

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

Joseph Dureau, Sébastien Ballesteros

July 6, 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  $z_t = [z_t^{(1)}, \dots, z_t^{(c)}]$ . 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  $\chi : \mathcal{R} \rightarrow [1; c]$  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  $\theta$ . In the remaining of this document, we will define compartmental models using the following formalism:

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

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

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

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  $t$ :

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

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:

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

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  $x_t^{\theta_t}$  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):

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

Under this formalism, the number of individuals in each compartment takes continuous 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  $\theta$ .

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 helpful to handle and represent different and independent sources of stochasticity:

$$dx_t = \mu_t(x_t, \theta)dt + LdB_t^{Q_t} \quad (4)$$

In this equation,  $\mu_t$  is referred to as the drift,  $L$  as the dispersion matrix, and  $Q_t$  as the diffusion matrix of the driving Brownian 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 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:

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

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:

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

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

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

The diffusion approximation relies on the limit of this expression when  $\varepsilon \rightarrow 0$  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:

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

which is equivalent to

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

Where

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

and

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

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

$$dz_t = \dot{A}(z_t)dt + LdB_t^{\dot{Q}_t^d} \quad (12)$$

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  $L\dot{Q}_t^d L = \dot{\Sigma}$ :

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

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  $Q^d = N^2\dot{Q}_t^d$ :

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

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  $dB_t$  being a multivariate Brownian motion with diffusion matrix  $Q^d = \text{diag}\{r^{(k)} z_t^{\chi^{(k)}}, k \in \mathcal{R}\}$ , 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:

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

Lastly, the resulting expression for  $\Sigma$  is the following:

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

### Diffusion approximation of the environmental stochasticity

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:

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

In the case of the SIR model:

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

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

$$\begin{cases} dS_t = -\beta S_t \frac{I_t}{N} d\Gamma_t \\ dI_t = \beta S_t \frac{I_t}{N} d\Gamma_t - \gamma I_t dt \\ dR_t = \gamma I_t dt \end{cases} \quad (19)$$

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:

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

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

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

$L^e$  is the  $c \times \text{Card}(\mathcal{R}^e)$  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  $Q^e = \text{diag}\{(\sigma^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi^{(k)}})^2, k \in \mathcal{R}^e\}$  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  $\varphi : \mathcal{R}^e \rightarrow [1 : n_g]$  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:

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

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

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

With  $Q^d = \text{diag}\{r^{(k)}(z_t, \theta) z_t^{\chi^{(k)}}, k \in \mathcal{R}\}$  and  $L^d = [l^{(1)}, \dots, l^{(c)}]$  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 stoichiometric vectors for noisy reactions.  $\varphi$  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  $\text{card}(\mathcal{R}^e) \times n_g$  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  $Q^g = \text{diag}\{(\sigma^{(p)})^2, p \in [1 : n_g]\}$ ,  $Q^e$  can be computed as  $Q^e = L^g Q^g L^{g'}$ . 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 at 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  $N \rightarrow \infty$ . 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:

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

Using the following notations:

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

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

$$M_i = \binom{z_t^{(i)}}{\{n_k\}_{\chi(k)=i} \bar{n}_i} \quad (\text{multinomial coefficient})$$

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

#### 3.1 Objectives of the inference process

Dynamic models can be seen as a hypothesised probabilistic relation between the trajectories of a system ( $x_t$  for  $t \in [t_0; t_n]$ , noted  $x_{0:n}$ ), and constant quantities grouped in a



parameter vector  $\theta$ . They induce the definition of a joint probability density  $p(x_{0:n}, \theta)$ . From a Bayesian perspective, the knowledge or the uncertainty over the components of  $\theta$  are enforced through the *a priori* density  $p(\theta)$ . For a given parameter vector  $\theta$ , the likely trajectories of the system are reflected by the density  $p(x_{0:n}|\theta)$ . The likelihood of the discrete dataset  $y_{1:n}$  under a given scenario is given by an observation model defining  $p(y_{1:n}|\theta, x_{0:n})$ .

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. Experience suggests that this process is likely to revise our understanding of infectious diseases (King et al., 2008). By exploring the joint posterior density  $p(x_{0:n}, \theta|y_{1:n})$ , information can be deducted respectively on the partially observed stochastic processes  $x_{0:n}$  and on the parameter vector  $\theta$ .

In any classical act of inference, the validity of this information 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  $\theta$ , 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 requiring attention when exploring the high-dimensional density  $p(x_{0:n}, \theta|y_{1:n})$ . Although this quantity can be computed pointwise up to a multiplicative constant (for a given trajectory  $x_{0:n}$  and parameter  $\theta$ ), through the Bayes rule, there is generally no direct way of deriving tractable formulas for the quantities of interest (i.e.  $p(\theta|y_{1:n})$ ,  $p(x_t|y_{1:n})$ , etc). 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 next one.

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

Distributions  $p(x_i|y_{1:i}, \theta)$  and  $p(x_{0:n}|y_{1:n}, \theta)$  are respectively termed *filtering* and *smoothing* conditional densities. They indicate what can be said about the state of the system at a given time  $t_i$  given all past observation, or what has been the trajectory of the system given all the available observations. An additional objective of the exploration of the filtered density is the estimation of the marginal likelihood  $p(y_{1:n}|\theta)$ . Particle

filtering methods provide an efficient solution to this problem.

A smoothing version of the most classic particle filter, referred to as Systematic Importance Resampling algorithm, is presented in Algorithm 1 (Doucet and Johansen, 2009). This algorithm can provide a sample  $\tilde{x}_{0:n}$  from  $\hat{p}_{pf}(x_{0:n}|y_{1:n})$ , and an unbiased estimator  $\hat{p}_{pf}(y_{1:n}|\theta)$  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:

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

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  $O(J^{1/2})$ , where  $J$  is the number of particles.

---

**Algorithm 1** Particle Smoothing algorithm for indirectly observed stochastic processes

---

```

Set  $L = 1$ ,  $W_0^{(j)} = \frac{1}{J}$ , sample  $(x_0^{(j)})_{j=1,\dots,J}$  from  $p(x_0|\theta)$  and calculate  $(z_0^{(j)})_{j=1,\dots,J}$ 
for  $k = 0$  to  $n - 1$  do
  for  $j = 1$  to  $J$  do
    Sample  $(x_{k:k+1}^{(j)})$  from  $p(x_{k:k+1}^{dis}|x_k, \theta)$  and calculate  $(z_{k:k+1}^{(j)})$ 
    Set  $\alpha^{(j)} = h(y_{k+1}, z_{k:k+1}^{(j)}, \theta)$ 
  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  $(z_{0:k+1}^{(j)}, x_{0:k+1}^{(j)})_{j=1,\dots,J}$  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  $k = 0$  to  $n - 1$  do
  for  $j = 1$  to  $J$  do
    Integrate between  $t_k$  and  $t_{k+1}$ :
       $\frac{dm_t}{dt} = \mu(m_t, \theta)$ 
       $\frac{dC_t}{dt} = \nabla \mu(m_t, \theta) C_t + C_t \nabla \mu(m_t, \theta)^T + L Q L^T$ 
    Compute the prediction error  $err = y_k - \psi(m_t, \theta)$ , and the following quantities:
       $S = \nabla \psi(m_t, \theta) C_k \nabla \psi(m_t, \theta)^T + R_k$ 
       $K = C_t \nabla \psi^T(m_t, \theta) S^{-1}$ 
    Update the mean state and Covariance:
       $m_t = m_t + K err$ 
       $C_t = C_t - K S K^T$ 
    Update the likelihood  $L(\theta) = L(\theta) \times \mathcal{N}(err; 0, S)$ 
  end for
end for

```

---

An approximate solution to the filtering problem for nonlinear and stochastic dynamic 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  $\mu$  corresponding to the drift component of the model (which Jacobian is noted  $\nabla\mu$ ), and diffusion and dispersion matrices being respectively noted  $Q$  and  $L$ . The EKF, described in Algorithm 2, is based on a gaussian approximation of the observation process  $\psi$  (which Jacobian is noted  $\nabla\psi$ ), 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, for any discretisation of time with step  $\delta$ .

### 3.2.1 Particle filter

### 3.2.2 Extended Kalman Filter

## 3.3 Full inference of paths and parameters

### 3.3.1 Particle Marginal Metropolis Hastings algorithm

### 3.3.2 Efficiently starting the PMMH close to the global mode

### 3.3.3 Efficient calibration of the PMMH sampling covariance matrix

## 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.
- 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.
- 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.
- Gelman, A. and Shalizi, C. (2012). Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology*.
- Guy, R., Larédo, C., and Vergu, E. (2013). Approximation of epidemic models by diffusion processes and their statistical inference. *arXiv preprint arXiv:1305.3492*.
- 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.
- 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.
- 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.
- Särkkä, S. (2006). Recursive bayesian inference on stochastic differential equations. *PhD dissertation*, 150:1–150.
- Viboud, C., Pakdaman, K., Boelle, 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.