S|S|M:
Inference for time series analysis with State Space Models

```
cat theta.json | ./ksimplex | ./kmcmc | ./pmcmc
```

Joseph Dureau, Sébastien Ballesteros, Tiffany Bogich

November 28, 2013

1   Introduction

The main motivation behind the open source library *SSM* is to reduce the technical friction that prevents modellers from sharing their work, quickly iterating in crisis situations, and making their work directly usable by public authorities to serve decision-making. An illustration of this problem is the fact that the scientific production on the 2009 H1N1 epidemic has peaked three years after, in 2012, and that the vast majority of this production takes the form of static `pdf` files. Even when the corresponding code is shared, it takes a substantial amount of time and effort for any other person than the author to at least reproduce the published results.

This document gathers the methodological aspects on which the *SSM* library is built on. Its first purpose is to provide full transparency on the general modeling framework that is proposed, and more precisely which are the available mathematical formulations for these models. It also describes the different inference algorithms implemented in the library, that are based on the state of the art of computational statistics. All the source code is open and available on github (https://github.com/standard-analytics/ssm), and all contributions, comments and suggestions are more than welcome!

The modelling facet of the library stemmed from a focus on epidemiology and ecology, but it is more widely targeted to all systems that can be represented as c systems of ordinary or stochastic differential equations, compartmental models and combinations thereof. While we refer the reader to the classic literature for ordinary and stochastic differential equations, the modelling framework for compartmental models is presented in Section 2 of this document, that proposes a grammar from which models can be defined in a simple and non-ambiguous way, freeing their fundamental description from their mathematical formulation and technical implementation. Stochastic differential equations are simply described through their deterministic drift and dispersion matrix. This definition of compartmental models simply relies on the definition of transformations of the system, called reactions, that occur at given rates. We extend this classic perspective with the possibility of introducing environmental stochasticity, that captures potential mis-specifications of the model. The later is based on the propositions made in **?**.

This mathematical formalisms available in this library to describe state space models are specially suited to particle-based methods as iterated filtering (**?**) and the particle MCMC (**?**). However, these methods remain computationally intensive which limits their applicability. Following the approach proposed in **?**, the *SSM* library also proposes algorithms based on the Extended Kalman Filter, that allow drastic cost reductions in preliminary explorations of targeted densities and efficient initialisations of particle-based methods. These aspects will be described in Section 3.

At last, we provide in Section 4 three illustrations of how compartmental models and inference methods can be used to explore historic datasets, monitor current epidemics, and forecast their future evolution.

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 c 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 c 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 and/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 characterise molecule types, electronic charge or radioactive states (**??**). Less classical illustrations of the use of compartmental models include tracking the spread of rumors among a population, spread of obesity, 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 (for individuals coming from outside the population of interest, a artificial source state can be introduced). This property 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 potential variations of external drivers of the system. They 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

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

Under some regularity conditions detailed in **?**, **?** or **?**, 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 **?**, to account for additional sources of uncertainty related to fluctuations in extrinsic determinants of the epidemic that are not explicitly included in the model. This uncertainty is reflected through additional sources of stochasticity, termed *environmental* stochasticity. The approach suggested in **?** is to consider stochastic transition rates $\tilde{r}_t^{(k)}$ for a subset $\mathcal{R}^e$, under the following constraints for all t:

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

The presence of white noise in a model with stochasticitay σ will be denoted $w.n.(\sigma)$ in the rows corresponding to noisy reactions. Yet, 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 (**?**). Following the work of **?** and **?**, the authors of **?** 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} respectively as the log or logit transformation of the quantity of interest.

2.3 Examples

We introduce three models that will be used in the last section in different contexts (plague, H1N1 and dengue). These models are defined following the formalism that has just been defined, from which different mathematical formulations can be derived as will be described in the following Section.

2.3.1 Plague: SI model with seasonal forcing

In this model, we consider two compartments: individuals are either infected or susceptible to be infected. Historical records show that plague is a seasonal forcing, consequently the reproduction rate is described as a periodic, sinusoidal function. If the life

expectancy with plague is denoted with μ_D^{-1}, the model can be described in the following way:

Reaction	Effect	Rate
infections	$(S_t, I_t) \to (\boldsymbol{S_t - 1}, \boldsymbol{I_t + 1})$	$R_0[1 + e\sin(t + \phi)]\mu_D I/N$
deaths	$(S_t, I_t) \to (S_t, \boldsymbol{I_t - 1})$	μ_D

2.3.2 H1N1: SEIR model with time-varying contact rate

We have illustrated in **?** how it is possible to capture the evolution of key parameters of an epidemic, on the specific example of the 2009 H1N1 epidemic in London. The latter exhibits two waves, which is an unusual trajectory and suggests that the drivers of the epidemic have changed over time. As a first approach to this problem, an SEIR model can be used to reflect the fact that after being infected, individuals spend some time in a latent state before developing symptoms and becoming infectious. At the end of the infectivity period, recovered individuals become resistant and can no longer become infected:

Reaction	Effect	Rate
infections	$(S_t, E_t, I_t, R_t) \to (\boldsymbol{S_t - 1}, \boldsymbol{E_t + 1}, I_t, R_t)$	$\beta_t I/N$
onset of symptoms	$(S_t, E_t, I_t, R_t) \to (S_t, \boldsymbol{E_t - 1}, \boldsymbol{I_t + 1}, R_t)$	k
recovery	$(S_t, E_t, I_t, R_t) \to (S_t, E_t, \boldsymbol{I_t - 1}, \boldsymbol{R_t + 1})$	γ

In order to capture the unknown variations of the effective contact rate β_t, it can be modelled using a random walk in the log space:

$$d\log \beta_t = \sigma dB_t \tag{3}$$

Note that in this example, transmissibility is directly reflected by the effective contact rate β_t, while a different parameterisation based on the reproduction rate R_0 and the infectivity period μ_D was used in the plague example. Both of these equivalent approaches can be found in the literature.

2.3.3 Dengue: parsmonious 2-strains model

Dengue is a seasonal disease with complex dynamics, for which four strains co-exist. It is believed that after recovering from infection, individuals go through a short period of cross-immunity that protects them from all strains. Typical dengue case records do not specify with which strain individuals have been infected, hence a parsimonious two-strains model has been proposed in **?** to temper identifiability issues:

Reaction	Effect	Rate
1st infections with str. 1	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (\boldsymbol{S_t - 1}, \boldsymbol{I1 + 1}, I2, R1, R2, S1, S2, I12, I21, R)$	$\beta_t(I1/N + i + \psi I21/N)$ $w.n.(\sigma)$
1st infections with str. 2	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (\boldsymbol{S_t - 1}, I1, \boldsymbol{I2 + 1}, R1, R2, S1, S2, I12, I21, R)$	$\beta_t(I2/N + i + \psi I12/N)$ $w.n.(\sigma)$
recovery from str. 1	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (S_t, \boldsymbol{I1 - 1}, I2, \boldsymbol{R1 + 1}, R2, S1, S2, I12, I21, R)$	γ
recovery from str. 2	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (S_t, I1, \boldsymbol{I2 - 1}, R1, \boldsymbol{R2 + 1}, S1, S2, I12, I21, R)$	γ
loss of cross-immunity	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (S_t, I1, I2, \boldsymbol{R1 - 1}, R2, \boldsymbol{S1 + 1}, S2, I12, I21, R)$	α
loss of cross-immunity	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (S_t, I1, I2, R1, \boldsymbol{R2 - 1}, S1, \boldsymbol{S2 + 1}, I12, I21, R)$	α
2nd infections with str. 2	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (S_t, I1, I2, R1, R2, \boldsymbol{S1 - 1}, S2, \boldsymbol{I12 + 1}, I21, R)$	$\beta_t(I2/N + i + \psi I12/N)$ $w.n.(\sigma)$
2nd infections with str. 1	$(S_t, I1, I2, R1, R2, S1, S2, I12, I21, R)$ $\rightarrow (S_t, I1, I2, R1, R2, S1, \boldsymbol{S2 - 1}, I12, \boldsymbol{I21 + 1}, R)$	$\beta_t(I1/N + i + \psi I21/N)$ $w.n.(\sigma)$

As with plague, seasonality is enforced explicitly through a sinusoidal factor:

$$\beta_t = \beta \times [1 + e \sin(t + \phi)]$$

This model allows for a different infectivity for individuals that have been infected for the second time, through the factor ψ. In addition, correlated *white* environmental noise is enforced on infection reactions as a mean to reflect potential mis-specifications of the model.

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):

$$\frac{dz_t}{dt} = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} \tag{4}$$

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 intrinsic and 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 other 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. Yet, 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 **?** that will be helpful to handle and represent different and independent sources of stochasticity:

$$dx_t = \mu_t(x_t, \theta)dt + LdB_t^{Q_t} \tag{5}$$

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 Eq. 5, utilising the notations that have been introduced earlier in this document:

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

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 $c \times n_{Q_t}$ (c is the number of compartments in the model). 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 reflected.

Diffusion approximation of the demographic stochasticity

In order to provide an SDE approximation of the demographic stochasticity, we rely on theoretical results of state-dependent Markov jump processes presented in **?**. The adaptation of these results to compartmental epidemic models models has been illustrated in **?**. Extensions of these results in non-homogeneous settings are provided in **?**.

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

$$\frac{\partial}{\partial t}P(z_t) = \sum_{k\in\mathcal{R}} r^{(k)}\tilde{z}_{k,t}^{\chi^{(k)}} P(z_t - l^{(k)}) - \sum_{k\in\mathcal{R}} r^{(k)}(z_t,\theta)z_t^{\chi^{(k)}}P(z_t) \tag{6}$$

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:

$$\begin{aligned}
\frac{\partial}{\partial t}P(S_t, I_t, R_t) = {} & \beta\frac{(S_t+1)}{N}(I_t-1)P(S_t+1, I_t-1, R_t) \\
& + \gamma(I_t+1)P(S_t, I_t+1, R_t-1) \\
& - \beta\frac{S_t}{N}I_t P(S_t, I_t, R_t) \\
& - \gamma I_t P(S_t, I_t, R_t)
\end{aligned} \tag{7}$$

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

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

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

$$\begin{aligned}
\frac{\partial}{\partial t}P(s_t, i_t, r_t) = {} & \frac{\partial}{\partial s}\beta s_t i_t P(s_t, i_t, r_t) - \frac{\partial}{\partial i}(\beta s_t i_t - \gamma i_t)P(s_t, i_t, r_t) \\
& + \frac{1}{2}\frac{\partial^2}{\partial s^2}\frac{1}{N}\beta s_t i_t P(s_t, i_t, r_t) \\
& - \frac{1}{2}\frac{\partial^2}{\partial i^2}\frac{1}{N}(\beta s_t i_t - \gamma i_t)P(s_t, i_t, r_t) \\
& - \frac{\partial^2}{\partial s\partial i}\frac{1}{N}\beta s_t i_t P(s_t, i_t, r_t),
\end{aligned} \tag{9}$$

which is equivalent to

$$\frac{\partial}{\partial t}P(s_t, i_t, r_t) = -\frac{\partial}{\partial x}[\dot{A}(s_t, i_t, r_t)P(s_t, i_t, r_t)] + \frac{1}{2}\frac{\partial}{\partial x}\frac{\partial}{\partial x}[\dot{\Sigma}(s_t, i_t, r_t)P(s_t, i_t, r_t)] \tag{10}$$

Where

$$\dot{A}(s_t, i_t, r_t) = \begin{pmatrix} -\beta s_t i_t \\ \beta s_t i_t - \gamma i_t \\ \gamma i_t \end{pmatrix} \tag{11}$$

and

$$\dot{\Sigma}(s_t, i_t, r_t) = \frac{1}{N} \begin{pmatrix} \beta s_t i_t & -\beta s_t i_t & 0 \\ -\beta s_t i_t & \beta s_t i_t + \gamma i_t & -\gamma i_t \\ 0 & -\gamma i_t & \gamma i_t \end{pmatrix} \tag{12}$$

Following **?**, Eq. 10 is a Fokker-Planck equation corresponding to a diffusion process that is a solution of

$$d\dot{z}_t = \dot{A}(\dot{z}_t)dt + LdB_t^{\dot{Q}_t^d} \tag{13}$$

Here, we follow the formalism of **?** 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 $L\dot{Q}_t^d L = \dot{\Sigma}$:

$$\dot{Q}^d(s_t, i_t) = \frac{1}{N} \begin{pmatrix} \beta s_t i_t & 0 \\ 0 & \gamma i_t \end{pmatrix} \quad and \quad L = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \tag{14}$$

Equation 13 can be transposed in the natural scale of $z_t = [S_t, I_t, R_t]^T$, with $A = N\dot{A}$ and $Q^d = N^2 \dot{Q}^d$:

$$dz_t = A(z_t)dt + LdB_t^{Q^d} \tag{15}$$

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 **?**. The authors demonstrate that the dynamic of a density-dependent Markov jump process can be approximated with Eq. 15 with dBt being a multivariate Brownian motion with diffusion matrix $\dot{Q}^d = diag\{r^{(k)} z_t^{\chi(k)}, k \in \mathcal{R}\}$, and L being the rectangular stoichiometric matrix which columns are the stoichiometric vectors $l^{(k)}$ with $k \in \mathcal{R}$. Additionally, the drift component $\dot{A}(t)$ is determined by:

$$\dot{A}(z_t) = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) \dot{z}_t^{\chi(k)} \tag{16}$$

Lastly, the resulting general expression for $\dot{\Sigma}$ is the following:

$$\Sigma(\dot{z_t}) = L\dot{Q}^d L' = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) \dot{z}_t^{\chi(k)} l^{(k)\prime} \tag{17}$$

Diffusion approximation of the environmental stochasticty

This section focuses on environmental stochasticity. In this perspective, we consider for the sake of illustration an infinite population leading to a deterministic behaviour in the absence of diffusing parameters:

$$dz_t = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt \tag{18}$$

In the case of the SIR model:

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

8

The framework proposed in **?** introduces environmental stochasticity by replacing deterministic time increments dt by random, 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:

$$
\begin{cases}
dS_t = -\beta S_t \dfrac{I_t}{N} d\Gamma_t \\[2mm]
dI_t = \beta S_t \dfrac{I_t}{N} d\Gamma_t - \gamma I_t dt \\[2mm]
dR_t = \gamma I_t dt
\end{cases}
\tag{20}
$$

We propose to derive a Gaussian formulation of epidemic models with *white* 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:

$$
\begin{cases}
dS_t = -\beta S_t \dfrac{I_t}{N} dt - \sigma \beta S_t \dfrac{I_t}{N} dB_t^{(1)} \\[2mm]
dI_t = (\beta S_t \dfrac{I_t}{N} d\Gamma_t - \gamma I_t) dt + \sigma \beta S_t \dfrac{I_t}{N} dB_t^{(1)} \\[2mm]
dR_t = \gamma I_t dt
\end{cases}
\tag{21}
$$

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

$$
dz_t = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt + L^e dB_t^{Q^e}
\tag{22}
$$

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 $Q^e = diag\{(\sigma^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)})^2, k \in \mathcal{R}^e\}$ 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 $\varphi : \mathcal{R}^e \to [1 : n_g]$ 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:

$$
\begin{cases}
dz_t = \sum_{k \in \mathcal{R}} l^{(k)} r^{(k)}(z_t, \theta) z_t^{\chi(k)} dt + L dB_t^Q \\[2mm]
dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta) dt + L^{\theta_t} dB_t^{Q^{\theta_t}}
\end{cases}
\tag{23}
$$

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

$$L = \left(\begin{array}{cc} L^d & L^e \end{array} \right) \quad and \quad Q = \left(\begin{array}{cc} Q^d & 0 \\ 0 & Q^e \end{array} \right) \tag{24}$$

On one hand, $Q^d = diag\{r^{(k)}(z_t, \theta)z_t^{\chi(k)}, k \in \mathcal{R}\}$ and $L^d = [l^{(1)}, .., l^{(c)}]$ accounts for demographic stochasticity. With regards to *white* environmental noise, $L^e = [l^{(k)}]_{k \in \mathcal{R}^e}$ 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 $Q^g = diag\{(\sigma^{(p)})^2, p \in [1 : n_g]\}$, 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 of 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 $N \to \infty$. Hence, the reference Markov jump process formalism quickly becomes intractable for other than small populations. The authors of **?** 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 **?** and extend it to the general framework for compartmental models proposed in SSM.

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:

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

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

$$M_i = \left(\begin{array}{c} z_t^{(i)} \\ \{n_k\}_{\chi(k)=i} \ \overline{n}_i \end{array} \right) \quad \text{(multinomial coefficient)}$$

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

In this section we will consider the more general class of state space models evolving in continuous time, with discrete observations. This definition encompasses systems of ordinary or stochastic differential equations as well as compartmental models and combinations thereof. The state of the system at time t is noted 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 getting to state x_{i+1} from state x_i cannot be computed. However, we consider that it is possible to simulate trajectories from the augmented prediction density $p(x_{i:i+1}|x_i,\theta)$. 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 (**?**).

As suggested by the now classic motto "all models are wrong, but some are useful" (**?**), 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 (**?**). 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 (**?**). 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 (**?**).

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 up to a multiplicative constant through 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 (**?**). In its current version, the *SSM* library provides a tailored and more efficient solution 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) (**?**). 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 a system conditioned on parameters θ and available observations $y_{1:n}$. They are targeted to problems where the target density 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 (**?**). If J is the number of particles, this algorithm can provide a sample $\tilde{x}_{0:n}$ from $\hat{p}_{pf}^J(x_{0:n}|y_{1:n})$, and an unbiased estimator $\hat{p}_{pf}^J(y_{1:n}|\theta)$ of $p(y_{1:n}|\theta)$. Under mild assumptions, the authors of **?** and **?** have proved the following properties:

$$\|\hat{p}_{pf}^J(x_{0:n}|y_{1:n}) - p(x_{0:n}|y_{1:n})\| \leq \frac{C_n}{J} \tag{25}$$

$$Var(\frac{\hat{p}_{pf}^J(y_{1:n}|\theta)}{p(y_{1:n}|\theta)}) \leq \frac{D_n}{J}$$

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 precision increasing as $O(J^{1/2})$.

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

Extended Kalman Filter

An approximate solution to the filtering problem for state space models is provided by the Extended Kalman Filter (EKF) algorithm (**??**). 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 filtered density for $p(x_i|y_{0:i})$ characterised by its mean m_t and covariance C_t. It provides with a deterministic and biased estimate $\hat{p}_{EKF}(y_{1:n}|\theta)$ of the marginal likelihood.

Note that in Algorithm 2, only one observation 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.

3.3 Full inference of paths and parameters

The central methodology utilised in SSM 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 to 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 generically 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

---
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

---

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:

$$\pi(x^i) K(x^{(i-1)}|x^{(i)}) = \pi(x^{(i-1)}) K(x^{(i)}|x^{(i-1)}) \tag{26}$$

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 (**??**). Here is one way to compute the ESS:

$$ESS(\{x^{(1)}, x^{(2)}, .., x^{(N)}\}) = \frac{N}{1 + 2 \sum_{k=1}^{k^{max}} Correl(\{x^{(1)}, .., x^{(N-k)}\}, \{x^{(k)}, .., x^{(N)}\})} \tag{27}$$

This indicator, along with other diagnostic tools proposed in the CODA package (**?**), are crucial in assessing 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 (**??**). At each iteration of the chain, a proposed value x^* is sampled from an importance distribution $q(.|x^{(i)})$, and accepted with probability:

$$1 \wedge \frac{\pi(x^*) q(x^{(i)}|x^*)}{\pi(x^{(i)}) q(x^*|x^{(i)})} \tag{28}$$

Otherwise, x^* is rejected and $x^{(i+1)}$ 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, although other choices are also possible. It is generally observed that increasing the dimension of x decreases the acceptance probability.

For example, the random walk Metropolis is based on a Metropolis-Hastings step 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% *(?)*

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

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

When the target distribution is not a multivariate normal density, these results are generally 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 $a \in [0; 1[$:

$$\lambda_{i+1} = \lambda_i \times a^i(AccRate_i - 0.234) \tag{29}$$

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 **?** is based on the following importance sampling distribution:

$$q(.|x^{(i)}) = \alpha \mathcal{N}\left(x^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(0)}\right) + (1-\alpha)\mathcal{N}\left(x^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(i)}\right), \tag{30}$$

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 $p(x_{0:n}, \theta|y_{1:n})$. The augmented path $x_{0:n}$, in particular, is a high-dimensional object. It contains the state of the system at each point of the discretised time $(t_0, t_0 + \delta, t_0 + 2\delta, \ldots, t_n - \delta, t_n)$. 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 (**?**). 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 $\hat{p}_{pf}^{J}(y_{1:n}|\theta)$ estimated by a particle filter conditioned on θ. The authors of **?** have shown that for any J the algorithm was asymptotically exact for a given discretisation of time. Under classic assumptions, when the number of iterations N^{θ} tends to infinity:

$$\|\hat{p}_{pf}^{J}(x_{0:n}^{(i)}, \theta^{(i)}|y_{1:n}) - p(x_{0:n}, \theta|y_{1:n})\| \to 0 \quad as\ i \to \infty \tag{31}$$

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_0^q is a central issue: we will explore in the next subsection a mean to automate 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 $\hat{p}_{pf}^{J}(y_{1:n}|\theta^{(0)})$ and sample $x_{0:n}^{(0)}$ from $\hat{p}_{pf}^{J}(x_{0:n}|y_{1:n}, \theta^{(0)})$

for $i = 1$ to N^{θ} **do**

  Sample θ^* from $q(.|\theta^{(i)})$

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

  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 $x_{1:n}^{(i+1)}$

end for

An alternative solution is the SMC^2 algorithm presented in **?**. 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. It additionally provides a estimate of $p(y_{1:n})$ under a given model, which can be used for model selection through the Bayes rule.

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 $p(\theta|y_{1:n})$. When MCMC chains are initialised from an arbitrary position, the likelihood classically follows an increasing trend before it stabilises, indicating that the chain has converged to a mode. During this phase, θ also follows a transient convergence phase. Naturally, due to this non-stationarity the generated samples are strongly correlated and weakly informative. 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 and dedicated tools, that we will now present. In addition, the pMCMC implementation proposed in *SSM* 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 that can be used 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). In *SSM*, 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 a 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, $p(\theta|y)$) through reflection, expansion, contraction or reduction transformations of the polygon. The use of this algorithm in SSM directly relies on its implementation in the GNU Scientific Library (**?**). 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 $p(\theta|y)$ 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 *SSM* 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 *SSM*.

The optimisation routines based on algorithm 5 require the use of ode or sde approximations 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 as a first step to initialise the Markov Chain that subsequently explores the posterior density:

```
cat theta.json | ./ksimplex | ./kmcmc | ./pmcmc
```

As a consequence, potential discrepancies between the likelihoods induced under the different formalisms will only have serious consequences in critical cases. In such situations, it is possible to rely on iterated filtering (`mif`), an asymptotically exact and plug-and-play solution to the frequentist problem of maximising the marginal likelihood $p(y_{1:n}|\theta)$ (**???**). This approach has already been used for numerous 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 **?**, 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 (**??**), 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 permit easier routine maximisation of the posterior density.

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

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

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

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

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

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

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

        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, $\tilde{\theta}^{(j)}(t_i) \sim \mathcal{N}\left(\tilde{\theta}^{(k_j)}(t_{i-1}), a^{m-1}(t_i - t_{i-1})\Sigma_\theta\right)$

        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

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 (see 2.4.2), 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 adaptation of the `kmcmc` algorithm.

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
_____

4   Application examples

The following examples have been obtained by routine application of the following sequence of algorithms:

```
cat theta.json | ./ksimplex | ./kmcmc | ./pmcmc
```

They illustrate how the results presented in **?** can be easily reproduced using *SSM* to capture the time-varying drivers of epidemics. They also introduce two novel applications of epidemic modeling to increase our understanding of past epidemics and predict their future evolution and serve decision-making.

4.1   Exploring the past: the Medieval Black Death

The analysis of historic and recent records of plague epidemics have brought into relief surprising discrepancies between the characteristics of past and present epidemics that we had so far been attributing to plague. For example, the authors of **?** have shown that while current laboratory-confirmed casesgenerally occur between November and April, Medieval Black Death epidemics used to burst between April and October. Have we been wrongly attributing the Black Death epidemics to the bubonic and pneumonic plagues? This question remains open, and we are simply going to illustrate here how mechanistic models could be used to provide further insight into the characteristics of present and historic epidemics.

We will be looking at two time series of 1665 epidemics in the UK, each indicating the monthly number of deaths caused by plague in London and Eyam. Contrasting these two cases is not only interesting due to the population size difference between London and Eyam, that had respectively 460000 and 350 inhabitants at the beginning of the epidemics, but also due to the peculiar story of the city of Eyam **?**. As some villagers started to die from plague, the clergyman William Monpesson decided to isolate the village in order to protect the neighbouring cities of Northern England. During one year, Eyam sacrificed and lived in quarantine. Food was cautiously supplied so that villagers did not starve. Yet, at the end of the epidemic 250 people had died.

This story sheds a particular light on the following time series that can be found in the *Bills of Mortality*. Since 1932, these records had been filled by English doctors,

who were required to monitor the deaths due to tuberculosis, small pox, measles, French pox, and plague:
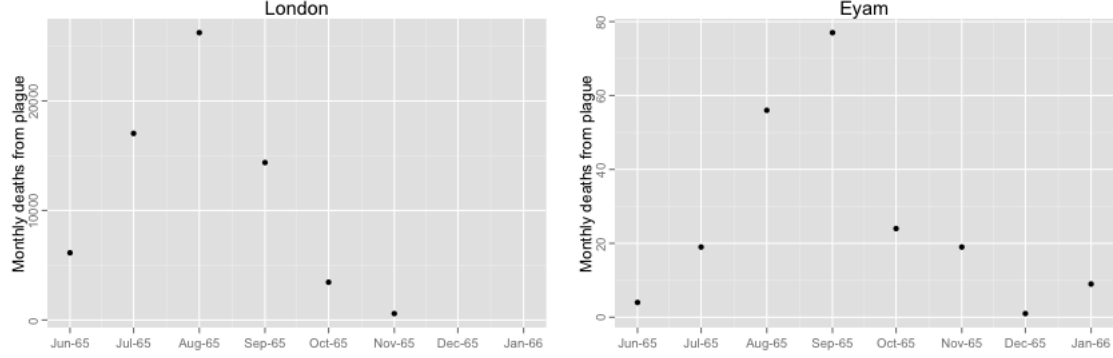

Figure 1: Monthly number of deaths caused by Plague in London and Eyam

To analyse this dataset we use the model introduced in section 2.3.1. The timing and amplitude of seasonal forcing, as well as initial conditions, reproduction rates in each city, and life expectancy with plague, are estimated. We make no assumption on the type of plague at stake, allowing the life expectancy after infection to lie between one and seven days (respectively corresponding to pneumonic and bubonic plague). The resulting estimates of the transmission potential of plague in each city, as well as life expectancy with plague, are the following:
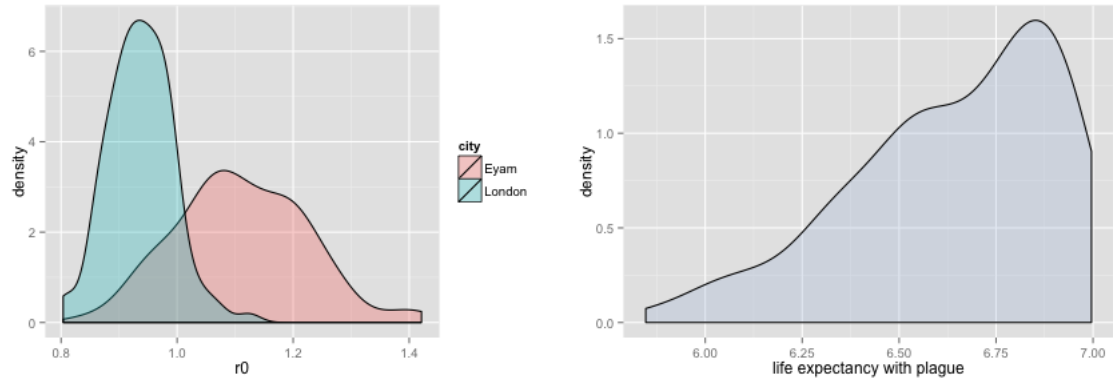

Figure 2: Posterior densities of R0 in each city, and of the life expectancy with plague.

These results provide information that could not have been inferred from direct observation of the time series of deaths in each city. First, they suggest that the isolation and living conditions in Eyam lead to a higher transmissibility of the disease. Furthermore, life expectancy after infection appears to be close to one week, suggesting that this epidemic was a bubonic plague rather than a pneumonic plague. The latter seems to be confirmed by historical records.

For the sake of transparency, and to foster further explorations of this problem and the data provided in **?**, the following repository provides the means to easily reproduce the presented results: https://github.com/JDureau/plague-UK-1665.

4.2 Real-time monitoring of the H1N1 transmission rate

We have proposed in **?** a generic solution to monitor the transmission rate of a pathogen during an epidemic from incomplete and uncertain measures of its spread among a population.

The proposed methodology relies on compartmental models in which some parameters are allowed to vary over time following a diffusion. It then helps to understand what are the underlying and unobserved causes of observed epidemic dynamics. We illustrated this approach on a time series of H1N1 cases recorded in London during the 2009 pandemic:
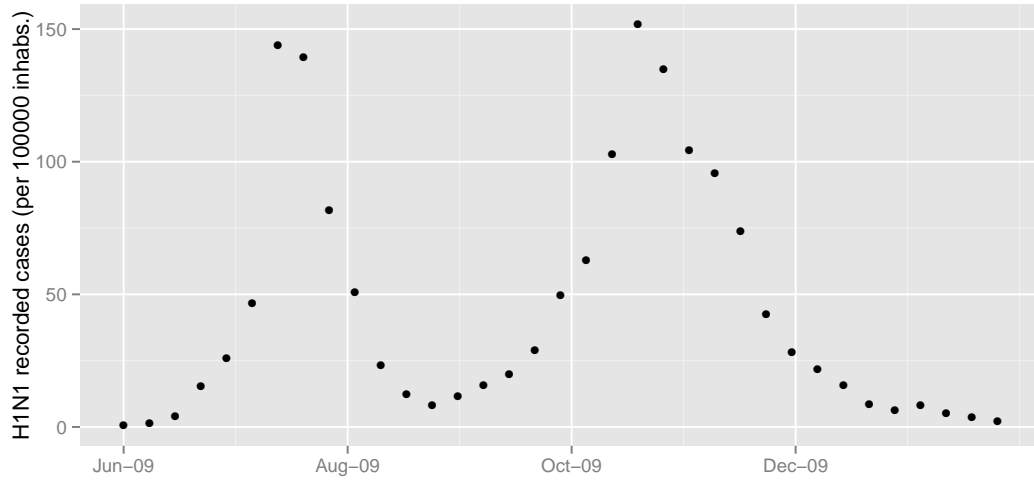

Figure 3: Weekly number of H1N1 cases recorded in London during the 2009 epidemic (courtesy of the Health Protection Agency)

The unusual shape of the epidemic trajectory, exhibiting two peaks, reflects variations of extrinsic quantities that drive its evolution. Holidays, and their subsequent impact on the frequency at which people meet and infect each other, provide a natural explanation to the decline of the first and second waves. However, the additional role of climate on the transmissibility of influenza is also debated, and media may have played an important role on individual awareness and behaviour. We show that according to the model we use, the effective transmission rate of H1N1 evolved in the way illustrated by Fig 4.

These results confirm that holidays have been the main driver of the epidemic, and further quantifies their impact on the transmission rate of influenza. For example, it shows that the impact of summer holidays is about twice more important than fall holidays, providing an indication on the potential impact of closing schools as a mean to mitigate an epidemic.

The following repository provides the means to easily reproduce the presented results: https://github.com/JDureau/H1N1-London-2009.
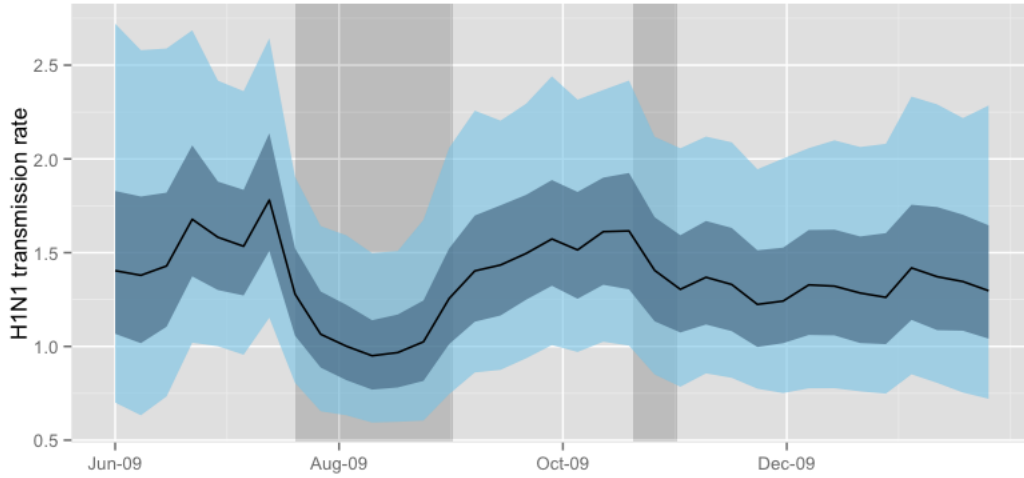
Figure 4: Estimated trajectory of the effective transmission rate. Darker grey areas indicate holiday periods. Light and dark blue areas respectively indicate 95% and 50% credible intervals

4.3 How many severe cases of dengue in Madeira next year?

Until last year, dengue had disappeared from the European continent. The last epidemic goes back to 1927-1928, in Greece. However, concerns of a return of dengue in Europe had started to rise in the recent years, due to the dissemination of *Aedes albopictus* across European countries. This mosquito plays a central role in dengue transmission, as it serves as a vector for the virus.

In September 2012, a first epidemic occured in Europe. 2159 cases were recorded over 3 months in the Portughese island of Madeira. Among these case, a few individuals were hospitalised for mild symptoms of fever but no severe case has been recorded. Fig. 5 shows the corresponding time series, supposing that no cases have been recorded between February and July 2013, which is when this document was written.

Although multiple and crucial aspects of dengue transmission are still to be explored, some epidemiologists argue that severe cases are more likely to correspond to secondary infections (**?**). After having previously been infected with one of the 4 dengue strains, an individual that is re-infected with another strain would have a much higher probability of developing severe symptoms as hemorragic dengue fever. If we follow this assumption, and consider that all infections that occured in 2012 were primary infection, there is a risk for severe cases in 2013 if any of the primary infected gets re-infected.

We illustrate here how mechanistic models can be used to forecast coming epidemics while reflecting the different sources of uncertainty. As described in 2.3.3, we have extended a multi-strain model that had been introduced in **?** to study dengue dynamics in South-East Asia.

Under these assumptions, the data from the 2012 epidemic can be used to reconstruct the current state of immunity of the population of Madeira, and to project its evolution. Following a Bayesian approach allows to reflect the available information on the respective lengths of the infectivity and cross-immunity periods, as well as the uncertainty on the proportion of asymptomatics and initial state of the population im-
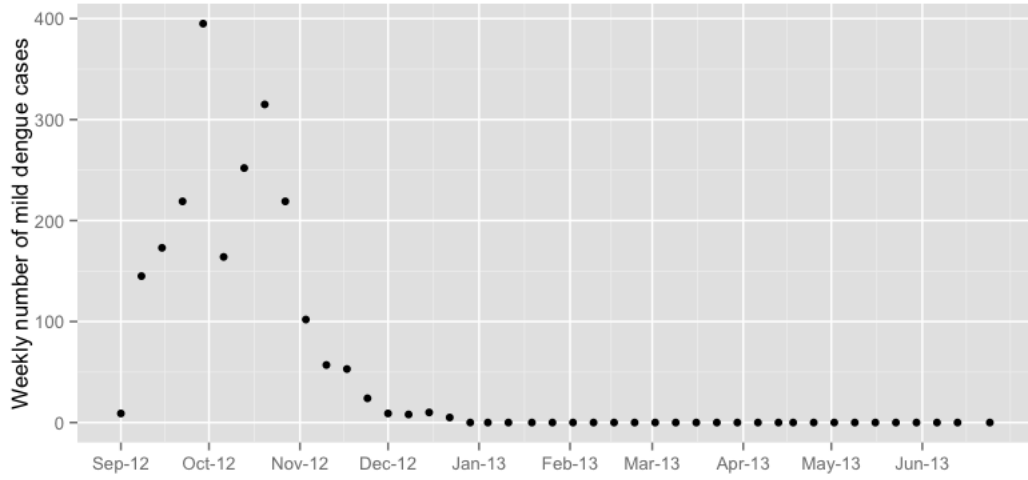
Figure 5: Weekly number of dengue cases recorded in Madeira

munity. We nonetheless consider that only less than 5% of the population had already
been infected with dengue before September 2012. Accordingly, the predicted number
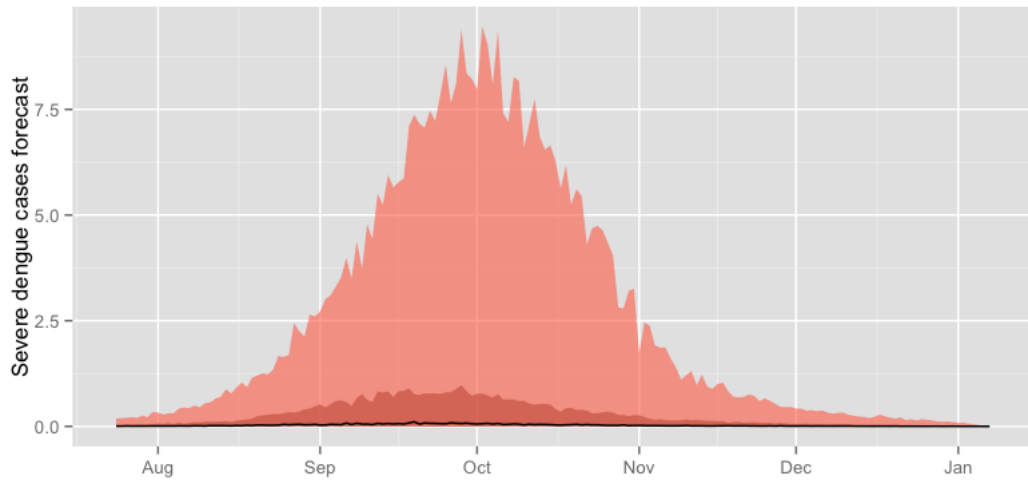of sever dengue cases occurring each week is illustrated in Fig. 6.

Figure 6: Forecasted evolution of the weekly number of sever cases in Madeira.

Naturally, these preliminary results shall be explored further to strengthen the ev-
idence provided by this analysis. To that end, the following repository provides the
means to easily reproduce the presented results: https://github.com/JDureau/dengue-
Madeira-2012.