (2012). Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology*.

Geyer, C. (1992). Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):473–483.

Guy, R., Larédo, C., and Vergu, E. (2013). Approximation of epidemic models by diffusion processes and their statistical inference. *arXiv preprint arXiv:1305.3492*.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97.

Ionides, E., Bhadra, A., Atchade, Y., and King, A. (2011). Iterated filtering. *The Annals of Statistics*, 39(3):1776–1802.

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

Ionides, E., Frydendall, J., Madsen, H., et al. (2012). Efficient iterated filtering. In *System Identification*, volume 16, pages 1785–1790.

Jazwinski, A. (1970). *Stochastic processes and filtering theory*, volume 63. Academic press.

King, A., Ionides, E., Pascual, M., and Bouma, M. (2008). Inapparent infections and cholera dynamics. *Nature*, 454(7206):877–880.

Kloeden, P. and Platen, E. (1999). *Numerical Solution to Stochastic Differential Equations*. Springer.

Lindström, E. (2013). Tuned iterated filtering. *Statistics & Probability Letters*.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337.

Metropolis, N., Rosenbluth, A., M.N. Rosenbluth and, A. T., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21:1087.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R news*, 6(1):7–11.

Popper, K. (2002). *The logic of scientific discovery*. Number 117 in A. Routledge.

Roberts, G., Gelman, A., and Gilks, W. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Roberts, G. and Rosenthal, J. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Särkkä, S. (2006). Recursive bayesian inference on stochastic differential equations. *PhD dissertation*, 150:1–150.

Spiegelhalter, D., Best, N., Carlin, B., and Linde, A. V. D. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Viboud, C., Pakdaman, K., Bo
"elle, P., Wilson, M., Myers, M., Valleron, A., and Flahault, A. (2004). Association of influenza epidemics with global climate variability. *European Journal of Epidemiology*, 19(11):1055–1059.

Algorithm 5 Simplex algorithm (a.k.a. Nelder-Mead algorithm)

Initialise $(\theta^{(1)}, \ldots, \theta^{(d+1)})$, with $d = dim(\theta)$.

Set $N = 0$.

Unless stated otherwise, set $\alpha = 1, \gamma = 2, \rho = -1/2$, and $\sigma = 1/2$

while convergence is not achieved and $N < N_{max}$ **do**

N = N+1

Order according to the values at the vertices $p(\theta^{(d+1)}|y) \leq \cdots \leq p(\theta^{(0)}|y)$

Calculate $\theta^{(0)}$, the center of gravity of all points but $\theta^{(d+1)}$

Reflection

Compute reflected point $\theta^{(r)} = \theta^{(0)} + \alpha(\theta^{(0)} - \theta^{(d+1)})$

If $p(\theta^{(d)}|y) \leq p(\theta^{(r)}|y)$ and $p(\theta^{(r)}|y) < p(\theta^{(0)}|y)$:

replace $\theta^{(d+1)}$ by $\theta^{(r)}$ and end iteration.

Expansion

If $p(\theta^{(1)}|y) \leq p(\theta^{(r)}|y)$:

If $p(\theta^{(r)}|y) \leq p(\theta^{(e)}|y)$:

Compute the expanded point $\theta^{(e)} = \theta^{(0)} + \alpha(\theta^{(0)} - \theta^{(d+1)})$,

and replace $\theta^{(d+1)}$ by $\theta^{(e)}$, and end iteration.

Else:

Replace $\theta^{(d+1)}$ by $\theta^{(r)}$, and end iteration.

Contraction

We know that $p(\theta^{(r)}|y) \leq p(\theta^{(d)}|y)$.

Compute the contracted point $\theta^{(c)} = \theta^{(0)} + \rho(\theta^{(0)} - \theta^{(d+1)})$,

If $p(\theta^{(d+1)}|y) \leq p(\theta^{(c)}|y)$:

Replace $\theta^{(d+1)}$ by $\theta^{(c)}$, and end iteration.

Reduction

For all points but $\theta^{(1)}$, replace $\theta^{(i)}$ by $\theta^{(1)} + \sigma(\theta^{(i)} - \theta^{(1)})$

end while

Algorithm 6 Posterior density Maximization by Iterated Filtering

Initialise $x_I^{(1)}$ and $\theta^{(1)}$.

Unless stated otherwise, set $a = 0.975, b = 2, \rho = -1/2$, and $L = round(0.75 \times n)$

for $m = 1$ to M **do**

Sample initial conditions, $\tilde{x}_I^{(j)}(t_0) \sim \mathcal{N}\big(x_I^{(m)}, a^{m-1}\Sigma_I\big), \qquad j = 1, \ldots, J$

Initialise filtered states, $\tilde{x}_F^{(j)}(t_0) = \tilde{x}_I^{(j)}(t_0)$

Rejuvenate parameters, $\tilde{\theta}^{(j)}(t_0) \sim \mathcal{N}\big(\theta^{(m)}, ba^{m-1}\Sigma_\theta\big)$

Set $\bar{\theta}(t_0) = \theta^{(m)}$

for $i = 1$ to n **do**

Propagate samples, $\tilde{x}_P^{(j)}(t_i) \sim p\big(x(t_i)|\tilde{x}_F^{(j)}(t_{i-1}), \tilde{\theta}^{(j)}(t_{i-1})\big)$

Compute weights, $w_i^{(j)} = p\big(y_i|\tilde{x}_P^{(j)}(t_i), \tilde{\theta}^{(j)}(t_{i-1})\big) \times p\big(\tilde{\theta}^{(j)}(t_{i-1})\big)^{\frac{1}{n}}$

Draw $k_1, \ldots, k_J$ such that $p(k_j = i) = w_i^{(i)} / \sum_l w_i^{(l)}$;

and filter the predicted states $\tilde{x}_F^{(j)}(t_i) = \tilde{x}_P^{(k_j)}(t_i)$

Filter the initial conditions $\tilde{x}_I^{(j)}(t_i) = \tilde{x}_I^{(k_j)}(t_{i-1})$

Filter and rejuvenate parameters, $\tilde{\theta}^{(j)}(t_i) \sim \mathcal{N}\big(\tilde{\theta}^{(k_j)}(t_{i-1}), a^{m-1}(t_i - t_{i-1})\Sigma_\theta\big)$

Set $\bar{\theta}(t_i)$ to be the sample mean of $\{\tilde{\theta}^{(k_j)}(t_{i-1})\}_{1 \le j \le J}$

Set $V(t_i)$ to be the sample mean of $\{\tilde{\theta}^{(j)}(t_i)\}_{1 \le j \le J}$

end for

Set $\theta^{(m+1)} = \theta^{(m)} + V(t_1) \sum_{i=1}^{n} V^{-1}(t_i)(\bar{\theta}(t_i) - \bar{\theta}(t_{i-1}))$

Set $x_I^{(m+1)}$ to be the sample mean of $\{\tilde{x}_I^{(j)}(t_L)\}_{1 \le j \le J}$

end for

Algorithm 7 Kalman MCMC algorithm

Initialise $\theta^{(0)}$.

Use the EKF algorithm to compute $\hat{p}_{EKF}(y_{1:n}|\theta^{(0)})$

for $i = 1$ to N^θ **do**

Sample θ^* from $q(.|\theta^{(i)})$

Use the EKF to compute $L(\theta^*) = \hat{p}_{EKF}(y_{1:n}|\theta^*)$

Accept θ^* with probability $1 \wedge \frac{L(\theta^*)q(\theta^{(i)}|\theta^*)}{L(\theta^{(i)})q(\theta^*|\theta^{(i)})}$

Record $\theta^{(i+1)}$

end for