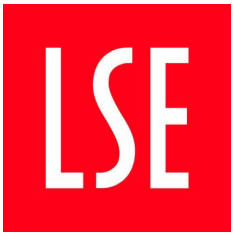

Bayesian Inference for Indirectly Observed Stochastic Processes, Applications to Epidemic Modelling

JOSEPH DUREAU

SUPERVISED BY KOSTAS KALOGEROPOULOS AND WICHER BERGSMA

Thesis submitted to the Department of Statistics
of the London School of Economics and Political Sciences
for the degree of Doctor of Philosophy

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others in which case the extent of any work carried out jointly by me and any other person is clearly identifed in it. The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgment is made. This thesis may not be reproduced without the prior written consent of the author. I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

The times after Copernicus were times in which there were great debates about whether the planets in fact went around the sun along with the earth, or whether the earth was at the centre of the universe and so on. Then a man named Tycho Brahe *evolved a way of answering the question. He thought that it might perhaps be a good idea to look very very carefully and to record exactly where the planets appear in the sky, and then the alternative theories might be distinguished from one another.*

Richard Feynman (1965). *The Character of Physical Law*.

Acknowledgements

I would first like to thank Kostas Kalogeropoulos for his patience, enthusiasm and constant implication in supervising my first steps as a statistician; it has truly been a pleasure. I also thank John Edmunds and Arnaud Doucet, who have taken the time to read, examine and discuss this manuscript.

I am grateful to Bernard Cazelles for introducing me to the world of epidemiology, inspiring this project and supporting me during the third year of the PhD. I also wish to thank all the researchers with whom I have collaborated on the questions explored in this thesis: Marc Baguelin, Marie-Claude Boily, Peter Vickerman, Michael Pickles and Alex Beskos. This work has also been possible thanks to enriching discussions with Sebastien Ballesteros, Anton Camacho, and Nikos Demiris, who have nourished my reflexion and motivation.

I have enjoyed being part of the Statistics department of the London School of Economics, and particularly thank Irini Moustaki, Chris Skinner, Whicher Bergsma, Pauline Barrieu and Ian Marshall for financial and human support.

I am deeply endebted to Cyrille Henry-Bonniot, Jérémie Triolet and Marine Ricci for making my visits to London restful, joyful and stimulating. Many thanks as well to all the other London folks: Hugo Maisonhaute, Flavia Giammarino, Yehuda Dayan, Mai Hafez, Raphaelle Metras and Guillaume Fournié.

I wish to express my gratitude to my friends and family for having born with me during these years of doubt and obsession. At last, this would not have been possible without the permanent support of Lucile. Thank you.

Abstract

Stochastic processes are mathematical objects that offer a probabilistic representation of how some quantities evolve in time. In this thesis we focus on estimating the trajectory and parameters of dynamical systems in cases where only indirect observations of the driving stochastic process are available. We have first explored means to use weekly recorded numbers of cases of Influenza to capture how the frequency and nature of contacts made with infected individuals evolved in time. The latter was modelled with diffusions and can be used to quantify the impact of varying drivers of epidemics as holidays, climate, or prevention interventions. Following this idea, we have estimated how the frequency of condom use has evolved during the intervention of the Gates Foundation against HIV in India. In this setting, the available estimates of the proportion of individuals infected with HIV were not only indirect but also very scarce observations, leading to specific difficulties. At last, we developed a methodology for fractional Brownian motions (fBM), here a fractional stochastic volatility model, indirectly observed through market prices.

The intractability of the likelihood function, requiring augmentation of the parameter space with the diffusion path, is ubiquitous in this thesis. We aimed for inference methods robust to refinements in time discretisations, made necessary to enforce accuracy of Euler schemes. The particle Marginal Metropolis Hastings (PMMH) algorithm exhibits this *mesh free* property. We propose the use of fast approximate filters as a pre-exploration tool to estimate the shape of the target density, for a quicker and more robust adaptation phase of the asymptotically exact algorithm. The fBM problem could not be treated with the PMMH, which required an alternative methodology based on reparameterisation and advanced Hamiltonian Monte Carlo techniques on the diffusion pathspace, that would also be applicable in the Markovian setting.

Table of Contents

List of tables

List of Figures

Notations

θ: vector of constant parameters

d: dimension of θ

y_i: observation made at time t_i

n: number of observations available

w_t: state of a system at time t ($w_t = \{x_t, z_t\}$)

x_t: state of the underlying driving process at time t

z_t: complementary compotents of a system at time t. In an epidemic modeling setting, z_t corresponds to the absolute number of individuals in each compartment. The vector containing the proportion of individuals in each compartment is noted $\dot{z}_t$

$w_{0:n}$: path of the system between times t_0 and t_n

$x_{0:n}$: path of the driving stochastic process between times t_0 and t_n

$\mathbb{P}_x$: prior distribution defined over $x_{0:n}$

S_t: absolute number of susceptibles at time t

s_t: proportion of susceptibles at time t

E_t: number of exposed individuals at time t

I_t: number of infectious individuals at time t

R_t: number of resistent or retired individuals at time t

β: effective contact rate

β_t: time-varying effective contact rate

k: symptoms development rate

γ: recovery rate

$r^{i,j}$: rate of reactions converting individuals of type i into individuals of type j

$\dot{r}^{i,j}$: normalised rate of reactions converting individuals of type i into individuals of type j, i.e. $r_{i,j}(z_t) = N\dot{r}_{i,j}(\dot{z}_t)$

$\mathcal{R}$: ensemble of couples (i, j) corresponding to states between which the model allows transfers

$\mathcal{R}^e$: ensemble of couples (i, j) corresponding to states between which the model allows transfers with stochastic rates

m: card($\mathcal{I}$)

c: number of compartments

$k^{(i,j)}$: stoichiometric vector of reaction (i, j)

$d\Gamma$: gamma increment, used to model environmental noise

$x_t^{\Gamma^{(i,j)}}$: components of the stochastic process corresponding to gamma increments for reaction (i, j)

$x_t^{\theta_t}$: components of the stochastic process corresponding to time-varying parameters

$x_t^{N_{(i,j)}}$: components of the stochastic process corresponding to the number of occurrences of reaction (i, j)

Q: diffusion matrix

Q^{θ_t}: diffusion matrix of the SDE followed by the time-varying parameters

μ^{θ_t}: drift term of the SDE followed by the time-varying parameters

Q^e: diffusion matrix of the SDE used to approximate the environmental stochasticity

Q^d: diffusion matrix of the SDE used to approximate the demographic stochasticity

J: number of particles used in a Sequential Monte Carlo algorithm

N^θ: number of iterations of an algorithm

δ: discretisation timestep

$p^{tr}_{M\to F}$: transmission probability from male to female in an unprotected act

$p^{tr}_{F\to M}$: transmission probability from female to male in an unprotected act

$NbActs$: mean number of acts per client

$Cond_{eff}$: condom efficacy per act

$NbClients^{HR}$: mean number of clients per month for high-risk sex workers

$NbClients^{LR}$: mean number of clients per month for low-risk sex workers

FSW: female sex worker

S^{LR}: number of susceptible low-risk FSWs

S^{HR}: number of susceptible high-risk FSWs

S^{M}: number of susceptible clients

I^{LR}: number of infected low-risk FSWs

I^{HR}: number of infected high-risk FSWs

I^{M}: number of infected clients

CU_t: time-varying condom use

μ_F: average length of commercial sex activity as a FSW

μ_M: average length of commercial sex activity as a client

α: average life expectancy

$prev^{obs}_i$: observed prevalence at time t_i

$prev^{model}_i$: modeled prevalence at time t_i

$\pi(.)$: MCMC target density

T: transition kernel

ESS: Effective Sample Size

$q(.|.)$: importance sampling density

Σ^q: covariance of the importance sampling density in random walk MCMCs

λ: scaling coefficient of the random walk steps

a: cooling rate of the adaptation scheme

p: momentum variable in the Hamiltonian setting

q: position variable in the Hamiltonian setting

t_H: time variable in the Hamiltonian setting

L: number of leapfrogs

u_t: asset price at time t

v_t: volatility at time t

σ_u: volatility of the SDE followed by u_t

σ_v: volatility of the SDE followed by v_t

μ_u: asset price asymptote

μ_v: volatility asymptote

κ: rate of the OU process followed by the fractional stochastic volatility

ρ: leverage parameter

dB_t^H: fractional Brownian motion

H: Hurst parameter

CHAPTER 1

Introduction

The topic of this thesis is the estimation of the full paths and parameters of stochastic processes in settings were only indirect observations are available. This observation scheme prevents the derivation of direct formulas providing information on the uncertain paths and parameters. A possible approach, explored in this thesis, is to contrast simulated trajectories of the system to the data. We concentrate on the problems encountered in epidemiology: in this field, statistical methodology provides with options that are limited, yet not fully exploited by practitioners. Hence, we explore means to diminish the computational burden of the current state-of the art Bayesian inference methods for indirectly observed stochastic processes, while also concentrating on practicality. In addition, we address the question of using diffusion processes to estimate key parameters evolving in a potentially fully unknown manner. We consider different observational schemes, and explore the applications of this approach in public health. Ultimately, we consider the problem of inference for non-Markovian stochastic processes through the example of fractional stochastic volatility models.

1.1 Modelling the spread of infectious diseases in human populations

This Section describes the questions explored by mathematical modellers in epidemiology, as an introduction to the questions treated in this thesis. We first describe the variety of challenges posed by infectious diseases, with examples of what has been achieved and of some persistent issues and rising threats in the field of public health. We then motivate the use of mathematical models to encompass the complex and nonlinear dynamics of epidemics. In addition, we comment on the use that can be made of models in this setting, stressing their utility and limitations. Lastly, we present the mathematical formalisms used to model the spread of epidemics among a population, concentrating on population-level

compartmental models.

1.1.1 Diverse situations and challenges

Infectious diseases, and the questions of design, monitoring and evaluation of public health interventions will be recurrent in this thesis. To start on a positive note on this matter, we could cite the example of smallpox, the first and only infectious disease affecting humans to ever be eradicated (Fenner et al., 1988). This major achievement was endorsed by the World Health Assembly on May 8th, 1980 (Breman and Arita, 1980). Yet, only thirteen years earlier the World Health Organisation (WHO) estimated that smallpox was still killing 2 million persons per year. It took a global and coordinated effort of surveillance, containment and vaccination to definitively free the world from smallpox. Nonetheless, infectious diseases remain responsible for about 26% of the total number of deaths every year, with striking disparities: infectious diseases cause 47% of deaths in African countries, and about 8% in Europe (WHO, 2012b). The socio-economic dimension of the problem appears clearly, and has motivated studies on poverty traps appearing through the two-ways retroaction mechanisms that link disease burden to economic health (Bonds et al., 2010).

Ambitious programs are being carried out, among which the project of polio eradication that started in 1985. The number of polio cases has greatly decreased, from over 350,000 cases in 1988 (Arita et al., 2006) to 214 confirmed cases in 2012 (WHO, 2012a). Yet, in 2012 polio remained endemic in Afhanistan, Nigeria and Pakistan, requiring a continuous and careful effort of vaccination to potentially achieve eradication (Kew, 2012). Vaccines may not be the only way to go: recent studies suggest that testing systematically all individuals over 15 every years old, and treating every person diagnosed as HIV positive with antiretroviral therapy (ART) even before symptoms start to develop could allow HIV elimination (zero incidence) in the coming decades (Granich et al., 2009). However, this idea is being debated and there is still a long way to elimination: the number of individuals accessing to treatment for HIV was multiplied by sixteen between 2001 and 2010, but it is estimated that over 2 million individuals are still infected every year, of which 75% live in Sub-Saharan countries (WHO, 2012b). This region is also severely affected by malaria, that caused over 600,000 deaths worldwide in 2010, among which 86% were children aged less than five years old (WHO, 2012b). Insecticide-treated bed nets and mosquitoes proliferation control remain the main means to limit the epidemic, until efficient vaccines are discovered (Raghavendra et al., 2011).

There is a risk that economic difficulties may lead to lower the efforts and investments towards the fight against infectious diseases (Suhrcke et al., 2011). However, new challenges will appear in the coming years. For example the tuberculosis virus that already kills over 1.4 million individuals every year is developing increasing resistance to current

antibiotics (Shah et al., 2007). Besides, perspectives of climate change raise concerns about an increase in the number of disease-bearing mosquitoes, that could augment the number cases of malaria and other mosquito-borne diseases as dengue or yellow fever (McMichael et al., 2006). Additionally, emerging diseases that appear for the first time, or may have appeared previously but are now being transmitted at a very rapid pace, are an important challenge to surveillance and global response systems. They generally appear following the adaptation of an animal virus or strain to humans, that constitute a *naive* population to invade (Jones et al., 2008). Careful surveillance systems are built to avoid the reproduction of dramatic past experiences. New strains of influenza, for example, killed 50 millions of individuals in the 1918 Spanish flu pandemic, between 1.5 and 2 million in 1957-58 (Asian flu) and about 1 million in 1968-69 (Hong-Kong flu) (Hilleman, 2002). Yet, the authors of Jones et al. (2008) have shown that the emergence of diseases has intensified over the past 60 years, in part due to the development of intensive breeding, transportation and urbanisation.

1.1.2 The dynamics of epidemics

The previous overview illustrates the need to achieve a good understanding of disease transmission dynamics, in order to make an optimal use of limited financial and human resources. We know, for a start, that the potential explosiveness of a disease can be measured by its basic reproduction number R_0 that corresponds to the mean number of secondary infections a newly infected individual causes in a fully susceptible population (Anderson et al., 1992). When R_0 is over 1, each individual starts infecting more than one individual and an epidemic bursts. The basic reproduction number can be seen as the ratio between the rate of infection, i.e. the number of individuals infected each day, and the length of the infectious period. This illustrates how evolutive compromises can be established: diseases with short infectivity periods generally lead to high levels of infectivity, we call them acute. Chronic diseases that induce life-long infectivity can allow for lower levels of infectivity. A related quantity is the efficient reproduction number, that additionally accounts for the proportion of remaining susceptible individuals at a given state of an epidemic (Anderson et al., 1992). It is denoted R_t and can be expressed in the following way:

$$R_t = \frac{\text{transmission rate}}{\text{length of infectivity period}} \times \text{proportion of remaining susceptibles} \tag{1.1}$$

Epidemics spread until R_t becomes smaller than 1, which can occur simply due to the depletion of individuals susceptible to be infected (they are generally referred to as susceptible individuals), but also to vaccination campaigns or to changes in the transmission rate due to prevention measures, holidays, climate, etc. In particular, epidemic waves generally wane before having infected the total population even in the absence of any external

intervention. Another consequence of these mechanisms is that not all the population needs to be vaccinated to maintain R_t below one (Anderson et al., 1992). This is called herd immunity.

Acute diseases are generally explosive and lead to quick epidemic bursts (Read and Keeling, 2006). Renewal of the susceptible population is then necessary for these diseases to persist in the population. For very transmittable diseases as measles, births suffice to maintain it at an endemic level. Other acute diseases evolve along time to escape existing immunity induced by previous infections, which leads to competing mechanisms between different strains, or versions of the virus. For example, various strains of influenza or dengue coexist (Ballegooijen and Boerlijst, 2004). A consequence of the absence or presence of immune escape is that the vaccine for measles used in 1960 would still be efficient nowadays (Rota and Bellini, 2003), while the vaccine for the H3N2 strain of influenza has been updated 19 times between 1972 and 2001 (Hay et al., 2001). These mechanisms classically lead to oscillating dynamics which chaotic properties have been illustrated (Bolker and Grenfell, 1993; Gupta et al., 1998; Stollenwerk et al., 2012). In this context, forecasting the impact of an intervention as a vaccination campaign, for example, is a challenging task. In conjunction to these complex dynamics, the effective reproduction rate R_t is highly contextual, which is also a key complexifying factor. Transmission potentials vary across the population, for example, in relation to age, risk behaviours, hygiene practices, etc. This heterogeneity can induce high rates of disease prevalence in some population clusters (Cori et al., 2009; Choi et al., 2010; Melegaro et al., 2011). Transmission potentials also vary in time, for example holidays generally tend to strongly decrease the numbers of contacts kids make with each other every day (Wallinga et al., 2006). Besides, transmission varies as awareness to a pathogen increases (Ferguson, 2007): this evolution was for example observed among the homosexual community when HIV emerged in the 80's (Cazelles and Chau, 1997).

At last, understanding and quantifying all these mechanisms is made harder by the stochastic nature of diseases transmissions, and by the fact that most of the transmission process is unobserved. Typically, the only available observations are noisy estimates of the number of individuals developing symptoms for a specific disease, obtained through medical diagnosis records (incidence data). It can alternatively consist of estimates of the proportion of the population infected with a given disease at a given time (we call this prevalence data). Observation regimes can vary depending on the type of disease and state of the health surveillance system, but incomplete observation is and will probably remain ubiquitous. Indeed, it is very rare to know whom infected whom, and when, neither do we generally know when individuals stop being infectious. Furthermore, all individuals do not necessarily develop symptoms or consult a doctor. Such characteristics of the

observation process itself are specifically hard to infer: for example, the estimates that can be found in the literature of the ratio between the proportion of asymptomatic and symptomatic cases for cholera range from 3 to 100 (King et al., 2008). Modelling these indirectly observed stochastic processes is an attempt to achieve a better understanding of epidemic dynamics and to inform public policies. It is a holistic approach aiming at capturing the elements that play a role on the diffusion of diseases among the population, their interactions and their dynamics.

1.1.3 Inference, a hypothetico-deductive approach

Dynamic models can be seen as a hypothesised probabilistic relation between the trajectories of a system, noted $w_{0:n}$, and constant quantities grouped in a parameter vector θ. They induce the definition of a joint probability density $p(w_{0:n}, \theta)$. Reflecting the fact that epidemics are only partially observed, the vector w_t can be separated into its driving underlying stochastic components x_t, and complementary ones z_t ($w_t = \{x_t, z_t\}$). 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 $p(w_{0:n}|\theta)$. The likelihood of the data $y_{1:n}$ under a given scenario is given by an observation model defining $p(y_{1:n}|\theta, z_{0:n})$.

As suggested in the previous section, 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 $p(x_{0:n}, z_{0:n}, \theta|y_{1:n})$, information can be deducted respectively on the indirectly observed driving processes $x_{0:n}$ and on the parameter vector θ. The validity of this information shall be critically examined from a three-fold perspective. First, the uncertainties associated with the data collection should be reflected in the observation model. Then, the limitations of the model itself should be acknowledged and questioned, while considering the practical feasibility of proposing extensions to palliate the imperfections of the model. A minimal condition requires the output of the model to be able to fit the available observations of mechanisms they are meant to reproduce (Gelman and Shalizi, 2012). At last, the information derived regarding $x_{0:n}$ and θ, reflected by the discrepancies between their marginal prior and posterior densities, should not be considered as hard truth but rather as plausible and testable hypothesis (Popper, 2002). When two models $\{M_1, M_2\}$ appear to efficiently fit the data, it is common practice to either discriminate between them from their Bayes factor ($K = p(y_{1:n}|M_1)/p(y_{1:n}|M_2)$) or using penalised likelihood indicators as the Akaike, Bayesian or Deviance Information Criteria (respectively noted AIC, BIC or DIC) (Akaike, 1973; Schwarz, 1978; Spiegelhalter

et al., 2002).

We have presented the process of inference under a Bayesian perspective, and will continue to do so in the remaining of this thesis, but a similar hypothetico-deductive approcach is generally followed when a frequentist perspective is taken. The main difference lies in the absence of a prior density over θ reflecting known or hypothesised information (although boundaries are classically enforced), and in the interpretations that are drawn from the exploration of the likelihood surface $p(y_{1:n}|x_{0:n}, z_{0:n}, \theta)$.

1.1.4 Applications in public health

On top of the plausible hypothesis inferred from available data regarding transmission dynamics and the values of the key parameters at stake, simulating scenarios of what will happen in the future or what would happen under a given intervention is central in public health applications (Garnett, 2002). Naturally, such forecast can only be as good as the hypothesised models they rely on, so thorough examination of the latter is crucial, and a critical perspective should be kept regarding projection results.

First, model forecasts are necessary to evaluate the impact of an intervention, which is not only important in order to ensure that it has achieved its goals, but also to inform future interventions (Pickles et al., 2010). Indeed, a decrease (or an increase) in the number of recorded cases is not necessarily the sign of an efficient (or inefficient) intervention. Epidemics have their own natural dynamic that should be accounted for (Brisson and Edmunds, 2006; Boily et al., 2007; Pickles et al., 2010; Baguelin et al., 2012). Hence, quantifying the impact of an intervention, that is the difference between what happened and what would have happened had the intervention not occurred, requires the simulation of a hypothetical scenario corresponding to the absence of intervention. The work presented in Baguelin et al. (2012) illustrates this approach: it simulates epidemics of seasonal influenza in the absence of vaccination campaigns, in order to assess the cost-efficiency of the yearly vaccination programme. Naturally, the validity of the generated results is subject to the quality of the transmission models used in this study, but they remain the only mean to assess the level of herd immunity indirectly induced by the vaccination of a segment of the population. This approach to impact evaluation has led to the definition of guidelines that encourage to thoroughly explore sources of uncertainty (both from models and parameters) and assess the robustness of the results drawn from simulations (Drummond and Jefferson, 1996; UNAIDS, 2000; Beutels et al., 2002). A different approach to evaluate interventions is the use of randomised trials. Originally, pharmaceutical clinical trials randomly assigned a treatment or a placebo version of it to a series of individuals, which would provide strong evidence of its efficiency or inefficiency. Because infectious diseases spread among the population, randomisation needs to be made at a community level (hence the

terminology Community Randomised Trials, CRT). These approaches are appealing, but they are generally very expensive, raise some ethical concerns, and require very careful logistics (Boily et al., 2007; Banerjee and Duflo, 2009). Hybrid approaches combining CRTs and modelling-based impact estimates are being experimented on interventions against HIV led in Botswana, Tanzania, Zambia and South Africa (Boily et al., 2012). In addition, models can be used to monitor in real time the impact of epidemics, providing estimates of the effect size that can be expected at the end of the intervention, contributing to informed real-time adjustments (Boily et al., 2012).

Alternatively, model forecasts are used as a tool to design optimal interventions by predicting a priori what could be the outcome of different intervention scenarios. An example is given in Choi et al. (2010), in which different vaccination strategies against human papillomavirus (HPV) are explored. This preliminary study was rendered critical by the high price of vaccines against HPV, and by the risk for HPV infections to slowly evolve into cervical cancer. Scenarios were simulated from a variety of models reflecting the uncertainties regarding the prevalence of HPV infections, patterns of sexual partnership, accuracy of screening as well as duration of infectiousness and immunity. The resulting conclusions were consequently conjectural, but they nevertheless permitted to discriminate a priori between different approaches, and to evaluate the sensitivity of the achieved conclusions. As such, it provides a robust evidence base for decision-making.

In some cases, the number of cases averted is not the only criteria for which an intervention is evaluated. When a campaign aims at modifying individual behaviour, estimating how the latter have evolved is crucial. However, the direct monitoring of this evolution can be biased or uncertain. This question will be addressed in Chapter 2 of this thesis, on the specific problem of estimating the evolution of condom use by female sex workers in Southern India following the intervention led by the Bill & Melinda Gates Foundation.

1.1.5 Mathematical formulation of epidemic dynamics

The canonic Susceptible - Infectious framework

The core of the dynamics of infectious diseases epidemics is the mechanism of individuals infecting each other. This fundamental principle leads to the distinction of infected and infectious individuals, and individuals that could be infected and become infectious who are called susceptibles. The way and the rate at which these infections occur can vary, but they can generally be characterised through the mean number of persons an infectious individual would infect over a given short period of time if all the other individuals were susceptible. It is called the transmission or effective contact rate, and reflects both the number of contacts and their nature, their propensity to effectively induce transmissions

of the disease. The effective contact rate is generally noted β.

Supposing that individuals can only be susceptible or infectious, the population can be divided into two groups also termed compartments. The size of each compartment at a time t is an integer value respectively noted S_t and I_t, and the size of the population is noted N. If the effective transmission rate is believed to be constant and uniform among the population, the transmission process can be described in the following way:

Reaction	**Effect**	**Rate**
Infection	$(S_t, I_t) \to (S_t - 1, I_t + 1)$	$\beta \frac{S_t}{N} I_t$

The term *reaction* will be used generically to refer to transformations of the system where the number of individuals in each compartment changes. For each reaction, its effect will be given along with its hypothesised rate of occurrence. In the present case, the rate of transition $\beta \frac{S_t}{N} I_t$ is motivated by the fact that each infectious individual transmits the diseases at a rate β, but only a fraction $\frac{S_t}{N}$ of the individuals receiving the disease are susceptible to become infected. This model makes the implicit assumption that infectiousness is memoryless: it does not depend on how long individuals have been infected for. A wide variety of models are derived from the Susceptible-Infectious (SI) framework. They are still based on the same principle of homogeneous status and behaviour for all individuals belonging to a compartment, independently of the time they have spent in it. Nevertheless, there are different ways in which the realism of compartmental models can be improved.

A first limitation of the SI framework is homogeneity: structured compartmental models are built to account for variations of given properties among the population. This heterogeneity can be related to a variety of factors including age, geographical location, risky behaviours, immunity conferred by past infections, etc. A structured version of the SI model would be based on the partition of the population into k groups labeled by indexes $\{1, 2, ..., k\}$. Hence, if infections among all groups are allowed a structured model can take the form of k^2 reactions:

Reaction	**Effect**	**Rate**
Infection in group 1 by group 2	$(S_t^{(1)}, I_t^{(2)}) \to (S_t^{(1)} - 1, I_t^{(2)} + 1)$	$\beta^{(1,2)} \frac{S_t^{(1)}}{N} I_t^{(2)}$
...	...	...
Infection in group i by group j	$(S_t^{(i)}, I_t^{(j)}) \to (S_t^{(i)} - 1, I_t^{(j)} + 1)$	$\beta^{(i,j)} \frac{S_t^{(i)}}{N} I_t^{(j)}$
...	...	...
Infection in group k by group $k-1$	$(S_t^{(k)}, I_t^{(k-1)}) \to (S_t^{(k)} - 1, I_t^{(k-1)} + 1)$	$\beta^{(k,k-1)} \frac{S_t^{(k)}}{N} I_t^{(k-1)}$

Although disease transmissions are the central process in epidemic dynamics, accounting only for the Susceptible and Infectious statuses is generally unrealistic. Biologically,

the transmission of a disease allows the invasion of the organism of a susceptible individual by a pathogen. The phase of invasion is a continuous and progressive process, which duration is not always negligible. For a given period of time, an exposed individual is neither susceptible nor infectious. This phase is called the latent period, and the corresponding compartment is generally denoted by E. It is also very common that individuals only remain infectious for a given period of time, until they recover. The recovery process can induce a permanent or temporary immunity, during which individuals are neither susceptible, exposed nor infectious: they are resistent (R). Models that account for the latent phase and allow for recovery inducing permanent immunity are called SEIR models. They take the following form:

Reaction	**Effect**	**Rate**
Infection	$(S_t, E_t, I_t, R_t) \to (S_t - 1, E_t + 1, I_t, R_t)$	$\beta \frac{S_t}{N} I_t$
Latent phase	$(S_t, E_t, I_t, R_t) \to (S_t, E_t - 1, I_t + 1, R_t)$	kE_t
Recovery	$(S_t, E_t, I_t, R_t) \to (S_t, E_t, I_t - 1, R_t + 1)$	γI_t

Additionally, the lack of memory of the SI framework is also a restricting assumption. The motivation for this assumption is the definition of a Markovian system that will be mathematically more tractable. A consequence of this choice is that, for example, the distribution of the duration of the latent phase in the SEIR is exponentially distributed with mean k^{-1}. The latter is known to be unrealistic, as studies have shown that a bell-shape distribution would be more adapted (Boelle et al., 2011). A classic solution is to artificially introduce p intermediary compartments, for example $E_t^{(1)}, .., E_t^{(p)}$, with leaving rates pk. This process results in an Erlang distribution for the duration of the latent phase, with preserved mean k^{-1} and variance $(pk)^{-1}$. A similar approach can be followed for other quantities, for example the duration of the infectivity period. We will denote SE^pIR a model relying on p compartments to model the latent phase. For illustration purposes, we formulate the SE^2IR model in a similar fashion than previous models:

Reaction	**Effect**	**Rate**
Infection	$(S_t, E_t^{(1)}, E_t^{(2)}, I_t, R_t) \to (S_t - 1, E_t^{(1)} + 1, E_t^{(2)}, I_t, R_t)$	$\beta \frac{S_t}{N} I_t$
Latent phase 1	$(S_t, E_t^{(1)}, E_t^{(2)}, I_t, R_t) \to (S_t, E_t^{(1)} - 1, E_t^{(2)} + 1, I_t, R_t)$	$2kE_t^{(1)}$
Latent phase 2	$(S_t, E_t^{(1)}, E_t^{(2)}, I_t, R_t) \to (S_t, E_t^{(1)}, E_t^{(2)} - 1, I_t + 1, R_t)$	$2kE_t^{(2)}$
Recovery	$(S_t, E_t^{(1)}, E_t^{(2)}, I_t, R_t) \to (S_t, E_t^{(1)}, E_t^{(2)}, I_t - 1, R_t + 1)$	γI_t

We focus in this thesis on population-level compartmental epidemic models. The formalism they define is common when studying epidemics spreading among a population. A classic alternative are individual-based models. They are also based on differentiated statuses as infected, exposed, etc., but the state vector is composed of the state each indi-

vidual, and potentially of details as the time spent in the current state, the history of previous infections, etc. Such models allow for a finer representation of heterogeneity among the population. For example, the effective contact rate of a given group in a population-level model is constant across the group, whereas individual-level models allow for each individual's transmission potential to be sampled from a continuous distribution, and be individual-specific at any time. Such models, however, require more information to be parameterised and often render inference computationally prohibitive unless the population at stake is small. Typical examples where individual-based models have permitted an effective exploration of transmission dynamics can nevertheless be found in the literature. For example, the authors of Cori et al. (2009) study both the temporal variability and social heterogeneity in SARS transmission potentials from Hong Kong data where the number of cases is limited and exactly measured. We could also cite Cauchemez et al. (2011), that explores influenza transmission dynamics during the 2009 influenza pandemic in a semi-rural community of Pennsylvania, where cases have been carefully recorded as well as details on household constitutions and even seating charts for some classes.

An opposite example of a situation where the use of individual-based models is both feasible and highly profitable is presented in Demiris and O'Neill (2005). It exhibits how information on the effective contact rates among a structured population can be inferred from only final outcome data indicating how many individuals have been infected in each household of an isolated and small community. Due to the very limited nature of data, it is beneficial to reconstruct the history of the epidemic with precision through a complete graph of whom infected whom.

Markovian jump processes as a reference model

We will now consider a general formulation to describe transmission dynamics from a population-level compartmental perspective that will serve as a reference in the remainder of the thesis. The description of the system is contained in a state vector z_t, which components represent the number of individuals in each compartment of the model. For example, in the SIR model, $z_t = (S_t \;\; I_t \;\; R_t)^T$. When the state vector no longer represents absolute numbers of individuals but fractions of the population in each compartment, the notation $\dot{z}_t$ will be used. Every reaction is associated with a couple of indexes (i, j), respectively corresponding to the states individuals leave and go to. All the couple of indexes are grouped in $\mathcal{R}$, of size m. Births and deaths can be defined in the same manner by artificially introducing source and sink compartments. The fluxes between compartments i and j are represented by a stoichiometric vector $k^{(i,j)} \in \mathbb{Z}^m$. For example, infections in an SIR model correspond to an individual leaving the susceptible (S) compartment and going to the infected (I) compartment. Hence, the stoichiometric vector $k^{(1,2)}$ associated

with this reaction is $(\ -1 \quad 1 \quad 0\ )^T$. Similarly, the vector $k^{(2,3)}$ associated with recoveries is $(\ 0 \quad -1 \quad 1\ )^T$. Reaction (i, j) occurs at a rate $r^{(i,j)}$, that depends both on the state of the system z_t and on the parameter values contained in θ. In the SIR model, for example, $\theta = (\beta, \gamma, S_0, I_0, N)$. With these notations, a general description of epidemic models can be made:

Reaction	**Effect**	**Rate**
reaction 1	$z_t \to z_t + k^{(i_1,j_1)}$	$r^{(i_1,j_1)}(z_t, \theta)$
...	...	...
reaction l	$z_t \to z_t + k^{(i_l,j_l)}$	$r^{(i_l,j_l)}(z_t, \theta)$
...	...	...
reaction m	$z_t \to z_t + k^{(i_m,j_m)}$	$r^{(i_m,j_m)}(z_t, \theta)$

Which defines a Markovian jump process:

Reference Markovian jump process epidemic model

$$\begin{aligned} P(z_{t+\delta} = z_t + k^{(i,j)} | z_t) &= r^{(i,j)}(z_t, \theta)\delta + o(\delta) \qquad \text{for any } (i,j) \in \mathcal{R} \\ P(z_{t+\delta} = z_t | z_t) &= (1 - \sum_{(i,j)\in\mathcal{R}} r^{(i,j)}(z_t, \theta)\delta) + o(\delta) \end{aligned} \tag{1.2}$$

This mathematical formalism is able to reproduce stochasticity observed among finite populations, that is termed demographic stochasticity. When transition rates can be written as $r(z_t, \theta) = N\dot{r}(z_t/N, \theta) = N\dot{r}(\dot{z}_t, \theta)$, they are said to be density-dependent. We will assume in this thesis that all transition rates are density-dependent, as it is generally the case when modelling epidemics and because transition rates can always be rewritten into this form.

The infinite population approximation

When studying systems with large populations or for theoretical studies, demographic stochasticity is commonly neglected. Under the density-dependance assumption, when the size N of the populations tends to infinity the dynamic of the normalised state variable $\dot{z}_t$ converges in distribution on any bounded time interval (Ethier and Kurtz, 1986) to the one of an Ordinary Differential Equation:

Epidemic model under
infinite population approximation

$$\frac{d\dot{z}_t}{dt} = \sum_{(i,j)\in\mathcal{R}} k^{(i,j)}\dot{r}^{(i,j)}(\dot{z}_t, \theta) \tag{1.3}$$

This system of equations, resulting from the mass action principle, is analog to models used in pharmaco-kinetics to monitor reactions among molecules in a solution. It is not always the case in epidemiology that populations are large enough to apply this approximation. Besides, even in large populations it may happen that the size of some compartments, typically I_t, gets close to 0 which is a limiting case of the approximation. In such cases, demographic stochasticity may play an important role.

Multinomial approximation of demographic stochasticity

The reference Markovian jump process model is generally not a tractable solution to account for demographic stochasticity. Indeed, it apprehends each reaction occurring in the system, which number increases with the number of individuals. As a consequence, tracking or simulating exactly the evolution of the system requires an increasingly fine discretisation of time, which quickly becomes computationally prohibitive (Breto et al., 2009). Different solutions have been proposed, among which the approximation introduced in Breto et al. (2009) that accounts for the possibility of multiple reactions over short periods of time through a multinomial approximation. If $N_t^{(i,j)}$ is the number of occurrences of reaction (i,j) up to time t, and $\Delta N_t^{(i,j)} = N_{t+\delta}^{(i,j)} - N_t^{(i,j)}$, a Markov chain can be defined in the following manner through its infinitesimal generator:

Epidemic model under
multinomial approximation

$$P\left(\Delta N_t^{(i,j)} = n^{(i,j)}, (i,j) \in \mathcal{R} | z_t\right) = E\left[\prod_{i=1}^{c}\left\{M^{(i)}\left(1 - \sum_{k\neq i} p^{(i,k)}\right)^{\overline{n}^i} \prod_{j\neq i}\left(p^{(i,j)}\right)^{n^{(i,j)}}\right\}\right] + o(\delta)$$

$$z_{t+\delta} = z_t + \sum_{(i,j)\in\mathcal{R}} k^{(i,j)} \Delta N_t^{(i,j)} \tag{1.4}$$

We use the following notations, with $r^{(i,j)} \equiv 0$ if $(i,j) \notin \mathcal{R}$:

$$\overline{n}^i = z_t^{(i)} - \sum_{k\neq i} n^{(i,k)}$$

$$p^{(i,j)} = p^{(i,j)}\left(r^{(i,j)}(z_t,\theta)\delta\right) = \left(1 - \exp\left\{-\sum_{k\neq i} r^{(i,k)}\delta\right\}\right) r^{(i,j)}\delta / \sum_{k\neq i} r^{(i,k)}\delta$$

$$M^{(i)} = \begin{pmatrix} z_t^{(i)} \\ n^{(i,1)} \ldots n^{(i,i-1)}\ n^{(i,i+1)} \ldots n^{(i,c)}\ \overline{n}^i \end{pmatrix} \quad \text{(multinomial coefficient)}$$

The authors of Breto et al. (2009) have proved that this infinitesimal generator characterises a properly defined Markov chain in the limit $\delta \to 0$. It allows efficient simulation independently of the size of the population, and has been used to introduce an additional source of stochasticity: *white* environmental noise.

Stochastic rates as a mean to capture environmental stochasticity

As stated in Section 1.1, although models can be made more complete by extending the number of compartments used to represent the state of an epidemic, they remain an approximation of a complex reality. Some factors may always be missing, or the role of these factors may be simplified. Drivers as climate or human behaviour can vary in an unpredictable way, modifying the rate at which the reactions of the model occur. For example, temperature and absolute humidity, that have been shown to be negatively correlated with transmission of influenza, exhibit intense variability (Shaman and Kohn, 2009). This source of variability is called environmental stochasticity. The authors of Breto et al. (2009) have motivated the modelling of this source of uncertainty through an integrated noise process with independent, stationary and nonnegative increments, that can be used as a multiplicative white noise over the transition rates. Although there are different alternatives, a Gamma distribution is generally used to model these independent increments.

Under the multinomial approximation setting, with $\Delta\Gamma_t^{(i,j)} = \Gamma_{t+\delta}^{(i,j)} - \Gamma_t^{(i,j)}$ being a noise increment with Gamma distribution of mean δ and standard deviation $\sqrt{\delta}\sigma^{(i,j)}$, a white environmental noise has been introduced in the following manner, by multiplying transition rates $r^{(i,j)}$ of a series of reactions $\mathcal{R}^e \in \mathcal{R}$ by random increments $\Delta\Gamma_t^{(i,j)}$ instead of δ:

$$\underline{\text{Epidemic model with stochastic rates}}$$
$$\underline{\text{under multinomial approximation}}$$

$$P\left(\Delta N_t^{(i,j)} = n^{(i,j)}, (i,j) \in \mathcal{R} | z_t\right) = E\left[\prod_{i=1}^{c}\left\{M^{(i)}\left(1 - \sum_{k\neq i} p^{(i,k)}\right)^{\overline{n}^i} \prod_{j\neq i}\left(p^{(i,j)}\right)^{n^{(i,j)}}\right\}\right] + o(\delta)$$
$$z_{t+\delta} = z_t + \sum_{(i,j)\in\mathcal{R}} k^{(i,j)} \Delta N_t^{(i,j)} \tag{1.5}$$

We use the following notations, with $r^{(i,j)} \equiv 0$ if $(i,j) \notin \mathcal{R}$ and $\Delta\Gamma_t^{(i,j)} \equiv \delta$ if $(i,j) \in$

$\mathcal{R} \setminus \mathcal{R}^e$:

$$\overline{n}^i = z_t^{(i)} - \sum_{k \neq i} n^{(i,k)}$$

$$p^{(i,j)} = p^{(i,j)}\left(r^{(i,j)}(z_t,\theta)\Delta\Gamma_t^{(i,j)}\right) = \left(1 - \exp\left\{-\sum_{k\neq i} r^{(i,k)}\Delta\Gamma_t^{(i,j)}\right\}\right) r^{(i,j)}\Delta\Gamma_t^{(i,j)} / \sum_{k\neq i} r^{(i,k)}\Delta\Gamma_t^{(i,j)}$$

$$M^{(i)} = \begin{pmatrix} z_t^{(i)} \\ n^{(i,1)} \ ... \ n^{(i,i-1)} \ n^{(i,i+1)} \ ... \ n^{(i,c)} \ \overline{n}^i \end{pmatrix} \quad \text{(multinomial coefficient)}$$

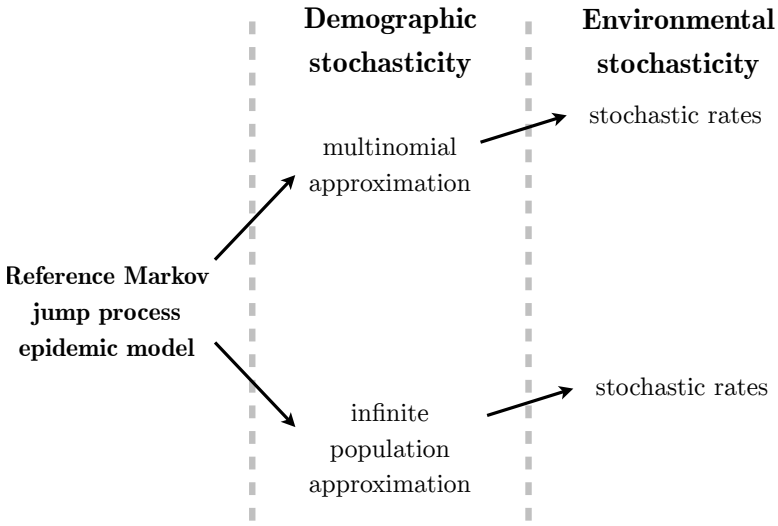

Figure 1.1: Representation of the different epidemic models introduced in the thesis

Again, the authors of Breto et al. (2009) have proved that this infinitesimal generator characterises a properly defined Markov chain in the limit $\delta \to 0$. Alternatively, environmental noise can also be introduced in the infinite population setting, defining a stochastic differential equation driven by Gamma noise:

Epidemic model with stochastic rates
under infinite population approximation

$$d\dot{z}_t = \sum_{(i,j)\in\mathcal{R}^e} k^{(i,j)}\dot{r}^{(i,j)}(\dot{z}_t)d\Gamma^{(i,j)} + \sum_{(i,j)\in\mathcal{R}\setminus\mathcal{R}^e} k^{(i,j)}\dot{r}^{(i,j)}(\dot{z}_t)dt \tag{1.6}$$

Variations in climate and human behaviour, however, cannot always be modelled as white noise as they have been shown to exhibit complex seasonal and inter-annual patterns (Meehl, 1987) influencing epidemic dynamics (Viboud et al., 2004). We explore in this thesis an alternative description of environmental noise by allowing certain parameters to vary in time in order to capture the influence of time-varying drivers of epidemics. A graphical representation of the different mathematical formulations of epidemic models that have just been introduced is provided in Figure 1.1. When the size N of the population tends to infinity, it is expected that the models based on the multinomial approximations converge to the ones defined under the infinite population approximation. A complete theoretical study of the nature of this convergence, however, has not yet been contributed to the literature to our knowledge.

Modelling observation regimes

Observation regimes of epidemics can take a variety of forms. Examples of data used to study their course are incidence, prevalence, and death counts. A substantial amount of uncertainty is generally associated with the collection of such data, principally due to the following reasons:

non-report: depending on disease severity, there can be a substantial proportion of individuals that do not consult a doctor, which leads to unrecorded cases.

asymptomatics: a proportion of the individuals infected with a disease will be infectious without developping any symptom. For influenza, virus inoculation to voluntary individuals have revealed that the proportion of asymptomatics could vary between 25% and 40% (Carrat et al., 2008).

false positives: different diseases can induce similar symptoms, which leads to uncertainties in diagnosis records. For example, it has been estimated that only 40% of patients that consult doctors for influenza-like symptoms are truly infected with influenza (Finkenstadt et al., 2005).

A general model for incidence data is the binomial distribution. The number of recorded cases on week i is denoted by y_i, α accounts for the rate of false positives and ρ is the reporting rate (which incorporates asymptomatics). Under these notations, the probability of recording y_i new cases over week i, if Inc_i is the true number of new infections having occurred over the same week, is given by $Binom((1-\alpha)y_i, Inc_i, \rho)$. In the $SEIR$ model, for example, incidence is defined as $\int kE_t dt$: cases can only be detected after the onset of symptoms. Similar observation models can be derived for prevalence or attack rates. Under infinite population approximations, a lognormal observation density can be used. To

preserve the Markovianity of the model and of the observation process, an artificial additional component can be added to the state vector to account for the accumulation of new cases over an observation period. For example, the state vector for an SEIR model that will be observed through incidence can be formulated as $\{S_t, E_t, I_t, R_t, Inc_t\}$.

1.2 Bayesian inference for indirectly observed stochastic processes

In this section, we propose a general formulation for indirectly observed stochastic processes, and present a series of notions that will serve as a basis for further development along the thesis. The problem we address is the efficient exploration of the joint posterior density $p(x_{0:n}, \theta|y_{0:n})$, where $x_{0:n}$ is the path of the driving stochastic process, and θ is a vector of constant parameters playing a role in the dynamic and observation of the system. More precisely, we will see that for any statistic that needs to be computed from the target density $p(x_{0:n}, \theta|y_{0:n})$, unbiased estimates relying on an approximate density $p_\delta(x^{dis}_{0:n}, \theta|y_{0:n})$ based on a time discretisation with resolution δ can be obtained up to Monte Carlo error. If the inference algorithm is robust to refinements of the resolution δ, both the bias due to discretisation and the variability induced by Monte Carlo error can always be reduced by increasing the computation time. Inference methods for which these sources of error can be rendered negligible will be called *asymptotically exact* in this thesis: in such cases, the hypothetical facts derived from the confrontation of a model to a given dataset are intrinsic to the model. Thus, models can be tested and compared, independently of the utilised inference algorithms.

Lastly, there appears to be a gap between the models and methods that are being used by practitioners to design and evaluate public health interventions, and the models and methods that have been appearing in the literature in the recent years (Brisson and Edmunds, 2006). This observation has motivated the development of plug-and-play algorithms; in this perspective, ease of use and automatic calibration will be central concerns in this thesis.

1.2.1 General framework for indirectly observed stochastic processes

We are interested in studying the paths and parameters of underlying stochastic processes that deterministically drive the trajectory of discretely observed auxiliary variables observed with noise. The driving process will be noted x_t. The auxiliary or complementary components, denoted z_t, are generally needed to obtain a complete description of the system and bridge the gap between the process of interest and the available observations of the system. The concatenation of vectors x_t and z_t form the state vector $w_t = \{x_t, z_t\}$. We allow in this general setting both x_t and z_t to be uni or multi-dimensional, and either real

or integer-valued. We note $x_{0:n}$ and $z_{0:n}$ the corresponding trajectories between times t_0 and t_n. More generally, state variables with subscript i will be an abuse of notation for their value at time t_i. The model can then be defined in the following way:

$$\begin{cases} x_{0:n} \sim p(.|\theta) \\ z_{0:n} = f(x_{0:n}, \theta) \\ p(y_i|z_i, \theta) = h_t(z_i, \theta) \end{cases} \tag{1.7}$$

This general formulation encompasses all the problems explored in this thesis. It naturally includes the case of directly observed stochastic processes, by setting $z_t = x_t$, although in that case specific solutions may be more efficient than the ones explored in this thesis. An assumption often made in this context imposes the Markov property on w_t. Markovian state-space model can simply be defined through their instantaneous dynamic and the observation process:

$$\begin{cases} x_{t+\delta} \sim p(.|x_t, z_t, \theta) \\ z_{t+\delta} \sim f(z_t, x_t, \theta) \\ p(y_i|z_{t_i}, x_{t_i}) = h_t(z_{t_i}, \theta) \end{cases} \tag{1.8}$$

Note that we make the potential interaction between z_t and x_t explicit in 1.8, as z_t may carry information about past values of the driving stochastic process. We will now give examples of such processes, starting with the general formulations of compartmental epidemic models introduced in the previous Section.

Compartmental epidemic models

We illustrate here how the different epidemic models that have been introduced in the previous section (see Figure 1.1) can be formulated as indirectly observed stochastic processes. The simplest stochastic model is defined under the infinite population approximation by introducing environmental stochasticity through stochastic rates. Noise increments can be contained in a vector $\left\{x^{\Gamma^{(i,j)}}, (i,j) \in \mathcal{R}^e\right\}$, with $\mathcal{R}^e \subset \mathcal{R}$ containing the indexes of reactions with stochastic rates:

$$\begin{cases} x_t^{\Gamma^{(i,j)}} = d\Gamma_t^{(i,j)} \quad \text{for } (i,j) \in \mathcal{R}^e \\ d\dot{z}_t = \displaystyle\sum_{(i,j)\in\mathcal{R}^e} k^{(i,j)} \dot{r}^{(i,j)}(\dot{z}_t, \theta) x_t^{\Gamma^{(i,j)}} + \sum_{(i,j)\in\mathcal{R}\setminus\mathcal{R}^e} k^{(i,j)} \dot{r}^{(i,j)}(\dot{z}_t, \theta) dt \end{cases}$$

For illustration, we also provide the formulation of epidemic models with stochastic rates under the multinomial approximation in the indirectly observed stochastic process framework. It additionally involves $x_t^{N^{(i,j)}}$, which denotes the total number of times reaction (i,j) has occurred up to time t:

$$\begin{cases} x_t^{\Gamma^{(i,j)}} = d\Gamma_t^{(i,j)} \quad \text{for } (i,j) \in \mathcal{R}^e \quad (x_t^{\Gamma^{(i,j)}} \equiv \delta \;\; if \;\; (i,j) \in \mathcal{R} \setminus \mathcal{R}_e) \\ p(x_{t+\delta}^{N^{(i,j)}} = x_t^{N^{(i,j)}} + n_{(i,j)} | x_t^{N^{(i,j)}}, x_t^{\Gamma^{(i,j)}}, z_t) = E\left[\prod_{i=1}^{c} \left\{ M^{(i)} \left(1 - \sum_{k \neq i} p^{(i,k)}\right)^{\overline{n}^i} \prod_{j \neq i} \left(p^{(i,j)}\right)^{n^{(i,j)}} \right\}\right] + o(\delta) \\ z_{t+\delta} = z_t + \sum_{(i,j) \in \mathcal{R}} k^{(i,j)} x_t^{N^{(i,j)}} \end{cases}$$

We use the following notations, with $r^{(i,j)} \equiv 0$ if $(i,j) \notin \mathcal{R}$:

$$\overline{n}^i = z_t^{(i)} - \sum_{k \neq i} n^{(i,k)}$$

$$p_t^{(i,j)} = p^{(i,j)}\left(r^{(i,j)}(z_t,\theta) x_t^{\Gamma^{(i,j)}}\right) = \left(1 - \exp\left\{-\sum_k r^{(i,k)} x_t^{\Gamma^{(i,j)}}\right\}\right) r^{(i,j)} x_t^{\Gamma^{(i,j)}} / \sum_k r^{(i,k)} x_t^{\Gamma^{(i,j)}}$$

$$M^{(i)} = \begin{pmatrix} z_t^{(i)} \\ n^{(i,1)} \; ... \; n^{(i,i-1)} \; n^{(i,i+1)} \; ... \; n^{(i,c)} \; \overline{n}^i \end{pmatrix} \quad \text{(multinomial coefficient)}$$

Hypoelliptic diffusions

Diffusion processes, that are a classic example of stochastic processes, will be recurrent along this thesis. They correspond to the solutions of stochastic differential equations written in the following manner:

$$dw_t = \mu_t(w_t,\theta)dt + L dB_t^Q(w_t,\theta) \tag{1.9}$$

The matrix L is referred to as a dispersion matrix, and dB_t^Q is a Brownian motion with diffusion matrix Q. This formalism, is also used in Särkkä (2006), and allows the flexible introduction of various types of correlated and uncorrelated noise that will be explored in Chapter 2. Hypoelliptic diffusions correspond to cases where the matrix LQL^T is degenerate. Without loss of generality, these systems can all be reformulated under the Markovian indirectly observed stochastic processes framework (Eq 1.8), with $\tilde{L}\tilde{Q}\tilde{L}^T$ being non-singular:

$$\begin{cases} dx_t = \mu_t^x(x_t,\theta)dt + \tilde{L} dB_t^{\tilde{Q}}(w_t,\theta) \\ dz_t = f(z_t,x_t,\theta) = \mu_t^z(x_t,z_t\theta)dt \end{cases} \tag{1.10}$$

Integrated Brownian Motion

The Integrated Brownian motion (IBM) will be utilised in this thesis as a way to derive a differentiable path from a diffusing object. It can be simply defined in the following manner:

$$\begin{cases} dx_t = \sigma dB_t \\ dz_{t+\delta t} = x_t dt \end{cases} \tag{1.11}$$

Volatility models and extensions thereof

The Stochastic Volatility models, ubiquitous in financial applications, are a classic example of indirectly observed diffusion processes formulated as stochastic differential equations. They model the price u_t of an asset as a diffusion process which volatility term is driven by a quantity v_t that is also stochastic and follows a diffusion process. Prices are considered to be exactly observed at discrete times $(t_i)_{i=1:n}$, defining an observation vector $(y_i)_{i=1:n}$. A general class of Stochastic Volatility models can be written in the following way (Heston, 1993), with $dB_t^{(1)}$ and $dB_t^{(2)}$ being independent Brownian motion increments:

$$\begin{cases} du_t = (\mu_u - \sigma_v(v_t)^2/2)dt + \sigma_u(v_t)dB_t^{(1)} \\ dv_t = \kappa(\mu_v - v_t)dt + \sigma_v dB_t^{(2)} \\ y_i = u_{t_i} \end{cases} \tag{1.12}$$

They can be put in our indirectly observed diffusions framework:

$$\begin{cases} dx_t = \kappa(\mu_v - x_t)dt + \sigma_v dB_t \\ z_t = \displaystyle\int_{t_{i-1}}^{t} \sigma_v(x_s)^2 ds \\ y_i = \mathcal{N}(y_{i-1} + \mu_u(t_i - t_{i-1}) - z_{t_i}/2, z_{t_i}) \end{cases} \tag{1.13}$$

In Chapter 4, we will develop efficient inference methods for fractional versions of the stochastic volatility models, where the stochastic volatility is driven by non-independent fractional Brownian motion increments.

1.2.2 Simulation schemes

Generally, the models that have just been considered are intractable: it is not possible to simulate trajectories from them without relying on time discretisations. We present here the schemes that can be used to simulate paths from the previously mentioned models. As

the discretisation time step tends to 0, it is expected that the probability densities implied by the discretised models converge to the one of the continuous-time model. The validity of this convergence result, as well as the nature and speed of convergence, will vary depending on the model and discretisation scheme.

Simulating paths of stochastic differential equations

One of the simplest ways to sample from a stochastic differential equation, whether it is hypoelliptic or not, is the Euler-Maruyama algorithm (Kloeden and Platen, 1999; Särkkä, 2006). The strong convergence of this algorithm is of order $O(\delta^{1/2})$, but as in the deterministic setting Runge-Kutta algorithms allow for higher orders of convergence.

Algorithm 1 Euler-Maruyama algorithm: sampling from an SDE (Eq. 1.9)

Initialise x_0, $t = 0$
while t<T **do**
  Draw ε from $\mathcal{N}(0, Q(t)\delta)$
  $w^{dis}_{t+\delta} = w^{dis}_t + \mu(w^{dis}_t, \theta)\delta + L\varepsilon$
  $t = t + \delta$
end while

Simulating from the reference Markov jump process epidemic model

Doob and Gillespie have explored means to generate exact simulations from the Markov jump process epidemic model of Eq. 1.2 (Doob, 1942, 1945; Gillespie, 1977). Given the state z_t of the system at a time t, the Doob-Gillespie algorithm relies on the simple probability distribution of the following event:

"no reaction occurs during $[t; t + \tau[$ and reaction i happens at time $t + \tau$"

Algorithm 2 follows, for the exact exploration of $p(x_t|\theta)$ in continuous time. However, its computational cost increases exponentially with the size of the population, through the necessary decrease of time increments τ. Alternative means to simulate from the Markov jump processes used in epidemiology have been explored. Gillespie, again, proposed a $\tau - leap$ algorihtm that serves as a basis for the multinomial approximation introduced in Breto et al. (2009).

Simulating epidemics through multinomial processes

The epidemic model resulting from the multinomial approximation, with stochastic rates (Eq 1.5), is formulated in terms of its infinitesimal generator. From this formulation, an Euler simulation scheme can be directly derived, that is presented in Algorithm 3.

Algorithm 2 Doobs-Gillespie algorithm

Initialise $z(0)$, $t = 0$
while t<T **do**
  Sample a lag τ with $p(\tau) = \left(\sum_{(i,j)\in\mathcal{R}} r^{(i,j)}(z_t,\theta)\right) \exp\left(-\tau \sum_{(i,j)\in\mathcal{R}} r^{(i,j)}(z_t,\theta)\right)$
  Sample a reaction (i,j) knowing that $p^{(i,j)} = r^{(i,j)}(z_t,\theta)/\sum_{(k,l)\in\mathcal{R}} r^{(k,l)}(z_t,\theta)$
  $z(t+\tau) = z(t) + k_i$
  $t = t + \tau$
end while

Algorithm 3 Euler scheme for Markov Jump processes

Initialise $z(0)$, $t = 0$
while t<T **do**
  for each incoming compartment i **do**
    Generate noise increments $\Delta\Gamma_t^{(i,j)} = \Gamma_{t+\delta}^{(i,j)} - \Gamma_t^{(i,j)} \sim \text{Gamma}(\delta/\sigma^2_{(i,j)}, \sigma^2_{(i,j)})$
    Generate process increments
      $(\Delta N_t^{(i,1)}, \ldots, \Delta N_t^{(i,i-1)}, \Delta N_t^{(i,i+1)}, \ldots, \Delta N_t^{(i,c)}, R_t^{(i)}) \sim$
        $\text{multin}(z_t^{(i)}, p_t^{(i,1)}, .., p_t^{(i,i-1)}, p_t^{(i,i+1)}, p_t^{(i,c)}, 1 - \sum_{j\neq i} p_t^{(i,j)})$
      with $p_t^{(i,j)} = (1-\exp\{-\sum_k r^{(i,k)}(z_t,\theta)\Delta\Gamma_t^{(i,k)}\}) r^{(i,j)}(z_t,\theta)\Delta\Gamma_t^{(i,j)} / \sum_k r^{(i,k)}(z_t,\theta)\Delta\Gamma_t^{(i,k)}$
  end for
  for each incoming compartment k **do**
    Set $x_{t+\delta}^{(i)} = R_t^{(i)} + \sum_{j\neq i} \Delta N_t^{(j,i)}$
  end for
  $t = t + \tau$
end while
(for notation simplicity, we consider $\mathcal{R} = \mathcal{R}^e = [1,c]^2$)

Non-Markovian stochastic processes : the fractional Brownian motion example

Because of the non-Markovian property of the fractional Brownian motion, sample paths cannot be iteratively generated as in the previous examples. To obtain a discretised path of length N and step size δ, a direct solution would be to sample a vector $\varepsilon \sim \mathcal{N}(0, I_N)$ and to multiply it by the Cholesky decomposition of the covariance matrix defined by:

$$\begin{cases} \Sigma_{l,m} = \dfrac{\delta^{2H}}{2}|l-m+1|^{2H} + \dfrac{\delta^{2H}}{2}|l-m-1|^{2H} - \delta^{2H}|l-m|^{2H} & if\ l \neq m \\ \Sigma_{l,l} = \delta^{2H} \end{cases} \tag{1.14}$$

However, both obtaining the Cholesky decomposition of Σ and multiplying it by ε have a $O(N^2)$ complexity. An $O(N \log N)$ implementation can be obtained by diagonalising Σ using a Fast Fourier Transform algorithm. This approach was first suggested in Davies and Harte (1987) and will be presented in further detail in Chapter 4.

1.2.3 The Monte Carlo Markov Chain machinery

Founding principles

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

$$\pi(x^i)K(x^{(i-1)}|x^{(i)}) = \pi(x^{(i-1)})K(x^{(i)}|x^{(i-1)}) \tag{1.15}$$

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

$$ESS(\{x^{(1)}, x^{(2)}, .., x^{(N)}\}) = \frac{N}{1 + 2\sum_{k=1}^{k^{max}} Correl(\{x^{(1)}, .., x^{(N-k)}\}, \{x^{(k)}, .., x^{(N)}\})} \tag{1.16}$$

This indicator will be used in this thesis to assess the efficiency of MCMC algorithms utilised to explore the complex and high-dimensional target density $p(x_{0:n}, \theta|y_{0:n})$. Before diving into the presentation of basic and more advanced MCMC algorithms, we introduce the most classic way to define transition kernels that respect the detailed balance condition: the Metropolis-Hastings step (Metropolis et al., 1953; Hastings, 1970). At each iteration of the chain, a proposed value x^* is sampled from an importance distribution $q(.|x^{(i)})$, and accepted with probability:

$$1 \wedge \frac{\pi(x^*)q(x^{(i)}|x^*)}{\pi(x^{(i)})q(x^*|x^{(i)})} \tag{1.17}$$

Otherwise, x^{i+1} is set equal to x^i. The proportion of proposed samples that have been accepted determine the acceptance rate. The Metropolis Hastings step allows the use of

any importance distribution q respecting the irreducibility and aperiodicity conditions, although other choices are also possible. It is generally observed that increasing the dimension of x decrease the acceptance probability. The Gibbs algorithm offers a first solution to this problem: at each iteration of the chain each component (or group of components) of x is updated independently, conditionally on the values of the others. This strategy can be useful when the marginal conditional density of the components being updated is tractable or easy to be sampled from. It can nevertheless induce poor mixing if the components being updated are strongly correlated to the ones they are being conditioned on.

Random walk Metropolis algorithm: theory and adaptive implementation

Algorithm 4 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

The random walk Metropolis is based on Metropolis-Hastings steps using a multivariate normal importance sampling distribution: $q(.|x^{(i)}) = \mathcal{N}(x^{(i)}, \Sigma^q)$ (see Algoritm 4). 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 1.1. *When π is a multivariate normal density, the acceptance rate that maximises the mixing efficiency of the random walk Metropolis algorithm is 23.4% (Roberts et al., 1997)*

Proposition 1.2. *When π is a multivariate normal density, optimal results are achieved by using $\Sigma^q = \frac{2.38^2}{d} \times Cov(\pi)$ (Roberts et al., 1997).*

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

and high acceptance rates. Therefore, the targeted acceptance rate can be approached by iteratively adapting λ with a cooling rate $a \in [0; 1[$:

$$\lambda_{i+1} = \lambda_i \times a^i(AccRate_i - 0.234) \tag{1.18}$$

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

$$q(.|x^{(i)}) = \alpha\mathcal{N}\left(x^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(0)}\right) + (1-\alpha)\mathcal{N}\left(x^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(i)}\right) \tag{1.19}$$

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.

Gradient-driven moves: the example of the Metropolis-Adjusted Langevin Algorithm

The Metropolis-Adjusted Langevin Algorithm (MALA) allows higher mixing performance than the random walk Metropolis algorithm by incorporating information about the gradient of the target distribution. The algorithm is based on the Langevin diffusion with stationary distribution π, defined by the following SDE:

$$dx_t = \frac{1}{2}\nabla \log \pi(x_t)dt + dB_t, \tag{1.20}$$

From this diffusion, an importance sampling distribution can be derived:

$$q(.|x) = \mathcal{N}\left(x^{(i)} + \frac{\varepsilon}{2}\nabla \log \pi(x^{(i)}), \varepsilon\Sigma^q\right) \tag{1.21}$$

This first-order Euler discretisation of the Langevin diffusion introduces an approximation that has to be corrected for by a Metropolis-Hastings step. Again, it has been proved that in the case of a multivariate normal target density the optimal acceptance rate for the MALA is 57.4% (Roberts and Rosenthal, 1998). The classic version of the MALA is presented in algorithm 5.

1.2.4 Exploring sequentially structured distributions: particle filtering

A companion alternative to Monte Carlo Markov Chain algorithms are Sequential Monte Carlo algorithms. Their founding principle is the exploration of probability densities that

Algorithm 5 MALA algorithm

Initialise $x^{(0)}$
for $i = 0$ to N **do**
  Sample $x^* \sim \mathcal{N}(x^{(i)} + \frac{\varepsilon}{2}\nabla \log \pi(^{(i)}), \varepsilon\Sigma^q)$
  Accept x^* with probability $1 \wedge \frac{\pi(x^*)}{\pi(x^{(i)})} \exp(-h(x^*) + h(x^{(i)}))$,
  where $h(x) = \nabla \log \pi(x)^T (\Sigma^q)^{-1} \nabla \log \pi(x)$
end for

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. The particle filter terminology is used when SMC approaches are applied to the exploration of a Markovian, timely structured density. A smoothing version of the most classic particle filter, referred to as Systematic Importance Resampling algorithm, is presented in Algorithm 6 (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:

$$\begin{aligned} \|\hat{p}^J_{pf}(x_{0:n}|y_{1:n}) - p(x_{0:n}|y_{1:n})\| &\leq \frac{C_n}{J} \\ Var(\frac{\hat{p}^J_{pf}(y_{1:n}|\theta)}{p(y_{1:n}|\theta)}) &\leq \frac{D_n}{J} \end{aligned} \tag{1.22}$$

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. In addition, due to the intractability of the models considered in this thesis it is necessary to discretise time with a resolution δ in order to sample from $p(x_{i:i+1}|x_i, \theta)$. When referring to the discrete representation of a path, the superscript dis will be used; for example for a time step δ, the discrete skeleton of x_t will be denoted by $x^{dis}_{0:n} = \{x_0, x_\delta, x_{2\delta}, \ldots, x_{t_n}\}$.

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 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. The EKF, described in Algorithm 7, is based on a gaussian approximation of the observation process ψ (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

Algorithm 6 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 $\alpha^{(j)} = h(y_{k+1}, z_{k+1}^{(j)}, \theta)$
  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 7 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}:
      $\frac{dm_t}{dt} = \mu(m_t, \theta)$
      $\frac{dC_t}{dt} = \nabla\mu(m_t, \theta)C_t + C_t\nabla\mu(m_t, \theta)^T + LQL^T$
    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:
      $m_t = m_t + Kerr$
      $C_t = C_t - KSK^T$
    Update the likelihood $L(\theta) = L(\theta) \times \mathcal{N}(err; 0, S)$
  end for
end for

provides with a deterministic and biased estimate $p^{EKF}(y_{1:n}|\theta)$ of the marginal likelihood, for any discretisation of time with step δ.

1.2.5 Full inference for stochastic processes

Sequential Monte Carlo techniques are a natural framework to explore $p(x_{0:n}^{dis}|y_{1:n}, \theta)$. In order to account for the uncertainties regarding the parameter vector θ, we are aiming for the exploration of the joint posterior density $p(x_{0:n}^{dis}, \theta|y_{1:n})$. Because of the high dimension of the target density, classic MCMC methods fail to be efficient and robust solutions. The particle MCMC algorithm offers a solution relying on the efficiency of particle filters (Andrieu et al., 2010). The principles of its particle marginal Metropolis Hastings version are illustrated in Algorithm 8: the high-dimensional density exploration problem is reduced to the design of an MCMC algorithm over θ, based on the likelihood $\hat{p}_{pf}(y_{1:n}|\theta)$ estimated

by a particle filter conditioned on θ. A second version of the PMCMC, the particle Gibbs algorithm, has been introduced in Andrieu et al. (2010). This algorithm will be presented in further details in Chapter 2 of this thesis. The authors of Andrieu et al. (2010) have shown that the algorithm was asymptotically exact for a given discretisation of time, as under classic assumptions when the number of iterations N^θ tends to infinity:

$$\|\hat{p}^J_{pf}(x^{(i)}_{0:n}, \theta^{(i)}|y_{1:n}) - p(x^{dis}_{0:n}, \theta|y_{1:n})\| \to 0 \quad as\ i \to\ \infty \tag{1.23}$$

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. 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 will be critical in the applicability of the algorithm. In the absence of techniques to efficiently estimate the marginal score $\nabla_\theta \log p(\theta|y_{1:n})$, random walk Metropolis algorithms are generally used. Even in its adaptive form, the parameterisation of its initial covariance matrix Σ^q_0 will be a central issue: we will explore in Chapter 2 a mean to automatise this process, rendering the PMCMC approach plug-and-play.

Algorithm 8 Particle MCMC algorithm (particle marginal Metropolis Hastings version)

Initialise $\theta^{(0)}$.
Use Particle Smoother to compute $\hat{p}(y_{1:n}|\theta^{(0)})$ and sample $x^{(0)}_{0:n}$ from $p(x^{dis}_{0:n}|y_{1:n}, \tilde{\theta})$
for $i = 1$ to N^θ **do**
  Sample θ^* from $q(.|\theta^{(i)})$
  Use PS to compute $L(\theta^*)$ and sample $x^*_{0:n}$ from $\hat{p}(x^{dis}_{0:n}|y_{1:n}, \theta^*)$
  Accept θ^* (and $x^*_{0:n}$) with probability $1 \wedge \frac{L(\theta^*)q(\theta^{(i)}|\theta^*)}{L(\theta^{(i)})q(\theta^*|\theta^{(i)})}$
  Record $\theta^{(i+1)}$ and $x^{(i+1)}_{1:n}$
end for

An alternative solution is the SMC^2 algorithm presented in Chopin et al. (2012). It explores both the probability density of $x_{0:n}$ and θ with an SMC algorithm, starting from the initial target $p(x_0, \theta)$ and progressively incorporating the available observations. The global complexity of this algorithm is similar to the PMCMC, but its ability to automatically adapt the number particles being utilised and to progressively learn from previous samples what could be seen as the equivalent of the covariance matrix Σ^q are promising features. At last, an asymptotically exact and plug-and-play solution to the frequentist problem of maximising the marginal likelihood $p(y|\theta)$ has been proposed in Ionides et al. (2006) and later in Ionides et al. (2011). This approach has been used for several applications in epidemiology, and it can be used to efficiently initialise the Markov chain of the PMCMC by incorporating the prior density into the maximised function of θ.

1.3 Plan and contributions of the thesis

1.3.1 Capturing the time-varying drivers of an epidemic using stochastic dynamical systems

The intrinsic dynamics of epidemics have been extensively studied: predictability remains limited by the intrinsic stochastic and sometimes chaotic nature of diseases transmission, but mathematical formalisms exist to address these aspects. The variability of human behaviour, however, remains an important source of uncertainty in epidemic models. Cross-sectional heterogeneity is a complex matter that has led to the development of structured epidemic models accounting for age classes, contact patterns specific to schools, workplaces, households, etc. Variations of of human behaviour in time, however, have been seldom considered. The need to explore this question further has been expressed in Ferguson (2007) and Funk et al. (2010), based on several examples of spontaneous changes in human behaviour also referred to as prevalence-elasticity of human behaviour. Variations of behaviour in time, as well as other time-varying drivers of epidemics as climate, are generally reflected by the evolution of some key parameters of diseases transmission models. The use of diffusion processes to model the variations of such parameters was proposed in Cazelles and Chau (1997) and Cori et al. (2009). The first study estimated the evolution of risk behaviours among the gay community in Paris from HIV cases data, using an approximate inference framework. In Cori et al. (2009), a more elaborate inference method was used that still required crude discretisation of time, and strongly relied on the exact observation scheme of SARS, the pathogen at stake. We propose in the second Chapter of this thesis a generalisation of these previous studies, and illustrate their applicability on weekly incidence data for influenza in the context of the 2009 H1N1 pandemic. This chapter is mainly based on an article published in *Biostatistics*, written jointly with Kostas Kalogeropoulos (LSE) and Marc Baguelin (Health Protection Agency). The development of an extended framework towards the combination of different types of environmental and demographic stochasticity, with corresponding stochastic differential equation approximations, is a subsequent development of this work.

1.3.2 Estimating changes in condom use from limited HIV prevalence data

Epidemic models are increasingly used to monitor the impact of interventions. In the third Chapter of this thesis, we explore the applicability of the novel methods introduced in Chapter 2 to capture the evolution of prevention behaviours, condom use by female sex workers in this case, from limited prevalence data. This study is conducted in the context of Avahan, the large-scale intervention of the Bill & Melinda Gates Foundation against HIV in India. However, the limited nature of data is common characteristic of large-scale

interventions, due to the cost of regularly and consistently collecting epidemic data. We illustrate that good levels of sensitivity can be achieved, allowing to effectively assess the impact of interventions while controlling the risk of over-estimation. Thereafter, we apply the method to evaluate the evolution of condom use in ten districts targeted by the intervention, and discuss the possibility of integrating survey-based estimates of condom use to reduce uncertainty. This chapter is mainly based on an article currently under revision, written jointly with Kostas Kalogeropoulos (LSE), Peter Vickerman (London School of Hygiene and Tropical Medecine), Michael Pickles (Imperial College) and Marie-Claude Boily (Imperial College). The generalisation of the method to different districts and incorporation of survey-based condom use estimates is a subsequent development of this work.

1.3.3 Bayesian inference with the advanced HMC algorithm

In spite of the recent developments for asymptotically exact and plug-and-play inference (Ionides et al., 2006; Andrieu et al., 2010; Chopin et al., 2012), computational costs remain a limiting factor in the exploration and application of models in epidemiology. Both the PMCMC and SMC^2 algorithms rely on the exploration of conditional densities $p(x_{0:n}|y_{1:n}, \theta)$ for series of values of θ. Hence, the computational complexity of both algorithms is of the order of $O(nJN^\theta)$ with n being the number of available observations, J the number of particles being used, and N^θ the number of considered values of θ. Furthermore, the number of necessary particles in the general case grows exponentially with the dimension of the driving stochastic process (Daum and Huang, 2003). We propose an alternative solution that allows to simultaneously update the full path of the driving process $x_{0:n}$ and the parameter θ through HMC steps with k leapfrogs. In this case, each of the iteration requires the simulation of only k scenarios of the system, with associated scores, instead of hundreds or thousands of particles. Moreover, the HMC algorithm is applicable to non-Markovian dynamic models: we apply it to a fractional stochastic volatility model, and illustrate how good mixing properties can be achieved even on high-dimensional objects such as the paths of indirectly observed stochastic processes. This work has been conducted in collaboration with Kostas Kalogeropoulos and Alex Beskos (University College London).

CHAPTER 2

Capturing the time-varying drivers of an epidemic using stochastic dynamical systems

2.1 Introduction

Epidemic models are often used to simulate disease transmission dynamics, detect emerging outbreaks (Unkel et al., 2012), and assess public health interventions (Boily et al., 2007). In order to capture the dynamics of epidemics, the main focus is generally made on their intrinsically dynamic elements such as the depletion of susceptibles or the population immunity evolution. Nevertheless, there are time-varying extrinsic factors that are crucial to the epidemic course. These may include social cycles (holidays), public interventions and climatic variations. This has been illustrated for diseases such as cholera, malaria (Cazelles et al., 2005; Ionides et al., 2006) or influenza (Shaman and Kohn, 2009). These studies were conducted either by relating climatic and incidence time-series (Cazelles et al., 2005), which does not disentangle the effect of intrinsic and extrinsic factors, or by experimentally assessing the virus resistance in different climatic conditions (Shaman and Kohn, 2009), requiring an extrapolation to the population scale. Overall, the time-varying nature of epidemics poses a challenging statistical problem stressing the need for suitable computational tools (Ferguson, 2007).

This Chapter considers a flexible modelling framework that encompasses time-varying aspects of the epidemic via stochastic differential equations. We aim at providing robust inferential procedures, incorporating the uncertainty associated with key parameters and accounting for data and model limitations. In order to provide an accurate and feasible computational toolbox, we provide Markov Chain Monte Carlo (MCMC) algorithms util-

ising recent developments such as particle MCMC (PMCMC) algorithms (Andrieu et al., 2010) and adaptive techniques (Roberts and Rosenthal, 2009). Modelling aspects are presented in Section 2, while the computational framework is presented in Section 3. In Section 4 we evaluate the performance of the proposed adaptive PMCMC schemes on simulated data. In Section 5 we present various applications of the methodology to the 2009 A/H1N1 pandemic. Section 6 proposes a general modelling framework allowing for the combination of time-varying parameters, stochastic rates and demographic stochasticity. We also provide with corresponding diffusion approximations to facilitate the calibration of this general class of models. Finally, Section 7 concludes with some relevant discussion. Further details can be found in the Supplementary Materials.

This Chapter is directly adapted from an article published in *Biostasticts*, co-written with Kostas Kalogeropoulos (London School of Economics) and Marc Baguelin (Health Protection Agency & London School of Hygiene and Tropical Medicine). Section 2.6 is a subsequent development of this work.

2.2 Modelling framework

2.2.1 Epidemic models with time-varying coefficients

We adopt a SEIR model as a guide in this Chapter, although the methodology can be applied to other dynamical systems. The model is set in (2.1), in which s_t accounts for the proportion of susceptible individuals at time t. Similarly, e_t, i_t and r_t respectively correspond to exposed (infected but not infective), infective, and removed (or resistant) individuals. New infections occur at a rate $\beta s_t i_t$, implying that the susceptible individuals make effective contacts at rate β (the effective contact rate), and only a fraction i_t of these contacts are made with infective individuals. The average period spent in the exposed (E) and infectious (I) compartments is respectively given by k^{-1} and γ^{-1}.

$$\begin{cases} \dfrac{ds_t}{dt} = -\beta s_t i_t \\ \dfrac{de_t}{dt} = \beta s_t i_t - k e_t \\ \dfrac{di_t}{dt} = k r_t - \gamma i_t \\ \dfrac{dr_t}{dt} = \gamma i_t \end{cases} \tag{2.1}$$

The basic reproduction number, R_0, represents the number of secondary infections from a primary infected individual in a fully susceptible population. A related quantity is

the effective reproduction number, R_t, that refers to the number of secondary cases from an infected individual at time t. R_t is a context-dependent quantity of high interest to policy makers as it indicates the possibility for the epidemic to grow ($R_t > 1$) or to decrease ($R_t < 1$) (Anderson et al., 1992).

Epidemic models can be quite detailed (including individual characteristics, geographic information etc.), or basic such as the SEIR model that geographically aggregates the cases and assumes deterministic transmission processes occurring at a given frequency each time infected and susceptible individuals meet. The simplest models allow easier inference and interpretation, but they are based on strong assumptions that could lead to misleading conclusions. In this Chapter we adopt stochastic extensions of the deterministic SEIR model. The additional dynamic error is likely to contain structural misspecifications and can subsequently be explored and potentially revised. We focus on large-scale epidemics, for which random effects in transmission processes can be considered to be well-approximated deterministically (Kurtz, 1981). The paradigm we adopt attributes the model limitations mainly to the time-varying nature of the effective contact rate, henceforth denoted as β_t, rather than to the variability in individual characteristics or in transmission processes.

An early approach to estimate R_t can be found in Fine and Clarkson (1982). It can be implemented through discrete generation models or by reconstructing the chain of transmission (Cauchemez et al., 2006; Griffin et al., 2011). However, as R_t estimates contain both the effects of evolving transmissibility and immunity, quantitative conclusions can hardly be generalised to cases where the immunological situation is different. We therefore concentrate on estimating β_t rather than R_t. A number of approaches use a finite-dimension function space for the trajectory of β_t. Low-dimensional examples can be found in Cauchemez et al. (2008), in which β_t is modeled as a piecewise linear function. In some higher-complexity models, as in Cauchemez and Ferguson (2008) and Ionides et al. (2006), β_t is estimated freely with a few-weeks resolutions. Loosely speaking, as the number of parameters for the trajectory of β_t increases, model-induced biases fade out at the expense of the variance. A compromise is required to improve robustness and is often controlled through a regularising parameter. For example, in He et al. (2011), β_t is estimated using cubic splines calibrated via AIC.

2.2.2 Diffusion driven epidemic models

We consider models where some coefficients in (2.2) are modelled with diffusion processes; in particular, focusing on the effective contact rate β_t:

$$\begin{cases} \dfrac{ds_t}{dt} = -\beta_t s_t i_t \\ \dfrac{de_t}{dt} = \beta_t s_t i_t - k e_t \\ \dfrac{di_t}{dt} = k e_t - \gamma i_t \\ \dfrac{dr_t}{dt} = \gamma i_t \\ dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + \sigma^{\theta_t}(x_t^{\theta_t}, \theta)dB_t, \quad x_t^{\theta_t} = g^{\theta_t}(\beta_t) \end{cases} \tag{2.2}$$

Here, $\mu^{\theta_t}(\cdot)$ denotes the drift, $\sigma^{\theta_t}(\cdot)$ the volatility and $g^{\theta_t}(\cdot)$ is a transformation meant to project the possible values of β_t on the real line. The assigned diffusion may capture features such as behaviour changes, preventive measures, seasonal effects, holidays etc. When prior knowledge on β_t is available, it can be reflected in $\mu^{\theta_t}(\cdot)$ and $\sigma^{\theta_t}(\cdot)$. For example, if the contact rate is expected to converge, an Ornstein Uhlenbeck process can be chosen. Other options may include a sigmoid or a sinusoidal form. In absence of prior information, or when the researcher wants to impose little restrictions, a Brownian motion can be used, with $\mu^{\theta_t}(\cdot) \equiv 0$ and $\sigma^{\theta_t}(\cdot) \equiv \sigma$. This model, with $g^{\theta_t}(\cdot) \equiv \log(\cdot)$, is henceforth denoted as BM. The obtained output can be either reported or used as an exploratory tool to construct a more structured model; see Section 2.5.3 and Chapter 3 for applications. The choice of BM implies a continuous, yet non-differentiable path satisfying the Markov property. In cases where β_t is believed to evolve as a smooth function of time, higher order Brownian motions could be used. The latter may be regarded as equivalent to non-parametric approaches such as cubic splines (Wahba, 1990) with the model in (2.2) imposing a prior on β_t, and σ being a regularising factor. The rate β_t can be perceived as a product of a smooth and a rough component; the former being a population average of the intrinsic transmission procedure and the latter containing extrinsic factors such as the amount of contact among individuals. It is therefore important to build a framework that contains both smooth and rough models.

The above model can be estimated with an Extended Kalman Filter (EKF), as in Cazelles and Chau (1997). The EKF, described in the Introduction of this thesis, allows for fast computations but is based on Taylor and Gaussian approximations whose error could be non-negligible; see Supplementary Materials for a relevant simulation experiment (Appendix

A). Nevertheless, the EKF can still be used as a tool to construct efficient proposal distributions for MCMC schemes. It can also be used to optimize sequential Monte Carlo (SMC) algorithms, but either at a strong computational cost (Särkkä and Sottinen, 2008) or crude time discretisations (Merwe et al., 2001; Dukic et al., 2009). Next, we develop a general framework for efficient MCMC schemes that allows fine time discretisations.

2.3 Data augmentation via MCMC for diffusion driven epidemic models

This section presents a general inferential framework for diffusion-driven epidemic models. We adopt the Bayesian paradigm to incorporate parameter uncertainty and prior information in the estimates of β_t trajectories. The problem can also be cast as estimating partially observed hypoelliptic diffusions, thus presenting various difficulties (Pokern et al., 2009). We begin by setting the model and justifying the need for data augmentation. Existing MCMC algorithms are considered but they can lead to extremely inefficient MCMC chains. We address the issue by taking advantage of the specific model structure to construct adaptive PMCMC schemes.

2.3.1 Model and data augmentation setup

For ease of exposition we focus on models satisfying (2.2), but the framework covers models with different ODE systems or more time-varying coefficients, as in Section 2.5.3. Being in continuous time, t can take any value between t_0 and t_n. We denote $z_{i:j}$ the path of the ODE states vector $z_t = \{s_t, e_t, i_t, r_t\}$ between observation times t_i and t_j. The data, $y_{1:n} = \{y_1, .., y_n\}$, usually provide information for i_t at specific times (prevalence data) or for integrals of z_t (incidence data). In either case, we assume that they are obtained with error as the collection procedure is typically associated with additional uncertainty. The observation distribution is denoted with $\mathbb{P}_y$ with density $h(y_{1:n}|z_{0:n}, \theta)$. Note that, in the model of (2.2), z_t can be written as a deterministic function $f(\cdot)$ of $x_t^{\theta_t}$ and the parameters $(k, \gamma, z_0) \subset \theta$. This function is the solution of the ODE and can be written as an intractable time integral involving $x_t^{\theta_t}$. Hence, the model becomes:

$$\begin{cases} dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + \sigma^{\theta_t}(x_t^{\theta_t}, \theta)dB_t \\ z_{0:n} = f(x_{0:n}^{\theta_t}, \theta) \\ y_{1:n}|z_{0:n}, \theta \sim \mathbb{P}_y(y_{1:n}|z_{0:n}, \theta) \end{cases} \tag{2.3}$$

We denote with $\mathbb{P}_{x^{\theta_t}}$ the distribution of the diffusion $x_t^{\theta_t}$ defined from the SDE above. We require the existence of a unique weak solution which translates into some mild assumptions on $\mu^{\theta_t}(.)$ and $\sigma^{\theta_t}(.)$; e.g. locally Lipschitz with a linear growth bound (Øksendal, 2003). The distribution of $\mathbb{P}_{x^{\theta_t}}$ may also be viewed as a prior on $x_t^{\theta_t}$, or else β_t. The model can now be defined from $\mathbb{P}_y$, $\mathbb{P}_{x^{\theta_t}}$, and the assigned priors on θ, denoted by $\pi(\theta)$:

$$\pi(x_{0:n}^{\theta_t}, \theta | y_{1:n}) \propto f(y_{1:n}|z_{0:n}, \theta) \times d\mathbb{P}_{x^{\theta_t}} \times \pi(\theta) \tag{2.4}$$

Given direct observations on $x_t^{\theta_t}$, it would have been possible to draw approximation-free inference on $d\mathbb{P}_{x^{\theta_t}}$ using the approach of Beskos et al. (2006). However, this is not possible in our case given the non-linear functionals in $f(\cdot)$ that render 2.4 intractable. We proceed by discretizing the path of x_t, and therefore of β_t and z_t. More specifically, we introduce m points between each pair of successive observation times t_i and t_{i+1} ($i = 0, 1, \ldots, n-1$). When referring to the discrete representation of a path, the superscript dis will be used; for example for a step $\delta = \frac{1}{m+1}$, the discrete skeleton of $x_t^{\theta_t}$ will be denoted by $x_{0:n}^{dis} = \{x_0, x_\delta, x_{2\delta}, \ldots, x_{t_n}\}$. The use of $x_{0:n}^{dis}$ allows for approximations of (2.4) through the Euler-Maruyama scheme to evaluate $d\mathbb{P}_{x^{\theta_t}}$:

$$\begin{cases} p\left(x_{\delta:n}^{dis}|x_0, \theta_x\right) = \prod_{i:\ t_0 < i\delta \leq t_n} p(x_{i\delta}^{dis}|x_{(i-1)\delta}^{dis}, \theta), \\ x_{i\delta}^{dis}|x_{(i-1)\delta}^{dis} \sim \mathcal{N}\left\{x_{(i-1)\delta}^{dis} + \delta\mu_{x^{\theta_t}}(x_{(i-1)\delta}^{dis}, \theta,\ \delta\sigma_{x^{\theta_t}}(x_{(i-1)\delta}^{dis}, \theta)^2\right\} \end{cases} \tag{2.5}$$

Moreover, given $x_{0:n}^{dis}$, the ODE can be solved numerically to obtain $z_{0:n}^{dis}$ and evaluate $p(y_{1:n}|z_{0:n}^{dis}, \theta)$. The approximation error can be made arbitrarily small by decreasing the user-specified parameter δ.

2.3.2 Data augmentation via Gibbs schemes

Model (2.3) can be put in the context of Chib et al. (2006), Kalogeropoulos (2007) or Golightly and Wilkinson (2008). In these approaches, a Gibbs scheme can be used to sample from the joint posterior in (2.4) of $x_{0:n}^{dis}$ and θ. The data augmentation algorithm alternates between drawing $x_{0:n}^{dis}$ given θ, and updating θ conditional on the augmented path $x_{0:n}^{dis}$. The MCMC protocol ensures that the chain provides samples from the marginal posteriors of $x_{0:n}^{dis}$ and θ. Nevertheless, the properties of the algorithm may become unacceptably poor. There are two essential issues associated with such schemes. The first concerns the non-trivial step of sampling on the diffusion pathspace of $x_t^{\theta_t}$. The second problem is caused by the high posterior correlations between $x_{0:n}^{dis}$ and θ, leading to reducible chains as δ decreases (Roberts and Stramer, 2001).

The majority of the literature on data augmentation schemes for diffusions handles the conditional updates of $x_{0:n}^{dis}$ with an independence sampler. As it is difficult to find good proposal distributions for the entire $x_{0:n}^{dis}$, the path is usually split into blocks. Overlapping blocking strategies are essential to ensure that all points are updated and continuity of the path is retained. An alternative way to update $x_{0:n}^{dis}$ is to use the particle filter via the Particle Gibbs algorithm of Andrieu et al. (2010). However, unless the issue of high posterior correlation between $x_{0:n}^{dis}$ and θ is resolved, none of these schemes will improve the overall MCMC performance. The problem is caused, for example, by the quadratic variation process of $x_t^{\theta_t}$ that identifies components of θ. For $\sigma^{\theta_t}(x_t^{\theta_t}, \theta) \equiv \sigma$ we get

$$\lim_{\delta \to 0} \sum_{i:\ t_0 < i\delta \leq t_n} (x_{i\delta}^{dis} - x_{(i-1)\delta}^{dis})^2 = \int_{t_0}^{t_n} \sigma^2 ds = \sigma^2 (t_n - t_0) \tag{2.6}$$

Thus, the conditional posterior of σ converges to a point mass as δ tends to 0. In practice this translates into an increasingly slow MCMC algorithm with a convergence rate of $O(\delta^{-1})$ (Roberts and Stramer, 2001). Schemes with a fixed δ (Cori et al., 2009) could work in some occasions but the approximation error could be substantial. In some cases, the problem can be tackled with suitable reparametrisation. The approach of Roberts and Stramer (2001) involves transforming $x_t^{\theta_t}$ to a diffusion $\dot{x}_t^{\theta_t}$ with unit volatility. An alternative scheme is offered by Chib et al. (2006) where the driving Brownian motion of $x_t^{\theta_t}$ is being used. In these algorithms the ODE states vector $z_{0:n}^{dis}$ becomes a function of σ, $\dot{x}_t^{\theta_t}$ and θ. Hence, in a Metropolis step, every proposed value of σ^* is associated with the corresponding values of $z_{0:n}^{dis^*}$. This succeeds into breaking the perfect dependence between $z_{0:n}^{dis}$ and σ, even for $\delta \to 0$. However, since components of $z_{0:n}^{dis}$ (or functionals thereof) are observed with error, the associated proposed values $z_{0:n}^{dis^*}$ should be close to the data for the move to be accepted. As the observation error becomes small and the data increase, this becomes increasingly difficult and leads to very small moves for σ and poor MCMC mixing. More details and simulations supporting this argument are provided in the Supplementary Materials (Appendix B). Consequently, we overcome this issue by updating $x_{0:n}^{dis}$ and θ jointly via the PMCMC algorithm, which is essential as it is not straightforward to implement joint updates with the other approaches mentioned in this section.

2.3.3 Adaptive particle Markov Chain Monte Carlo algorithms

Particle filters are SMC algorithms used to recursively explore conditional densities in state space models (Doucet and Johansen, 2009). For given values of θ, J particles $(x_t^{(j)})$ are sequentially propagated from t_0 to t_n. In various time steps t_i, the trajectories that best fit the data $y_{1:i}$ are given more weight through resampling. Algorithm 9 shows how they can be applied in our context. The quantity $L^{i+1}(\theta)$ provides estimates of $p(y_{1:i}|\theta)$ and the

Algorithm 9 Particle Smoothing algorithm

Initialise: Set $L^0(\theta) = 1$, $W_0^{(j)} = \frac{1}{J}$, sample $(x_0^{(j)})_{j=1,\dots,J}$ from $p(x_0|\theta)$ and calculate $(z_0^{(j)})_{j=1,\dots,N}$ by solving the ODE (for example with the Euler scheme)
for $i = 0$ to $n-1$ **do**
  for $j = 1$ to J **do**
    Sample $(x_{i:i+1}^{(j)})$ from $p(x_{k:k+1}^{dis}|x_k, \theta)$ (see Eq. 2.5) and calculate $(z_{i:i+1}^{(j)})$ by solving the ODE
    Set $\alpha^{(j)} = h(y_{i+1}|z_{0:i+1}^{(j)}, \theta)$
  end for
  Set $W_{i+1}^{(j)} = \frac{\alpha^{(j)}}{\sum_{k=1}^{J} \alpha^{(k)}}$, and $L^{i+1}(\theta) = L^i(\theta) \times \frac{1}{J} \sum \alpha^{(j)}$
  Resample $(z_{0:i+1}^{(j)}, x_{0:i+1}^{(j)})_{j=1,\dots,J}$ according to $(W_{i+1}^{(j)})$,
end for

resampling step is essential to control the variance of that estimate over time. Algorithm 9 also provides a random sample from $p(x_{1:i}^{dis}|y_{1:n}, \theta)$. In order to sample from $p(x_{1:n}^{dis}, \theta|y_{1:n})$, the PMCMC algorithm can be used. PMCMC was introduced in Andrieu et al. (2010) and successfully integrates particle filters in MCMC algorithms. Its implementation is presented in Algorithm 10. The issues raised in Section 2.3.2 are now addressed as $x_{0:n}^{dis}$

Algorithm 10 Particle MCMC algorithm (particle Marginal Metropolis Hastings version)

Initialise: $\theta^{(0)}$, to an initial value.
Use Particle Smoother to compute $L(\theta^{(0)}) = \hat{p}(y_{1:n}|\theta^{(0)})$, and sample $x_{1:n}^{(0)}$ from $p(x_{1:n}^{dis}|y_{1:n}, \theta^{(0)})$
for $k = 1$ to N^θ **do**
  Sample θ^* from $q(.|\theta^{(i)})$
  Use PS to compute $L(\theta^*)$ and sample $x_{1:n}^*$ from $\hat{p}(x_{1:n}^{dis}|y_{1:n}, \theta^*)$
  Accept θ^* (and $x_{1:n}^*$) 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

and θ are sampled jointly. In other words, $x_{0:n}^{dis}$ is being numerically integrated out while a sample from its posterior is obtained at each MCMC iteration.

Although the PMCMC algorithm is theoretically valid even for a single particle (Andrieu et al., 2010), large values of J are usually required for reasonably stable acceptance rates and large moves in the θ space. It is therefore essential to update the d-dimensional θ at once, making the proposal $q(.|\theta)$ crucial to the overall MCMC performance. In this chapter we propose to use the adaptive Metropolis algorithm of Roberts and Rosenthal (2009). After transforming the parameters to take values in the real line we use a Normal distribution centered at the current value of θ and with covariance given by $\lambda\Sigma$. Static random walk metropolis proposals set $\Sigma = I_d$ or $\Sigma = \hat{\Sigma}$. Adaptive schemes change the value

λ at each iteration i to obtain acceptance rate of 0.234 through diminishing adaptation:

$$\lambda^{(i+1)} = \exp\left\{\log(\lambda^{(i)}) + a^n(AccRate - 0.234)\right\}, \tag{2.7}$$

where the cooling rate $a < 1$ can be set to 0.999 and $AccRate$ denotes the acceptance rate up to iteration i. The general shape of the importance sampling distribution $q(.|\theta^{(i)})$ can also be progressively adapted:

$$q(.|\theta^{(i)}) = \alpha\mathcal{N}\left(\theta^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(0)}\right) + (1-\alpha)\mathcal{N}\left(\theta^{(i)}, \lambda\frac{2.38^2}{d}\Sigma^{(i)}\right) \tag{2.8}$$

In this alternative adaptive scheme, α is usually set to 0.05, $\Sigma^{(i)}$ is the posterior covariance matrix estimated by the draws up to i, and $\Sigma^{(0)}$ should be specified in advance. In this chapter we enhance the above adaptive algorithms utilising information from the EKF to estimate the covariance $\hat{\Sigma}$ or $\Sigma^{(0)}$. One choice, EK-Mode, is the observed information matrix evaluated through numerical differentiation at the mode of the approximate posterior density proportional to $p^{EKF}(y_{1:n}|\theta)p(\theta)$. Another choice, EK-MCMC, is to run an MCMC scheme based on the EKF approximation of the likelihood $p^{EKF}(y_{1:n}|\theta)$ and compute the posterior covariance from the draws. Note that the computational burden of these methods is marginal with regards to the PMCMC. As demonstrated in Section 2.4, the use of these initialisations can result in substantial improvement.

2.4 Simulation experiments

The proposed algorithms are illustrated and tested on simulated data in this section. We focus on the BM model, where $\log(\beta_t)$ follows a Brownian motion with volatility σ, corresponding to the case of little information on the shape of β_t. The trajectories of β_t were drawn either from the BM model itself (experiment 1) or from a deterministic sigmoid curve (experiment 2). The data y_i, $i = 1, \ldots, 50$ represent noisy observations of weekly new cases of the epidemic $\left(\int_{week\ i} ke_t dt\right)$. We complete the model by assigning a Normal distribution to each $\log(y_i)$ with mean $\log(\int_{week\ i} ke_t dt)$ and variance τ^2. The parameters were tuned to obtain realistic epidemic incidence curves, and observations were generated setting $\tau = 0.1$ (experiments 1.a and 2.a), and $\tau = 0.05$ (experiments 1.b and 2.b). The assigned priors were informative for k, γ and r_{t_0} and vague for e_{t_0}, i_{t_0}, σ and τ. We used 3,000 particles and 100,000 MCMC iterations after a long burn-in period, in order to estimate σ, τ and $\beta_{1:n}$. The number of particles was determined by preliminary exploration of the impact of J on the acceptance rate, illustrated by figure 2.1 for two values of the measurement error parameter τ. Similarly, the discretisation step δ is determined by monitoring quantities such as $E[\hat{p}_\delta(\sigma|y_{1:n})]$ or $E[\hat{p}_\delta(\tau|y_{1:n})]$ for different values of δ. As shown

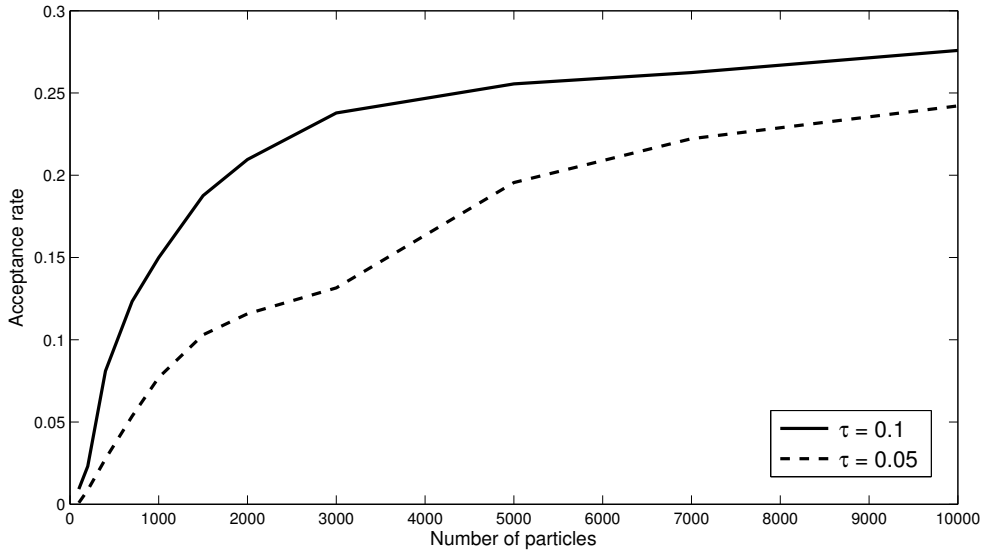

Figure 2.1: Acceptance rate as a function of J, in two situations where the noise amplitude is respectively 10% (full line) and 5% (dotted line).

		Exp 1.a	**Exp 1.b**	**Exp 2.a**	**Exp 2.b**
τ	Simulation value	**0.1**	**0.05**	**0.1**	**0.05**
	Posterior mean	0.103	0.083	0.078	0.050
	Posterior median	0.103	0.084	0.077	0.050
	Posterior 95% c.i.	$[0.051; 0.152]$	$[0.027; 0.137]$	$[0.063; 0.96]$	$[0.042; 0.060]$
σ	Simulation value	**0.07**	**0.07**	**n.d.**	**n.d.**
	Posterior mean	0.066	0.083	0.016	0.014
	Posterior median	0.064	0.084	0.015	0.014
	Posterior 95% c.i.	$[0.048; 0.090]$	$[0.046; 0.089]$	$[0.010; 0.027]$	$[0.001; 0.021]$

Table 2.1: Mean, median and 95% confidence intervals for τ and σ estimates in four experiments.

in figure 2.3, convergence is observed as δ tends to 0. In the present series of experiments, δ was fixed to 0.1.

Figure 2.3 shows estimates and 95% pointwise credible intervals of the path for experiments 1.a and 1.b provided by the adaptive PMCMC initialized with EK-MCMC. The posterior output is in good agreement with the simulation trajectories suggesting that the underlying trajectory of β_t can be estimated reasonably well from the partial and noisy observations considered (experiments 2.a and 2.b with $\tau = 0.05$ lead to similar results).

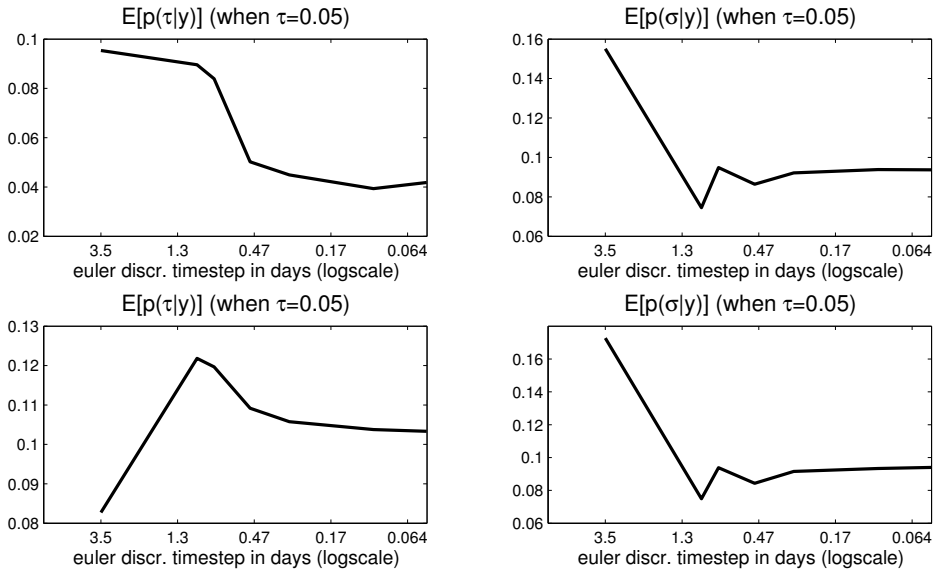

Figure 2.2: Convergence of the posterior density as the Euler discretization time-step δ decreases (x-axis in the log-scale)

Moreover, Table 2.1 presents the mean, median and 95% credible intervals for the estimates of σ and τ in each experiment. True values for σ and τ are contained in the 95% credible intervals in all cases but τ in experiment 2.a. A potential explanation for this slight underestimation may be the fact that we estimate a smooth trajectory with a Brownian motion which artificially captures part of the noise.

Next, we use the data of experiment 1 to compare the proposed adaptive PMCMC schemes. Comparison is made in terms of the effective sample size $\left(1 + 2\sum_{i\geq 1}\eta(i)\right)^{-1}$, with $\sum_i \eta(i)$ being the sum of the lagged sample auto-correlations, as in Geyer (1992) or Brooks and Roberts (1998). We record the minimum ESS among the MCMC components and multiply by 100 to monitor the percentage of the total iterations that can be considered as independent. We consider three covariance matrices for each of the two adaptive algorithms defined in Section 2.3.3: I_d and the ones from EK-Mode and EK-MCMC. Simulation results are presented in table 2.2. Combining the initialisation based on EK-MCMC or EK-Mode to the adaptation of Σ^q leads to sampling performance close or equivalent to what is achieved when using the optimal but a priori unknown matrix $Cov[p(\theta|y_{1:n})]$. Additionally, initialisations with EK-Mode and EK-MCMC directly allow for effective exploration of the target density, and thus a quicker burn-in adaptation period than with Id for which, even when adapting ε, more than 100,000 iterations are required to generate 10

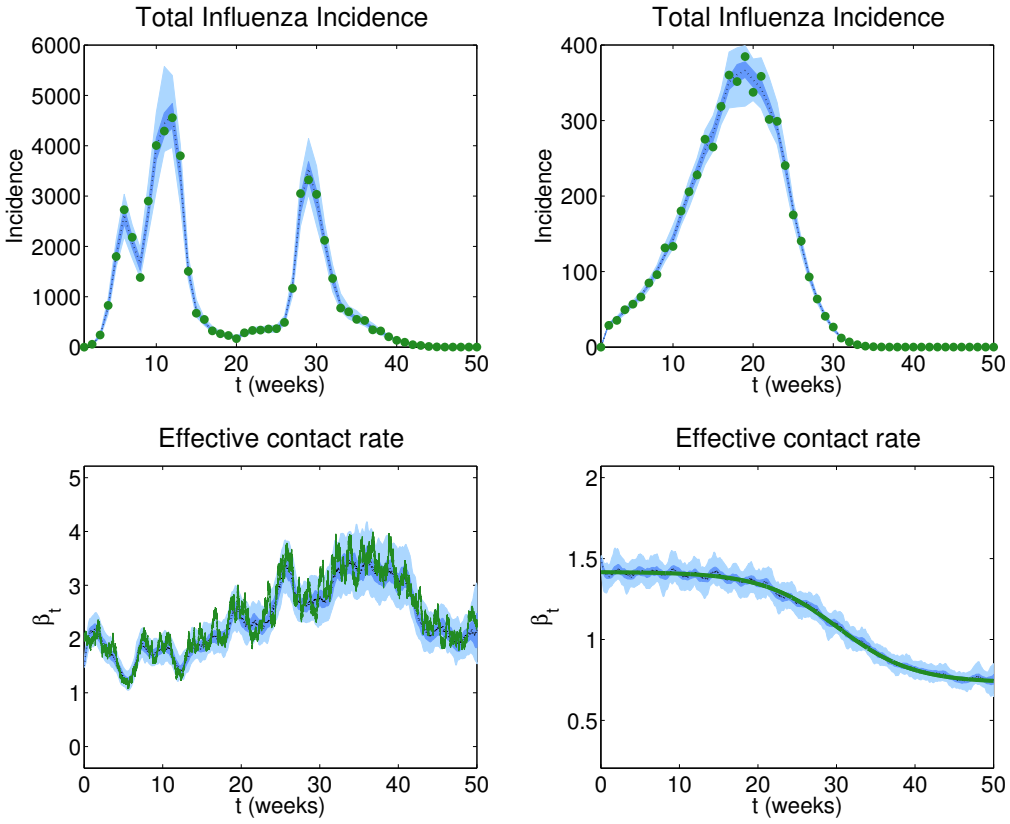

Figure 2.3: Illustration of how the underlying dynamic of the effective contact rate can be estimated from weekly recorded cases. Green dots indicate simulated observed incidence (top panels). Green lines indicate simulated effective contact rate trajectories (bottom panels). Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas show credible intervals, respectively at 50% and 95% levels.Top panels: simulated weekly numbers of cases observed with noise, and corresponding model-based offline reconstructions (left: experiment 1, right: experiment 2) *Bottom panels*: simulated and estimated trajectory of the effective contact rate (left: experiment 1, right: experiment 2)

effectively independent samples. The substantial improvement induced by the proposed adaptive algorithms is expected to intensify as the dimension of θ increases.

Σ_0^q	Algorithm parameter being adapted	$min_\theta(ESS)$	$min_\theta(ESS)$ relative to optimal initialisation $\Sigma_0^q = Cov[p(\theta\|y_{1:n})]$
Id	ε	0.008%	0.01
EK-Mode	ε	0.19%	0.14
EK-MCMC	ε	0.54%	0.39
Id	Σ^q	0.57%	0.41
EK-Mode	Σ^q	1.24%	0.89
EK-MCMC	Σ^q	1.38%	1.00

Table 2.2: Relative efficiency of the different versions and initialisations of the adaptive PMCMC algorithm, via the minimum ESS (%) after adaptation.

2.5 The 2009 A/H1N1 pandemic

2.5.1 Data, model and estimates

The proposed methodology is illustrated on data from the A/H1N1(2009) pandemic in England between June and December 2009. The data consists of estimates of weekly ILI cases $y_{1:n}$ provided by the Health Protection Agency (Baguelin et al., 2010). The estimates were obtained from the recorded ILI cases among a selected sample of GPs. They accounted for over-reporting due to similarities in symptoms with other respiratory diseases, based on subsequent virological positivity tests. Corrections for asymptomatic infections and the propensity of each patient to consult were also made. Overall these corrections lead to a multiplicative correction coefficient $c = 10$, whose value is also supported by a further serological survey (Miller et al., 2010). In our analysis c is initially held fixed to 10, but this choice is explored further in Section 2.5.2. We adopt a model that admits noisy data to reflect the associated uncertainty. The noise model of Section 2.4 was used, combined with a BM formulation of $\mathbb{P}_x$. Vague priors, $\mathcal{N}_{>0}(0, 10^6)$, were put on τ, σ and β_0. The priors for k and γ were obtained from additional data sources (Baguelin et al., 2010), the results of which are summarised through Normal distributions that place 95% probability in a symmetric manner between 1.55 and 1.63 days for the latent period k^{-1}, and between 0.93 and 1.23 days for the infectious period γ^{-1}. A Dirichlet distribution was used for the initial proportions in compartments S, E, I, R, constraining the mean of the one in R to be 0.15, its variance 0.15^2, and the means of the other initial proportions to be equal.

The adaptive EK-MCMC algorithm was applied to the data and Fig. 2.4 depicts the incidence curve together with the posterior mean and pointwise 95% credible intervals (corresponding traceplots for the components of θ are shown in figure 2.5). Estimates of β_t are also displayed indicating various changes over time. The changes in β_t are consistent with the argument that schools closure for holidays have been driving the epidemic: differ-

ent values are observed during school and holidays periods, appearing to be synchronised with schools opening and closing. Posterior summaries for the static parameters, as well as a sensitivity analysis on the priors can be found in table 2.3. They suggest that estimates are quite sensitive to the choice of prior for k and γ, but not to the remaining parameters.

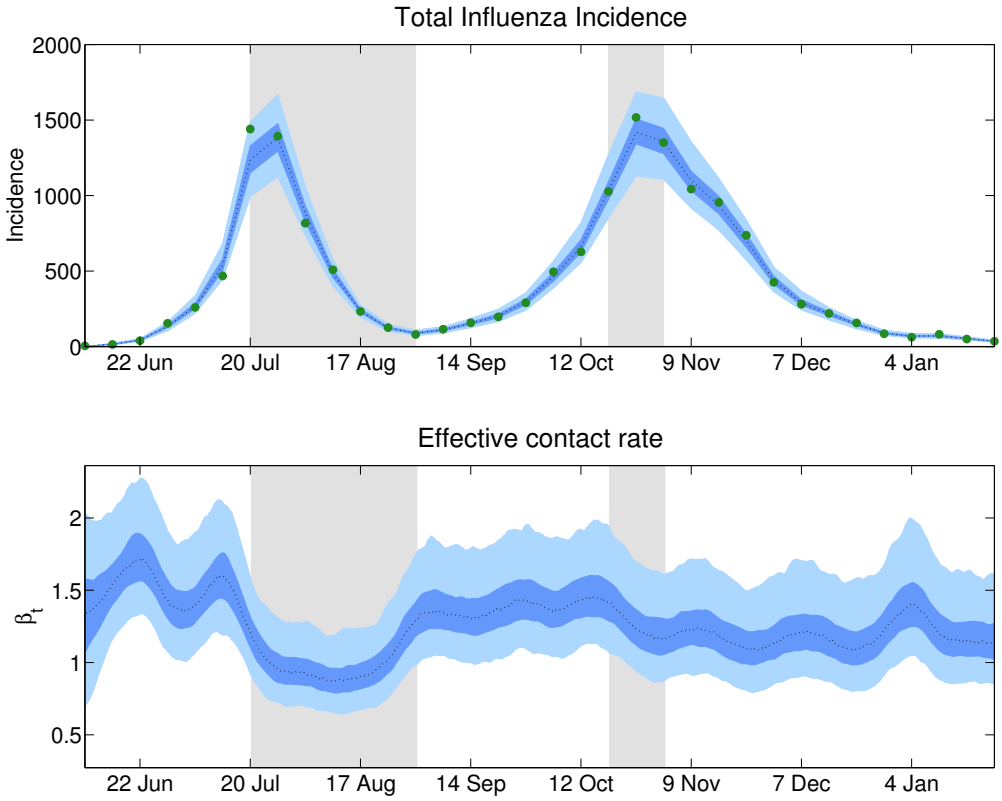

Figure 2.4: Weekly incidence data from the A/H1N1 2009 influenza pandemic and corresponding offline estimates of the effective contact rate. Green dots indicate incidence estimates provided by the Health Protection Agency. Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas show credible intervals, respectively at 50% and 95% levels. Holidays are indicated by a light grey area. *Top* : observations of the weekly total number of A/H1N1 influenza cases in London (per 100,000 inhabs.) and model-based offline reconstruction. *Bottom*: offline estimates of the effective contact rate.

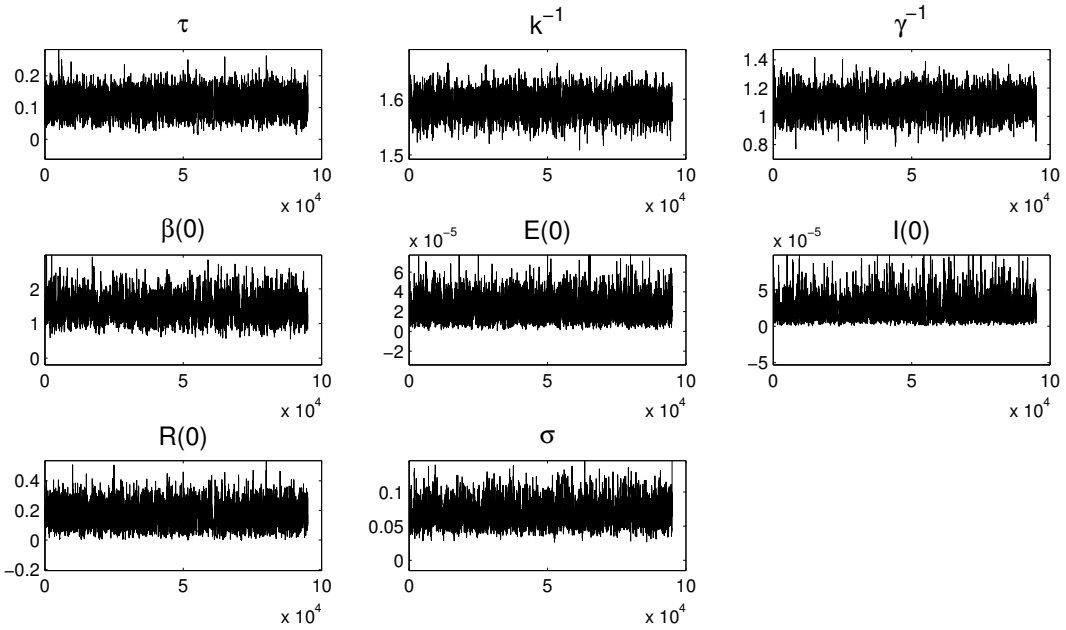

Figure 2.5: MCMC traceplots for each component of θ

2.5.2 Application in real time. Was the first wave waning due to depletion of susceptibles?

In this section the proposed methodology is applied in real time, i.e. considering partial datasets from June 2009 up to the 20th of July, the 7th of September and the 26th of October. Each time the algorithm is run from scratch to provide samples from the joint posterior $\pi(x_{1:i}, \theta | y_{1:i})$. From a computational cost point of view this procedure can be improved further by utilising previous MCMC runs, for example under the SMC^2 framework (Chopin et al., 2012). We did not pursue this direction further, as the PMCMC algorithm runs quite fast (less than 2 hours on a standard PC). In order to reduce uncertainty, especially at early stages, the value of τ was set to 0.1 rather than being estimated as in Section 2.5.1. We otherwise use the same model as before: main results are shown in Fig. 2.6. A model with integrated Brownian motion was also fit but BM was chosen in terms of Deviance Information Criterion; see figure 2.7. In addition, figure 2.7 also illustrates what is obtained from a Frequentist approach relying on an optimisation of the marginal likelihood, which can be done with the maximum likelihood via iterated filtering algorithm (MIF) introduced in (Ionides et al., 2006). As the density $p(\beta_{0:n} | y_{1:n})$ does not account for parameter uncer-

	τ	k^{-1}	γ^{-1}	β_0	$E(0)$	$I(0)$	$R(0)$	σ
2.5% quantile	0.04	1.55	0.93	0.80	5.2×10^{-6}	1.6×10^{-6}	0.02	0.04
25% quantile	0.09	1.57	1.03	1.16	1.6×10^{-5}	7.0×10^{-6}	0.12	0.05
Median estimate	0.11	1.59	1.08	1.35	2.3×10^{-5}	1.6×10^{-5}	0.17	0.06
75% quantile	0.13	1.60	1.13	1.56	3.1×10^{-5}	2.8×10^{-5}	0.22	0.07
97.5% quantile	0.17	1.63	1.23	2.13	5.2×10^{-5}	6.5×10^{-5}	0.33	0.10
Median when r_{t_0} shifted +10%	0.11	1.59	1.09	1.41	2.0×10^{-5}	1.9×10^{-5}	0.19	0.06
Median when r_{t_0} shifted +20%	0.11	1.59	1.09	1.44	1.8×10^{-5}	2.1×10^{-5}	0.24	0.06
Median when r_{t_0} shifted -10%	0.11	1.59	1.08	1.31	2.2×10^{-5}	2.2×10^{-5}	0.15	0.07
Median when r_{t_0} shifted -20%	0.12	1.59	1.09	1.27	1.9×10^{-5}	2.1×10^{-5}	0.13	0.07
Median when k^{-1} shifted +10%	0.11	1.59	1.09	1.33	1.8×10^{-5}	2.2×10^{-5}	0.15	0.06
Median when k^{-1} shifted +20%	0.12	1.60	1.08	1.34	1.8×10^{-5}	2.3×10^{-5}	0.17	0.06
Median when k^{-1} shifted -10%	0.12	1.58	1.08	1.28	2.0×10^{-5}	2.0×10^{-5}	0.16	0.06
Median when k^{-1} shifted -20%	0.12	1.57	1.09	1.31	1.9×10^{-5}	2.0×10^{-5}	0.16	0.06
Median when γ^{-1} shifted +10%	0.11	1.59	1.12	1.23	2.2×10^{-5}	2.1×10^{-5}	0.15	0.07
Median when γ^{-1} shifted +20%	0.10	1.59	1.14	1.18	2.0×10^{-5}	2.4×10^{-5}	0.15	0.07
Median when γ^{-1} shifted -10%	0.11	1.59	1.06	1.37	1.9×10^{-5}	2.0×10^{-5}	0.16	0.07
Median when γ^{-1} shifted -20%	0.10	1.59	1.02	1.46	2.0×10^{-5}	1.8×10^{-5}	0.17	0.07

Table 2.3: Original estimates compared to the ones resulting from respectively tilting the priors on r_{t_0}, γ^{-1} or k^{-1} by +10, +20, -10 or -20%

tainty, the resulting pointwise 95% credible intervals are narrower; roughly 50% on the 6-month dataset and even more at early stages with less information on θ.

On August 1^{st}, the first wave of the epidemic had waned, incidence rates were decreasing and schools had closed. Two scenarios were competing to explain the epidemic decline: (i) holidays had caused the waning of the epidemic by lowering the effective contact rate. Hence, a similar or stronger wave could occur when schools would reopen in September in colder climatic conditions. (ii) The epidemic had stopped independently of holidays because a critical proportion of the population had been infected, conferring a sufficient level of herd immunity to stop the epidemic. In this case, no second wave was to be expected in September. On August 1^{st} there was great uncertainty around the value of c (Baguelin et al., 2010), which is crucial in distinguishing between the two scenarios. We therefore conducted the following exercise.

The PMCMC algorithm, run up to August 1st, provides samples from the posterior of the difference in β_t between August 1^{st} and July 13^{th} (before the decrease in incidence). For $c = 10$, the 97.5% point of this posterior is -0.32, indicating a decrease in β_t. The latter supports scenario (i), as the competing scenario is associated with a zero-decrease in β_t. Nevertheless, as this value depends on c, the algorithm was run for different values of it

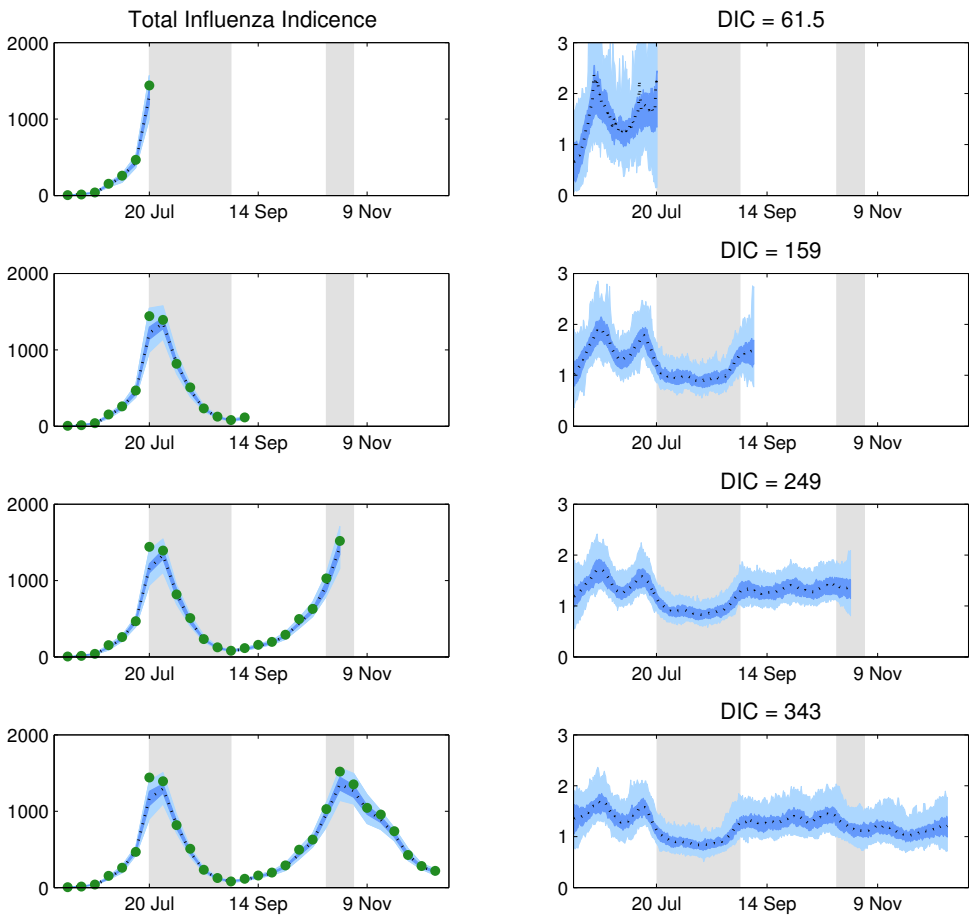

Figure 2.6: What could have been inferred by carefully following the epidemic in real time? Green dots indicate observed incidence estimates provided by the Health Protection Agency (left panels). Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas respectively indicate 50% and 95% credible intervals of the posterior density. Holidays are indicated by a light grey area. Left panels: HPA estimates of the weekly total number of A/H1N1 influenza cases in London (per 100,000 inhabs.) Right panels: "real-time" estimates of the effective contact rate.

ranging from 20 to 150. The results appear on Fig. 2.8. Note that the 97.5% point of interest increases as a function of c and reaches 0 for a correction factor close to 70. As this level seemed unrealistic (Baguelin et al., 2010), the experiment provides evidence in favour of scenario (i) highlighting the danger of a second wave in September, that actually occurred. Such evidence can be important for decision-makers, especially when considering implementations of preventive measures as vaccines.

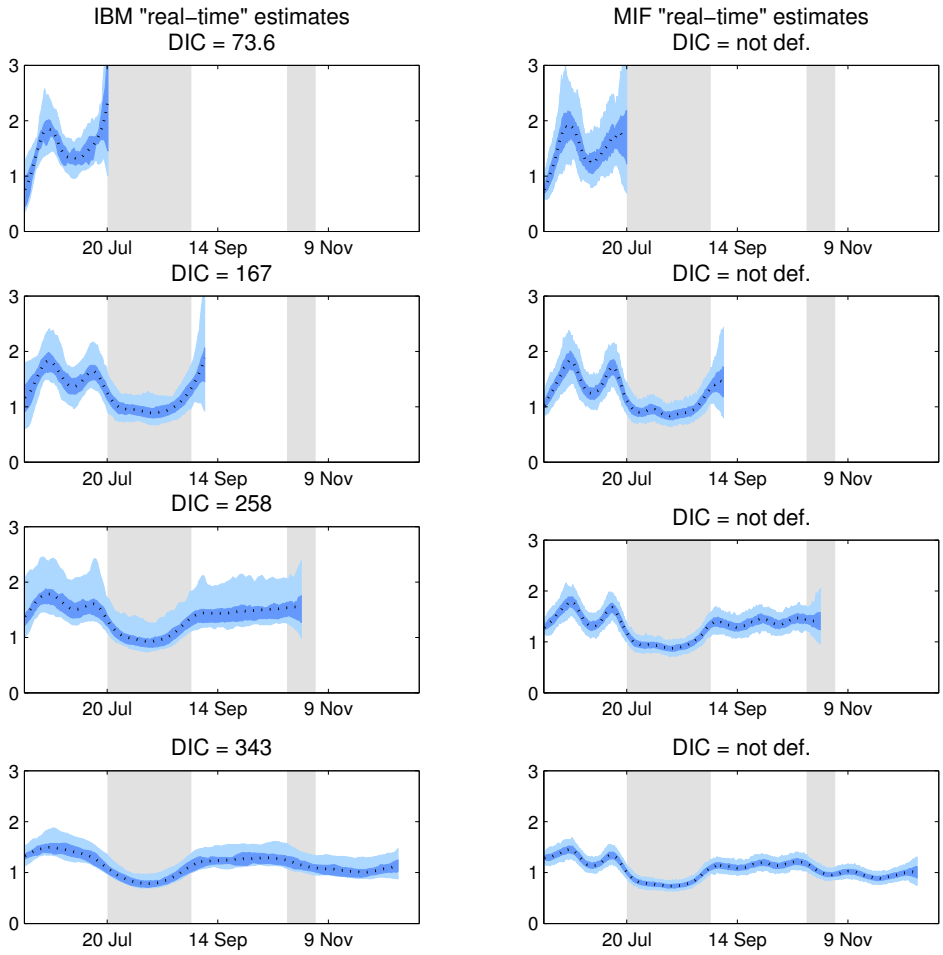

Figure 2.7: Modeling choices and implications, aiming for robustness. Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue show credible intervals, respectively at 50% and 95% levels. *Left panels*: estimates from an alternative modelling approach: exploring the full posterior density of an IBM diffusion model (left). *Right panels*: estimates from an alternative methodological approach: exploring the posterior density of a BM diffusion model conditioned on a likelihood-maximizing parameter θ^* provided by the MIF algorithm (right)

2.5.3 A multiple age group diffusion driven SEIR model

The analysis of Section 2.5.1 can be used to construct more structured models. For example, the effect of holidays is evident and may differ from children to adults, thus casting doubts on the assumption of a homogeneous population. It seems more natural to consider a

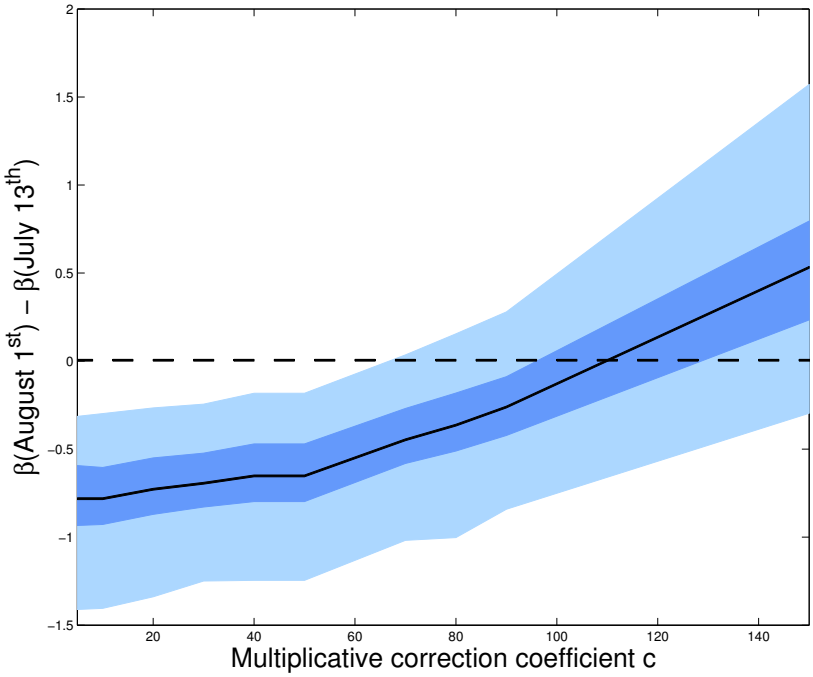

Figure 2.8: The implication of different scenarios for the real value of underreporting on the decrease of the effective contact rate between July 13^{th} and August 1^{st}. For each value of c, the mean of the posterior density for $\beta(August\ 1^{st}) - \beta(July\ 13^{th})$ is plotted in black. Dark and light blue areas respectively indicate 50% and 95% credible intervals of the posterior density. The dotted line locates the scenario whith no change in the effective contact rate.

model with two age groups (c: children and a: adults) and target all possible effective contact rates among them. In our notation β^{ca} refers to the effective contact rate from children to adults and s_t^c denotes the proportion of susceptible children at time t. For reasons of parsimony we assign Brownian motions to $\log(\beta_t^{cc})$, $\log(\beta_t^{aa})$ and treat β^{ca}, β^{ac}

as constant. The dynamic part of the model is now given by

$$\begin{cases} \frac{ds_t^c}{dt} = -s_t^c \left(\beta_t^{cc} i_t^c \frac{N}{N^c} + \beta^{ca} i_t^a \frac{N}{N^a} \right) \\ \frac{de_t^c}{dt} = s_t^c \left(\beta_t^{cc} i_t^c \frac{N}{N^c} + \beta^{ca} i_t^a \frac{N}{N^a} \right) - k e_t^c \\ \frac{ds_t^a}{dt} = -s_t^a \left(\beta_t^{aa} i_t^a \frac{N}{N^a} + \beta^{ac} i_t^c \frac{N}{N^c} \right) \\ \frac{de_t^a}{dt} = s_t^a \left(\beta_t^{aa} i_t^a \frac{N}{N^a} + \beta^{ac} i_t^c \frac{N}{N^c} \right) - k e_t^a \\ \frac{di_t^c}{dt} = k e_t^c - \gamma i_t^c \\ \frac{dr_t^c}{dt} = \gamma i_t^c \\ \frac{di_t^a}{dt} = k e_t^a - \gamma i_t^a \\ \frac{dr_t^a}{dt} = \gamma i_t^a. \end{cases} \tag{2.9}$$

The data from the A/H1N1(2009) pandemic provide incidence estimates for children and adults separately so they can be used to estimate the model of (2.9). If only final outcome data were available, not all effective contact rate parameters would be estimable. However, the temporal dataset provides extra information by the relative variation of susceptible and infective population in adults versus children. We applied the EK-MCMC scheme, which was essential in order to obtain reasonable MCMC performance. Fig. 2.9 depicts the results. Unlike earlier attempts with versions of a multi-group model with a single diffusion driving all contact rates, the fit appears to be good. The trajectory of children seems to be similar with that of Fig. 2.4 thus stressing their role to the evolution of the epidemic. Posterior summaries for the parameters can be found in table 2.4 .

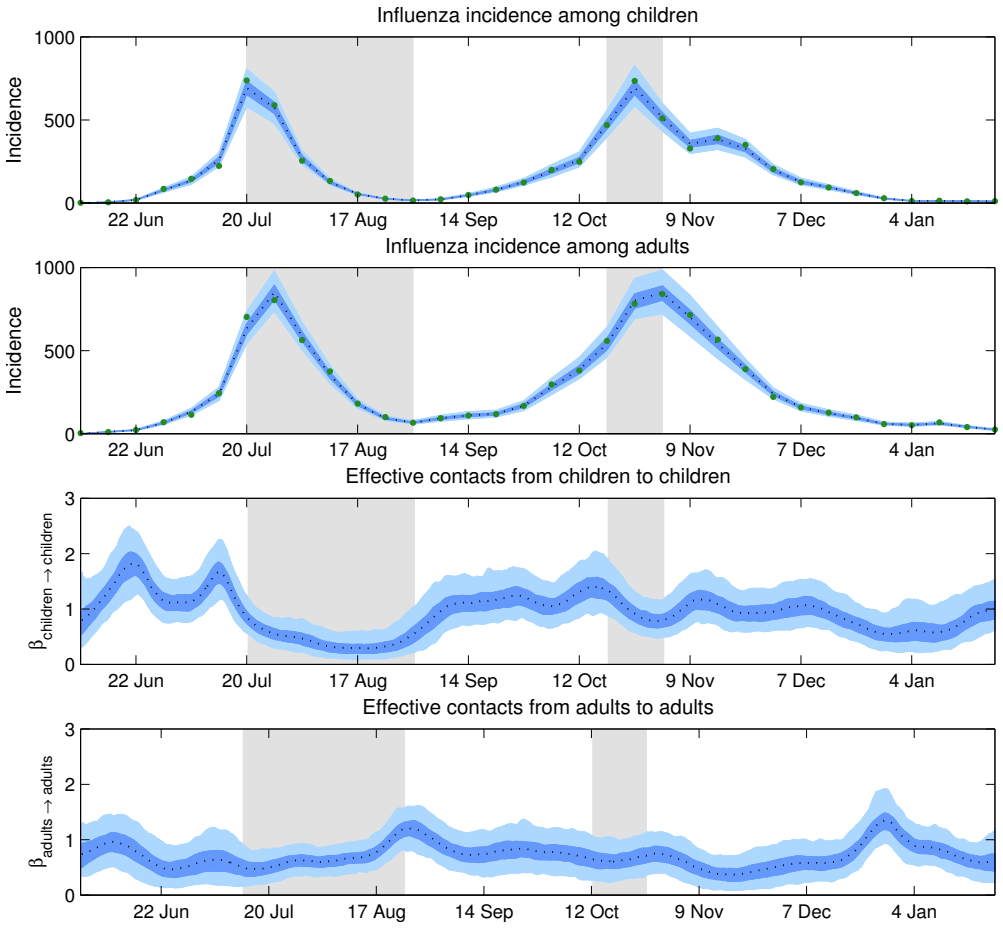

Figure 2.9: Offline estimates of the effective contact rate among children and adults during the A/H1N1 2009 influenza pandemic using a 2-classes age-structured model and age-specific incidence data. Green dots indicate observed incidence estimates among each age group provided by the Health Protection Agency (first and second panels). Black dotted lines indicate the mean of the pointwise posterior density. Dark and light blue areas respectively indicate 50% and 95% credible intervals of the posterior density. Holidays are indicated by a light grey area. *First panel*: HPA estimates of the weekly total number of A/H1N1 influenza cases among children in London (per 100,000 inhabs.) *Second panel*: HPA estimates of the weekly total number of A/H1N1 influenza cases among adults in London (per 100,000 inhabs.) *Third panel*: offline estimates of the effective contact rate from children to children. *Fourth panel*: offline estimates of the effective contact rate from adults to adults.

	Posterior mean	Posterior median	Posterior 95% c.i.
k^{-1}	1.56	1.55	$[1.53; 1.60]$
γ^{-1}	1.00	1.00	$[0.92; 1.08]$
$\beta_{cc}(0)$	1.44	1.36	$[0.89, 2.30]$
$\beta_{aa}(0)$	1.40	1.39	$[1.25; 1.64]$
β_{ca}	0.30	0.31	$[0.16; 0.50]$
β_{ac}	0.32	0.32	$[0.18; 0.48]$
$E_c(0)$	2.1×10^{-5}	1.9×10^{-5}	$[1.3 \times 10^{-5}; 3.8 \times 10^{-5}]$
$I_c(0)$	1.2×10^{-5}	1.4×10^{-5}	$[0.6 \times 10^{-5}; 2.9 \times 10^{-5}]$
$R_c(0)$	0.13	0.12	$[0.06; 0.26]$
$E_a(0)$	2.1×10^{-5}	2.0×10^{-5}	$[1.1 \times 10^{-5}; 3.5 \times 10^{-5}]$
$I_a(0)$	1.0×10^{-5}	1.0×10^{-5}	$[0.3 \times 10^{-5}; 1.5 \times 10^{-5}]$
$R_a(0)$	0.16×10^{-5}	0.16	$[0.09; 0.28]$
σ_c	0.11	0.10	$[0.08; 0.15]$
σ_a	0.08	0.08	$[0.05; 0.11]$

Table 2.4: Mean, median and 95% confidence intervals for the parameters of the structured model applied to the A/H1N1 pandemic data

2.6 Coupling time-varying parameters, stochastic rates and demographic stochasticity: towards an extended framework

As presented in the Introduction, the authors of Breto et al. (2009) have proposed an alternative to the infinite population approximation made in this Chapter. Under this framework, the finite and discrete nature of the population is preserved and the demographic stochasticity is reflected with a multinomial distribution allowing for multiple reactions over short periods of time δ. In this Section, we extend the use of time-varying parameters in this alternative setting to allow the exploration of parameter variations in conjunction with demographic stochasticity. We also derive a general formulation allowing for both stochastic rates and time-varying parameters for a complete description of environmental stochasticity. Although identifiability issues could arise if too many sources of uncertainty are introduced in the model, in the presence of strongly informative data and in specific cases it may be profitable to capture environmental stochasticity in different manners. Lastly, we propose a generic way to extend the use of the continuous-time Extended Kalman Filter in these more general settings by providing SDE approximations driven by Brownian motion for the different models and types of stochasticity. Generic expressions will be provided, and they will be illustrated on the following SIR density-dependent jump process model:

Reaction	Effect	Rate
Infection	$(S_t, I_t, R_t) \to (S_t - 1, I_t + 1, R_t)$	$\beta \frac{S_t}{N} I_t$
Recovery	$(S_t, I_t, R_t) \to (S_t, I_t - 1, R_t + 1)$	γI_t

The model above is generalised with the following notations:

Reaction	Effect	Rate
reaction 1	$z_t \to z_t + k^{(i_1,j_1)}$	$r^{(i_1,j_1)}(z_t, \theta)$
...	...	...
reaction l	$z_t \to z_t + k^{(i_l,j_l)}$	$r^{(i_l,j_l)}(z_t, \theta)$
...	...	...
reaction m	$z_t \to z_t + k^{(i_m,j_m)}$	$r^{(i_m,j_m)}(z_t, \theta)$

As previously mentioned, we concentrate on density-dependent models where rates can be reformulated as $r_i(z_t) = N\dot{r}_i(\dot{z}_t)$ with $\dot{z}_t = z_t/N$. For example, by noting $s_t = S_t/N$ and $i_t = I_t/N$ in the SIR model:

$$\begin{aligned} \dot{r}_1(s_t, i_t, r_t) &= \beta s_t i_t \\ \dot{r}_2(s_t, i_t, r_t) &= \gamma i_t \end{aligned} \tag{2.10}$$

2.6.1 Generalisation of the use of time-varying parameters

We start by proposing a general model under the infinite population assumption that combines stochastic rates and time-varying parameters:

$$\underline{\text{General framework for environmental stochasticity}}$$
$$\underline{\text{under infinite population approximation}}$$

$$\begin{cases} x_t^{\Gamma^{(i,j)}} = d\Gamma_t^{(i,j)} \quad for\ (i,j) \in \mathcal{R}^e \\ x_{t+\delta}^{\theta_t} = x_t^{\theta_t} + \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + L^{\theta_t} dB_t^{Q^{\theta_t}} \\ d\dot{z}_t = \sum\limits_{(i,j)\in\mathcal{R}^e} k^{(i,j)} \dot{r}^{(i,j)}(\dot{z}_t, x_t^{\theta_t}, \theta) x_t^{\Gamma^{(i,j)}} + \sum\limits_{(i,j)\in\mathcal{R}\setminus\mathcal{R}^e} k^{(i,j)} \dot{r}^{(i,j)}(\dot{z}_t, x_t^{\theta_t}, \theta)dt \end{cases} \tag{2.11}$$

This model takes the form of a hybrid stochastic differential equation driven by Gaussian and Gamma noise. Alternatively, under the multinomial approximation of the demographic stochasticity, we propose to extend the definition of the infinitesimal generator introduced in Breto et al. (2009) by combining stochastic rates and time-varying parameters to model environmental stochasticity:

General framework for environmental stochasticity
under multinomial approximation

$$\begin{cases} x_t^{\Gamma^{(i,j)}} = d\Gamma_t^{(i,j)} \quad \text{for } (i,j) \in \mathcal{R}^e \quad (x_t^{\Gamma^{(i,j)}} \equiv \delta \;\; if \;\; (i,j) \in \mathcal{R} \setminus \mathcal{R}_e) \\ x_{t+\delta}^{\theta_t} = x_t^{\theta_t} + \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + L^{\theta_t} dB_t^{Q^{\theta_t}} \\ p(x_{t+\delta}^{N^{(i,j)}} = x_t^{N^{(i,j)}} + n_{(i,j)} | x_t^{N^{(i,j)}}, x_t^{\Gamma^{(i,j)}}, z_t) = E\left[\prod_{i=1}^{c}\left\{M^{(i)}\left(1 - \sum_{k\neq i} p^{(i,k)}\right)^{\overline{n}^i} \prod_{j\neq i}\left(p^{(i,j)}\right)^{n^{(i,j)}}\right\}\right] + o(\delta) \\ z_{t+\delta} = z_t + \sum_{(i,j)\in\mathcal{R}} k^{(i,j)} x_t^{N^{(i,j)}} \end{cases}$$

using the following notations, with $r^{(i,j)} \equiv 0$ if $(i,j) \notin \mathcal{R}$:

$$\overline{n}^i = z_t^{(i)} - \sum_{k\neq i} n^{(i,k)}$$

$$p_t^{(i,j)} = p^{(i,j)}\left(r^{(i,j)}(z_t, , x_t^{\theta_t}, \theta) x_t^{\Gamma^{(i,j)}}\right) = \left(1 - \exp\left\{-\sum_k r^{(i,k)} x_t^{\Gamma^{(i,j)}}\right\}\right) r^{(i,j)} x_t^{\Gamma^{(i,j)}} / \sum_k r^{(i,k)} x_t^{\Gamma^{(i,j)}}$$

$$M^{(i)} = \begin{pmatrix} z_t^{(i)} \\ n^{(i,1)} \; ... \; n^{(i,i-1)} \; n^{(i,i+1)} \; ... \; n^{(i,c)} \; \overline{n}^i \end{pmatrix} \quad \text{(multinomial coefficient)}$$

The validity of the Markov chain defined by this infinitesimal generator under the additional introduction of a time-varying component $x_t^{\theta_t}$ in the rates $r^{(i,j)}$ will be the subject of subsequent theoretical work, to ensure that the stable properties of numerical solutions as δ tends to 0 are the properties of a properly defined limiting continuous time process.

2.6.2 Diffusion approximation of the demographic stochasticity

In order to provide an SDE approximation of the demographic stochasticity, we rely on the theoretical results on state-dependent Markov jump processes presented in Ethier and Kurtz (1986). They have been more recently adapted in a structured SIR epidemic modelling setting in Dargatz (2007), and are generalised in this section. Jump process models can be characterised by their master equation:

$$\frac{\partial}{\partial t} P(z_t) = \sum_{(i,j)\in\mathcal{R}} r^{(i,j)}(z_t - k^{(i,j)}, \theta) P(z_t - k^{(i,j)}) - \sum_{(i,j)\in\mathcal{R}} r^{(i,j)}(z_t, \theta) P(z_t) \quad (2.12)$$

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,

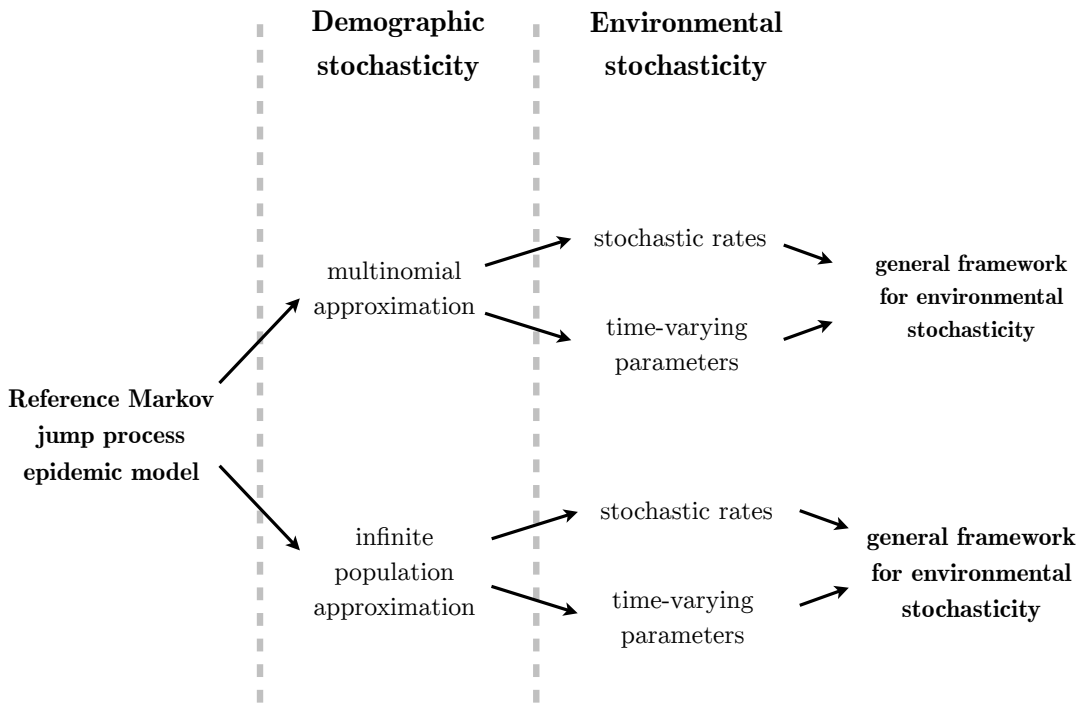

Figure 2.10: Representation of the different models introduced in the thesis

the master equation becomes:

$$\begin{aligned}\frac{\partial}{\partial t}P(S_t, I_t, R_t) = &\beta\frac{(S_t+1)}{N}(I_t-1)P(S_t+1, I_t-1, R_t)\\ &+\gamma(I_t+1)P(S_t, I_t+1, R_t-1)\\ &-\beta\frac{S_t}{N}I_tP(S_t, I_t, R_t)\\ &-\gamma I_tP(S_t, I_t, R_t)\end{aligned} \tag{2.13}$$

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

$$\begin{aligned}\frac{\partial}{\partial t}P(s_t, i_t, r_t) = &\frac{1}{\varepsilon}\beta(s_t+\varepsilon)(i_t-\varepsilon)P(s_t+\varepsilon, i_t-\varepsilon, , r_t)\\ &+\frac{1}{\varepsilon}\gamma(i_t+\varepsilon)P(s_t, i_t+\varepsilon, r_t-\varepsilon)\\ &-\frac{1}{\varepsilon}\beta s_ti_tP(s_t, i_t, r_t)\\ &-\frac{1}{\varepsilon}\gamma i_tP(s_t, i_t, r_t)\end{aligned} \tag{2.14}$$

The diffusion approximation relies on the limit of this expression when $\varepsilon \to 0$ while N is kept constant. The author of Dargatz (2007) shows that in this case, the former master

equation converges to the following partial differential equation:

$$\begin{aligned}\frac{\partial}{\partial t}P(s_t,i_t,r_t) = &\frac{\partial}{\partial s}\beta s_t i_t P(s_t,i_t,r_t) - \frac{\partial}{\partial i}(\beta s_t i_t - \gamma i_t)P(s_t,i_t,r_t)\\ &+ \frac{1}{2}\frac{\partial^2}{\partial s^2}\frac{1}{N}\beta s_t i_t P(s_t,i_t,r_t)\\ &- \frac{1}{2}\frac{\partial^2}{\partial i^2}\frac{1}{N}(\beta s_t i_t - \gamma i_t)P(s_t,i_t,r_t)\\ &- \frac{\partial^2}{\partial s \partial i}\frac{1}{N}\beta s_t i_t P(s_t,i_t,r_t),\end{aligned} \tag{2.15}$$

which is equivalent to

$$\frac{\partial}{\partial t}P(s_t,i_t,r_t) = -\frac{\partial}{\partial x}[\dot{A}(s_t,i_t,r_t)P(s_t,i_t,r_t)] + \frac{1}{2}\frac{\partial}{\partial x}\frac{\partial}{\partial x}[\dot{\Sigma}(s_t,i_t,r_t)P(s_t,i_t,r_t)] \tag{2.16}$$

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

Following Kloeden and Platen (1999), 2.16 is a Fokker-Planck equation corresponding to a diffusion process that is a solution of

$$d\dot{z}_t = \dot{A}(\dot{z}_t)dt + Ld\dot{B}_t^{Q^d} \tag{2.17}$$

Here, we follow the formalism of Särkkä (2007) where $d\dot{B}_t^{Q^d}$ is a Brownian motion with diffusion matrix $\dot{Q}^d$ and L is a stoichiometric dispersion matrix such that $L\dot{Q}^d L' = \dot{\Sigma}$:

$$\dot{Q}^d(s_t,i_t) = \frac{1}{N}\begin{pmatrix} \beta s_t i_t & 0 \\ 0 & \gamma i_t \end{pmatrix} \quad and \quad L = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \tag{2.18}$$

Equation 2.17 can be transposed in the natural scale of $z_t = [S_t, I_t, R_t]^T$, with $A = N\dot{A}$ and $Q^d = N^2\dot{Q}^d$:

$$dz_t = A(z_t)dt + LdB_t^{Q^d} \tag{2.19}$$

This result can be generalised based on the density-dependance property of rates $(r^{(i,j)})$. 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 2.19 with dBt being a multivariate Brownian motion with diffusion matrix $Q^d = Ndiag\{\dot{r}^{(i,j)}, (i,j) \in \mathcal{R}\} = diag\{r^{(i,j)},$

$(i,j) \in \mathcal{R}\}$, and L being the $c \times m$ stoichiometric matrix which columns are the stoichiometric vectors $k^{(i,j)}$ with $(i,j) \in \mathcal{R}$. Additionally, the vector $A(t)$ is determined by:

$$A(z_t) = \sum_{(i,j)\in\mathcal{R}} k^{(i,j)} r^{(i,j)}(z_t, \theta) \tag{2.20}$$

Lastly, the resulting expression for Σ is the following:

$$\Sigma(z_t) = LQ^d L' = \sum_{(i,j)\in\mathcal{R}} k^{(i,j)} r^{(i,j)}(z_t, \theta) k^{(i,j)T} \tag{2.21}$$

2.6.3 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:

$$d\dot{z}_t = \dot{A}(\dot{z}_t)dt \tag{2.22}$$

In the case of the SIR model:

$$\begin{cases} ds_t = -\beta s_t i_t dt \\ di_t = (\beta s_t i_t - \gamma i_t)dt \\ dr_t = \gamma i_t dt \end{cases} \tag{2.23}$$

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

$$\begin{cases} ds_t = -\beta s_t i_t d\Gamma_t \\ di_t = \beta s_t i_t d\Gamma_t - \gamma i_t dt \\ dr_t = \gamma i_t dt \end{cases} \tag{2.24}$$

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:

$$\begin{cases} ds_t = -\beta s_t i_t dt - \sigma\beta s_t i_t dB_t^{(1)} \\ di_t = (\beta s_t i_t d\Gamma_t - \gamma i_t)dt + \sigma\beta s_t i_t dB_t^{(1)} \\ dr_t = \gamma i_t dt \end{cases} \tag{2.25}$$

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

$$dz_t = A(z_t)dt + L^e dB_t^{Q^e} \tag{2.26}$$

L^e is the $c \times Card(\mathcal{R}^e)$ stoichiometric matrix which columns are the stoichiometric vectors $k^{(i,j)}$ with $(i,j) \in \mathcal{R}^e$. In addition, $dB_t^{Q^e}$ is a Brownian motion with diffusion matrix $Q^e = diag\{\sigma^2_{(i,j)} r_{(i,j)}(z_t)^2, (i,j) \in \mathcal{R}^e\}$ containing the variance of the different environmental noises imposed upon the system.

2.6.4 Gaussian SDE approximation in the general case

From the previous results, the general epidemic model under the infinite population assumption can be approximated with the following SDE:

$$\begin{cases} dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + L^{\theta_t} dB_t^{Q^{\theta_t}} \\ dz_t = A(z_t, x_t^{\theta_t}, \theta)dt + L^e dB_t^{Q^e}, \end{cases} \tag{2.27}$$

where A is the following vector :

$$A(z_t) = \begin{pmatrix} \sum_{(i,j)\in\mathcal{R}} k^{(i,j)} r^{(i,j)}(z_t) \end{pmatrix}. \tag{2.28}$$

Additionally, if demographic stochasticity is accounted for (typically through the multinomial approximation), the SDE approximation becomes

$$\begin{cases} dx_t^{\theta_t} = \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + L^{\theta_t} dB_t^{Q^{\theta_t}} \\ dz_t = A(z_t, x_t^{\theta_t}, \theta)dt + L dB_t^{Q}, \end{cases}$$

In this case, the matrices L and Q are constructed by concatenating the dispersion and diffusion matrices of the different sources of independent noises:

$$L = \begin{pmatrix} L^d & L^e \end{pmatrix} \quad and \quad Q = \begin{pmatrix} Q^d & 0 \\ 0 & Q^e \end{pmatrix} \tag{2.29}$$

2.7 Discussion

In this chapter, we examined epidemic models where some of the parameters are modelled with diffusions or integrals thereof. The main motivation was to account for time-varying drivers of epidemics (virus evolution, seasonality, schools closure, etc), while maintaining a simple interpretation. We presented a unified framework that supports

data-augmentation MCMC schemes based on fine partitions of the diffusion path; the associated approximation error can be controlled by the user without affecting the MCMC performance. This solution can be viewed as an extension of the approaches by Roberts and Stramer (2001) and Chib et al. (2006) to the more challenging observation regime of this Chapter. The consideration of the algorithms in a continuous-time setting revealed major issues associated with Gibbs data-augmentation schemes. These difficulties justifie the use of particle MCMC, which updates paths and parameters jointly, while pointing directions for future research on Gibbs schemes. We also presented a computational machinery based on the PMCMC algorithm (Andrieu et al., 2010), that was integrated in an adaptive MCMC context. In particular, we consider EKF-based initialisations of the adaptive algorithms that can offer substantial improvement, especially in cases with many static parameters. Increased stability and improved approximation of non-linear dynamics could be achieved be replacing replacing the Extended Kalman Filter by a Square Root Uscented Kalman Filter (Van Der Merwe and Wan, 2001). Nevertheless, the application of the PMCMC remains computationally demanding, and relying on a random walk exploration of the parameter space can be critically limiting, in particular for complex epidemic models. Situations of structural identifiability leading to hyperbolic shapes of the posterior distribution (see for example Figure 2.11) also illustrate a limitation of the robustness of the random walk Metropolis scheme: as the orientation of the posterior distribution of the two components θ_1 and θ_2 varies over the parameter space, it is necessary to rely on region-specific covariance matrices Σ_q driving the random walk, following the ideas of Roberts and Rosenthal (2009) or Girolami and Calderhead (2011).

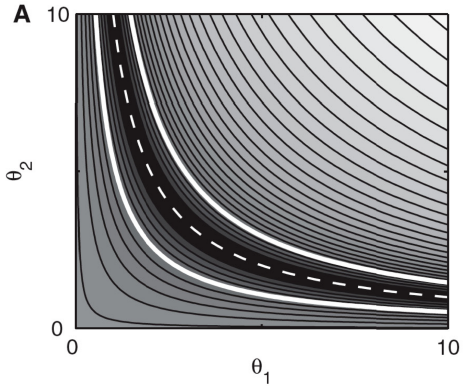

Figure 2.11: Posterior covariance of structurally unidentifiable parameters. Figure originating from Raue et al. (2009).

On a modelling perspective, our study starts from a simple SEIR model under the in-

finite population approximation. The proposed methodology can be viewed as an exploratory tool towards more structured models; e.g. the age-structured model of Section 5.3 that appears to be an improved representation of reality. This approach can help in developing richer models and testing alternative scenarios for public health interventions, or to bring further insights on extrinsic factors such as climate on the dynamics of epidemics. Moreover, this framework can support multiple sources of data, of potentially different nature: Rasmussen et al. (2011) has shown how epidemiological time series and genetic sequences can be combined for more informative estimates using the PMCMC algorithm. While we worked mainly with influenza time series, the developed methodology can be applied to other cases as illustrated in Chapter 3.

Time-varying parameters can be seen as an alternative to the stochastic rates introduced in Breto et al. (2009) to model environmental stochasticity. The combination of these two approaches allows to model high-frequency variations as well as longer term trends in time-varying drivers of epidemics. In the last Section, we have proposed a unified framework allowing for the combination of different types of environmental and demographic stochasticity, along with a generic and adaptive inferential framework. The application of the Extended Kalman Filter to this general setting relies on additional approximations: the diffusion approximation (Ethier and Kurtz, 1986) and a Gaussian approximation of independent Gamma increments. The robustness of these approximations and their impact on the contribution of the EKF-based pre-exploration of the posterior density will be the object of further research.

CHAPTER 3

A Bayesian approach to estimate changes in condom use from limited HIV prevalence data

3.1 Introduction

Significant resources are being committed to implement large-scale interventions against infectious diseases such as HIV/AIDS, that killed an estimated two million individuals in 2008 (UNAIDS, 2009). Although such interventions are implemented on a large scale because they are expected to work, increasing attention is given to their evaluation to understand what still needs to be done to control the epidemic and eventually achieve elimination, ensuring that resources are not waisted on strategies that do not work.

Even if antiretroviral therapy has become an important component of large scale prevention interventions, condom use and circumcision remain important strategies for reducing HIV transmission. While there are difficulties in estimating condom use trends accurately, due to biases inherent in self-reported behaviour (Turner and Miller, 1997; Zenilman et al., 1995; Hanck et al., 2008), its average level closely determines the spread of HIV (Boily et al., 2007). Thus, it is important to assess if trends in epidemiological data such as HIV prevalence can be used to infer the impact of interventions on risk behaviours that are susceptible to self-reported bias. This is motivated by the fact that directly observed quantities as HIV prevalence do not provide straightforward indications on the impact of an intervention. Indeed, an epidemic has an intrinsic dynamic, which can cause the prevalence to grow although an efficient intervention is being led if the intervention is introduced early in an epidemic. Alternatively, in a mature epidemic the prevalence can decrease even though on-going interventions are inefficient (Boily et al., 2002). How-

ever, the trajectory of CU over time, and especially since the beginning of a prevention programme, can shed light on the impact of the intervention and on the future trajectory of the epidemic. In this light, we apply a Bayesian methodology to trends in HIV prevalence data, focusing on the specific example of Avahan, India AIDS initiative, a large-scale HIV/AIDS intervention targeted to high risk groups.

The Avahan intervention was motivated by high levels of HIV prevalence amongst high-risk groups observed in southern India (typically over 20%) (Ramesh, et al., 2008), which lead to concerns about infections bridging to their long-term partners and the general population. The programme was launched by the Bill & Melinda Gates Foundation in 2003 (BMGF, 2008), and has targeted high-risk groups for HIV infection, in particular female sex workers (FSWs), by promoting and distributing free condoms. Different studies have been conducted to examine the impact of Avahan (Boily et al., 2007; Deering et al., 2008; Lowndes et al., 2010; Pickles et al., 2010), and to learn from it in order to inform future large-scale interventions. A key part of such evaluations is the examination of how risk behaviours, chiefly condom use (CU) defined as the proportion of sex acts protected by condoms at a given time have changed over the course of the intervention. However, this can be difficult to measure in practice. Baseline CU may be difficult to record when an intervention needs to be implemented rapidly, as happened with Avahan, or may be recorded only on few occasions. While those targeted by the intervention may be asked about their CU history (Lowndes et al., 2010), their answers may be subject to social desirability and recall biases. In principle, the total number of condoms sold or distributed can be enumerated (Bradley et al., 2010), but accurate records may not be available, condoms may be used for family planning by lower-risk individuals, and the distribution of condoms is not a guarantee of their correct usage (Bradley et al., 2010; Kumar et al., 2011). Thus, in addition to direct approaches through quantitative behavioural surveys or records of condom availability, model-based methods can be used to infer unobserved quantities of interest, such as CU, from observed quantities such as HIV prevalence using knowledge of the dynamics of large-scale epidemics. A first study in the context of Avahan was presented in Pickles et al. (2010). In this work, a deterministic dynamic model for HIV/sexually transmitted infections was formulated based on a compartmental representation incorporating heterogeneous sexual behaviour. The model included various parameters for which informative prior distributions were used. Prior elicitation was based on various data sources, such as previous literature (see Pickles et al. (2010) for more details) and serial cross-sectional surveys termed integrated behavioural and biological assessment (IBBA) conducted in the districts of India targeted by the intervention. The objective was to utilise this model and assess its ability to fit the available prevalence observations under three different hypothesised scenarios of evolution of CU.

The work we present in this Chapter operates in the same context as in Pickles et al. (2010), but focus is given on exploring the entire space of CU trajectories rather than considering three scenarios regarding its evolution. Similarly, the model formulation can also be put in a state-space setting where an underlying latent process (CU trajectory) is indirectly observed through the prevalence data, and the link between these quantities is given by the deterministic model for HIV infections. Inference in this context is a challenging task given the limited amount of HIV prevalence data aside from initial conditions (three or four observations in total) that are concentrated over a period of 6 years and are utilised to estimate a 25-years long trajectory. Various models for the CU trajectories were considered, including smooth and non-differentiable (yet continuous) choices. In the remainder of this Chapter, the term *trajectory prior* is used to refer to these models in order to avoid confusion with the deterministic HIV model. Focus is given on estimating the amplitude of the change in CU since 2003 (the start of Avahan) in order to assess the impact of the Avahan intervention on CU. The properties of the estimators arising from the methodology introduced in Chapter 2 are studied via simulations, and the performance is assessed from a decision-making perspective through their sensitivity and specificity in detecting strong changes in CU.

The next section presents the models introduced in this Chapter, the data that are typically available for such studies, and the way in which prior information is incorporated. The computational techniques, mainly the particle MCMC algorithm, are also presented. The methodology developed to compare the performance of the proposed trajectory priors is presented in Section 3.3, and the results from this study are introduced in Section 3.4 along with an application to real data from the Indian AIDS initiative Avahan. In Section 3.5, the analysis is extended to ten districts targeted by the Avahan intervention, and to the Bayesian synthesis of model-based and survey-based CU estimates. Finally Section 3.6 concludes with some relevant discussion.

This chapter is directly adapted from an article currently under review, co-written with Kostas Kalogeropoulos (London School of Economics), Peter Vickerman (London School of Hygiene and Tropical Medicine), Michael Pickles (Imperial College) and Marie-Claude Boily (Imperial College). Section 3.5 is a subsequent development of this work.

3.2 Models and methods

3.2.1 HIV transmission model for female sex workers

We use a structured deterministic model of HIV transmission in a stable but open population of sex workers and their clients, under the infinite population approximation. The model structure accounts for high-risk (HR) and low-risk (LR) FSWs, who have different

numbers of clients (M). State variables $S^{HR}, I^{HR}, S^{LR}, I^{LR}, S^{M}$ and I^{M} respectively represent the absolute number of susceptible and infected individuals among high-risk FSWs, low-risk FSWs, and clients. This model is parameterised using data from serial cross-sectional bio-behavioural surveys (IBBAs) in Mysore district in southern India (Ramesh, et al., 2008). Some uncertainty remains about these biological and behavioural parameters, which is reflected on the estimates of CU using a Bayesian approach (De Angelis et al., 1998). As motivated in Vickerman et al. (2010), low-risk individuals uninvolved directly in sex work are ignored as they have little influence on the dynamics of the epidemic. Each individual in these three groups is either susceptible to HIV infection, infectious, or retired either due to death or ceasing commercial sexual activity. The flow-diagram corresponding to high-risk FSWs is shown in Figure 3.1. In addition, individuals that either decease or stop being involved in commercial sex are replaced by susceptible ones, maintaining the population at risk at a constant size. As illustrated in Figure 3.1, the force of infection β is a function of a number of different parameters:

- $NbClients^{HR}$ and $NbClients^{LR}$: number of clients of FSWs per month, which differs for high risk and low risk FSW
- $NbEncounters$: mean number of encounters with a FSW per client per month
- $NbActs$: number of acts per client encounter
- $p^{tr}_{M\to F}$ or $p^{tr}_{F\to M}$: probability of HIV transmission from male to female or female to male respectively during an unprotected sex act
- $Cond_{eff}$: efficacy of condoms in protecting against transmission of HIV per sex act
- $CU(t)$: proportion of commercial sex acts with FSWs that are protected by condoms, that we allow to vary in time (parameter that we want to estimate)

Additionally, the transmission dynamics of HIV will depend on the following durations:

- μ^{-1}_{FSW} or μ^{-1}_{M}: average duration of sexual activity as a sex worker / client
- α^{-1}: average life expectancy with HIV

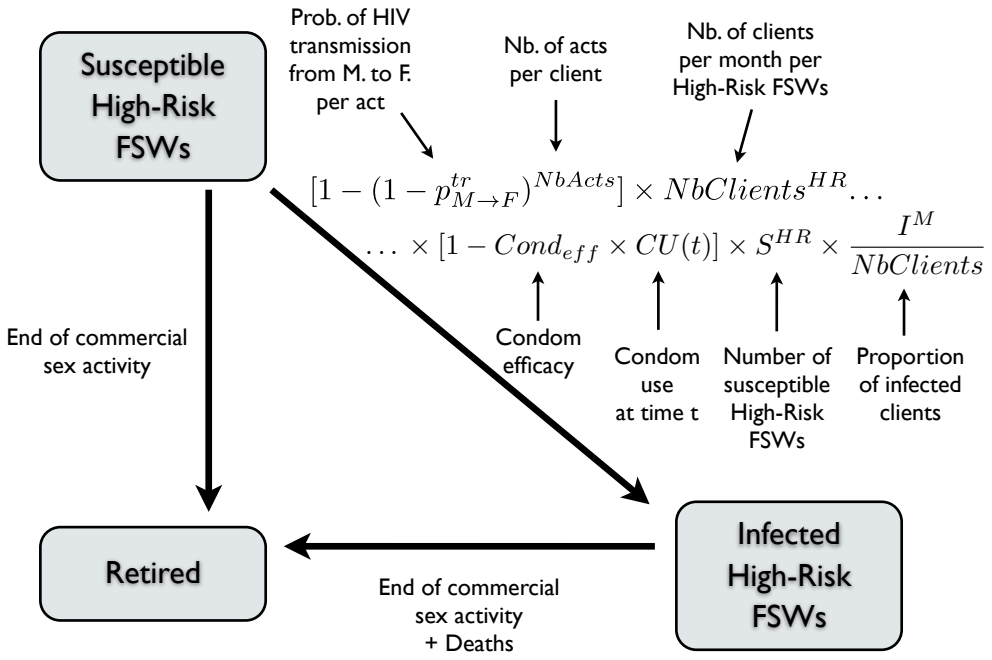

Figure 3.1: Flow-diagram of the model for high-risk FSWs. Transmission dynamics for low-risk FSWs and clients are defined similarly

In mathematical terms, the model can be defined with a set of differential equations:

$$\begin{cases} \frac{dS_t^{LR}}{dt} = -\beta^{LR} S_t^{LR} \frac{I_t^M}{Tot_M} + (\mu_{FSW} + \alpha) I_t^{LR} \\ \frac{dI_t^{LR}}{dt} = \beta^{LR} S_t^{LR} \frac{I_t^M}{Tot_M} - (\mu_{FSW} + \alpha) I_t^{LR} \\ \frac{dS_t^{HR}}{dt} = -\beta^{HR} S_t^{HR} \frac{I_t^M}{Tot_M} + (\mu_{FSW} + \alpha) I_t^{HR} \\ \frac{dI_t^{HR}}{dt} = \beta^{HR} S_t^{HR} \frac{I_t^M}{Tot_M} - (\mu_{FSW} + \alpha) I_t^{HR} \\ \frac{dS_t^M}{dt} = -\beta^M S_t^M \left(Prop_{LR}(\frac{I_t^{LR}}{Tot_{LR}}) + Prop_{HR}(\frac{I_t^{HR}}{Tot_{HR}}) \right) + (\mu_M + \alpha) I_t^M \\ \frac{dI_t^M}{dt} = \beta^M S_t^M \left(Prop_{LR}(\frac{I_t^{LR}}{Tot_{LR}}) + Prop_{HR}(\frac{I_t^{HR}}{Tot_{HR}}) \right) - (\mu_M + \alpha) I_t^M \end{cases} \tag{3.1}$$

Using the following notations:

$$\beta^{LR} = [1-(1-p^{tr}_{M\to F})^{NbActs}]NbClients^{HR}(1-Cond_{eff}CU_t)$$
$$\beta^{HR} = [1-(1-p^{tr}_{M\to F})^{NbActs}]NbClients^{LR}(1-Cond_{eff}CU_t)$$
$$\beta^{M} = [1-(1-p^{tr}_{F\to M})^{NbActs}](\frac{NbClients^{HR}+NbClients^{LR}}{2})\frac{Tot_F}{Tot_M}(1-Cond_{eff}CU_t)$$
$$Prop_{LR} = \frac{NbClients^{HR}Tot_{LR}}{NbClients^{HR}Tot_{LR}+NbClients^{LR}Tot_{HR}}$$
$$Prop_{HR} = \frac{NbClients^{LR}Tot_{LR}}{NbClients^{HR}Tot_{LR}+NbClients^{LR}Tot_{HR}}$$

In the model above, three types of constant parameters are involved:

- Initial prevalence among the different groups of interest in 1985:

$$\theta_{i.c.} = \{S^{LR}_{1985}, I^{LR}_{1985}, S^{HR}_{1985}, I^{HR}_{1985}, S^{M}_{1985}, I^{M}_{1985}\}$$

- Constant parameters describing the biological and behavioural determinants of HIV:

$$\theta_{tr.} = \{\mu_{FSW}, \mu_M, \alpha, NbClients^{HR}, NbClients^{LR}, p^{tr}_{M\to F}, p^{tr}_{F\to M}, NbActs, Cond_{eff}\}$$

- Parameters that play a role in the CU trajectory priors for: θ_{CU}

All three components $\theta_{i.c.}, \theta_{tr.}$, and θ_{CU} are integrated into a global vector of constant parameters, denoted by θ. Under this notation, the trajectory $z_{0:n}$ of the space vector $z_t = \{S^{LR}_t, I^{LR}_t, S^{HR}_t, I^{HR}_t, S^{M}_t, I^{M}_t\}$ is defined as a deterministic function of θ and $CU_{0:n}$ (i.e. $z_{0:n} = f(\theta, CU_{0:n})$) through an HIV transmission model, and is compared with the available observations, denoted by $y_{1:n}$. Note that the function $f(.)$ is not available in closed form but can be obtained given the trajectory of $CU_{0:n}$ by solving the above ordinary differential equations (ODE). More specifically, we introduce a time discretisation with equidistant points of time step δ resulting in a discretised skeleton of CU denoted $CU^{discr}_{0:n} = \{CU_{t_0}, CU_{t_0+\delta}, CU_{t_0+2\delta}, .., CU_{t_n}\}$. The partition of the CU trajectory can be made arbitrarily fine by the user-specified parameter δ to limit the approximation error induced by the time discretisation.

Assigning a model for the observation error provides the likelihood of the observation $y_{1:n}$ conditionally on the CU trajectory: $p(y_{1:n}|\theta, CU_{0:n})$. In this Chapter, we use a binomial distribution considering that prevalence estimates are derived from a random sample of 425 FSWs or clients in the Mysore district. More specifically if we denote $prev^{model}$ the value of prevalence estimated through the HIV transmission model ($prev^{model}_i = (I^{LR}_i +$

$I_i^{HR})/2$ if prevalence among FSWs is observed at time t_i, $prev_i^{model} = I_i^M$ if prevalence among clients is observed at time t_i), the observation model is the following:

$$425 \times prev_i^{obs} \sim Bin(425, prev_i^{model}) \tag{3.2}$$

Overall, the model appearing of Figure 3.1 and Equations 3.1 is a simplified version of the one in in (Pickles, et al., 2010). This was done mainly for parsimony reasons; models of increased complexity can be used provided that there is adequate information on their parameters. More details on the informative priors that have been used are provided in Section 3.2.3.

3.2.2 Trajectory priors for condom use

In this Chapter, we introduce three different formulations for the evolution of the CU trajectory. Our first trajectory prior, denoted BM, assigns a Brownian motion to $CU_{0:n}$, transformed to take values in the real line. Initial considerations in Pickles et al. (2010) and classic literature on smoothly growing quantities also motivated the introduction of alternative formulations based on sigmoid-shaped growth curves. Hence, the second trajectory prior, denoted by dBR, is based on the generalised Bertalanffy-Richards model; see for example (Garcia, 1983; Yuancai et al., 1997). In order to enrich this context and address estimation issues that can be encountered with the dBR (Lei and Zhang, 2004), we also consider an alternative empirical sigmoid curve (dSigm). In what follows, we denote with x_t the latent process that drives the CU trajectory, which in turn is linked with the prevalence observations through the model in Section 3.2.1.

Brownian motion (BM)

The first formulation assigns a Brownian motion to a transformed version of the CU trajectory. As the latter has to be constrained in the [0,1] region, we work with the *logit* transformation of CU_t, denoted by:

$$\begin{aligned} CU_t &= \frac{\exp(x_t)}{1+\exp(x_t)} \\ dx_t &= \sigma dB_t \end{aligned} \tag{3.3}$$

The use of diffusion processes to describe time varying quantities in contexts associated with uncertainty has been explored in Chapter 2 of this thesis. It can also be seen as a prior according to which x_t is a random walk with continuous, yet non-differentiable trajectories. It is used here as an attempt to incorporate a limited amount of prior information on the shape of the trajectory. It can also be used as an exploration tool for potential modelling-remodelling steps towards more informative formulations. Variations of this

formulation may include smoother diffusion models, by taking integrals of the Brownian motion, or alternative transformations such as the *probit* link. We note at this point that very little information is available on the volatility in (3.3) which is determined mostly by its prior. More details are provided in Sections 3.2.3 and 3.3.

Deterministic Bertallanfy-Richards function (dBR)

Qualitatively, reconstruction of CU trends by alternative methods (Lowndes et al., 2010; Bradley et al., 2010) suggest that CU was quite low in 1985, and has grown over the recent year. The above motivated the use of a growth curve parametric model instead of the Brownian motion diffusion. This is in line with various approaches in modelling quantities that are smoothly growing in time in different contexts such as biology (Zwietering et al., 1990), marketing (Lessne and Hanumara, 1988) and epidemiology (Omran, 1971). We use the generalised Bertalanffy-Richards (BR) family (Richards, 1959; Garcia, 1983) that can be written as:

$$CU_t = \eta(1 - Be^{-kt})^{\frac{1}{1-m}} \tag{3.4}$$

or else, in differential equation framework:

$$\begin{aligned} CU_t &= [(1-m)x_t + \eta^{1-m}]^{\frac{1}{1-m}} \\ dx_t &= -kx_t dt \end{aligned} \tag{3.5}$$

This family contains various growth curves, including the logistic ($m = 2$) and Gompertz ($m \to \infty$) functions. The growth curve can be parameterised by four quantities: the initial value of CU_0, the time of inflection t_{in}, the value of CU after an infinite time (η, also termed as the asymptote), and the *shape* or *allometric* parameter m. Note that the time of inflection can be related to the parameter k by the following equation:

$$k \times t_{in} = \log(\frac{B}{1-m}) \tag{3.6}$$

Furthermore, this definition implies that the initial value CU_0 is lower than $m^{\frac{1}{1-m}}\eta$. In order to focus on sigmoid-shaped growth curves, we restrict our attention to cases where $m \geq 1$ (Yuancai et al., 1997). For illustration, the slope of the curve at its inflection point is a complex and non-monotonous function of m:

$$\frac{CU}{dt}(t_{in}) = \eta k m^{\frac{m}{1-m}} = \frac{\eta}{t_{in}} \log(\frac{1 - (\frac{CU_0}{\eta})(1-m)}{1-m}) m^{\frac{1}{1-m}} \tag{3.7}$$

Deterministic empirical sigmoid curve (dSigm)

An empirical sigmoid model is also considered to address the potential difficulties that can arise with the parameterisation of the dBR. Since growth models are used to study intrinsically growing objects, trajectories that are inexplicably stable for a long period of time and

that eventually start picking at a rapid pace are not typical under the BR formulations. It may however be the case that the extrinsic influence of the Avahan intervention induces a late increase of CU after fifteen years of stability or slow increase. Moreover, inference on the allometric parameter m in dBR can be problematic (Lei and Zhang, 2004). These may lead to underestimating the amplitude of a shift in CU under the potential influence of the Avahan intervention. For this reason, we also consider an alternative sigmoid curve, defined in the following way:

$$\begin{aligned} CU_t &= a + \frac{b}{c(1+x_t)} \\ dx_t &= -kx_t dt \end{aligned} \tag{3.8}$$

Here the model is parameterised by its baseline (CU_0), its asymptote (η), its time of inflection (t_{in}), and the increase rate (r), from which a, b and c can be computed:

$$\begin{aligned} a &= CU_0 - b \\ b &= (\eta - CU_0)c \\ c &= \frac{1}{1+e^{t_{in}/\tau}} \end{aligned} \tag{3.9}$$

The slope of the curve at its inflection point is now a simpler function of the model parameters than with the previous model (Equation 3.7):

$$\frac{dCU}{dt}(t_{in}) = \frac{\eta - CU_0}{4r} \tag{3.10}$$

Stochastic growth curves

It is also possible to combine the Brownian motion and the growth curve approaches using diffusions. Stochastic extensions of the dBR and dSigm model can be considered, in which the mean behaviour remains intact while some random perturbations are introduced through a stochastic differential equation. In order to ensure positivity, restrict CU_t below one and retain the link with deterministic dBR curve, a geometric Brownian motion can be used to replace equations (3.6) and (3.8)

$$dx_t = -kx_t dt + \sigma x_t dB_t \tag{3.11}$$

The stochastic growth curve defined by (3.5) and (3.11) was also mentioned in Garcia (1983). A convenient feature for both stochastic extension of dBR and dSigm is the fact that since

$$x_t = \frac{1}{1-m}(CU_t^{1-m} - \eta^{1-m}), \tag{3.12}$$

and x_t is strictly negative, the resulting CU trajectory is maintained strictly below η. Given the limited data at our disposal, these models can hardly be fitted in the context of this

Chapter. Nevertheless, they may be helpful in cases where more observations are available.

3.2.3 Priors

The parameters contained in $\theta_{i.c.}$ and $\theta_{tr.}$ cannot be identified from the prevalence observations only, so we assign informative priors on them. These are summarised in Table 3.1 and are similar to the priors used in Pickles et al. (2010). General quantities as transmission probability for unprotected acts, or life expectancy with HIV are based on previous literature, whereas priors concerning quantities that are more sociologically and geographically specific were estimated from cross-sectional individual-based surveys (IBBAs) in the district of Mysore.

The parameter vector θ includes an additional component, θ_{CU}, that contains the parameters for different models describing the CU trajectories, $\theta = \{\theta_{i.c.}, \theta_{tr.}, \theta_{CU}\}$. Although there is some information in the data for θ_{CU}, the posterior will depend on the prior to a large extent. As mentioned earlier, there is very little information on the volatility parameter of the BM formulation. Throughout this Chapter we used a Uniform prior between 0 and 0.5. As explained in more detail in Section 3.3.1, the parameter of main interest in this study is the quantity $\Delta CU = CU_{2009} - CU_{2003}$. Simulations suggest that, if we combine the BM approach with a $Unif(0, 1)$ prior for CU_0 (CU in 1985), this results in a symmetric prior on ΔCU that is centered around 0 with 2.5% and 97.5% points at ± 0.6 respectively. We considered it as a reasonably vague prior for ΔCU and evaluated the performance of the resulting model via the simulation experiments of Section 3.3. More diffuse priors can also be used by setting a larger value for the upper limit of the Uniform prior for σ. Regarding the parameters of the sigmoid curves, we used vague priors that are also shown in Table 3.1.

3.2.4 Computational schemes for implementation

The joint posterior density $p(\theta, CU^{dis}_{0:n}|y_{1:n})$ can be computed pointwise up to proportionality, using the HIV transmission model of Section 3.2.1 which links the prevalence observations with the CU trajectories, the trajectory priors of Section 3.2.2 and the remaining priors of Section 3.2.3. For the dBR and dSigm trajectory priors, it can be put in a non-linear regression framework, with the non-linear function being the solution of the ODE, and can therefore be implemented with standard software such as WinBUGS through WBDiff (Lunn, 2004). However this is not possible for the BM case where more involved techniques are required. Since the posterior probability density function is intractable, a data augmentation scheme can be utilised. This inference problem poses some challenges

HIV transmission model parameters definition	Notation	Range of uniform priors for the district of Mysore (Pickles et al. 2010)
Probability of transmission from M. to F. per act	$p^{tr}_{M\to F}$	0.0006-0.0055
Probability of transmission from M. to F. per act	$p^{tr}_{F\to M}$	0.0001-0.007
Condom efficacy per act	$Cond_{eff}$	80%-95%
Mean number of acts per clients	$NbActs$	1-2
Mean number of clients per high-risk FSW	$NbClients^{HR}$	46.6-54.0 clients/month
Mean number of clients per low-risk FSW	$NbClients^{LR}$	20-23.7 clients/month
Toral number of FSWs	$Tot_{LR} + Tot_{HR}$	2144
Cliens/FSW population ratio	$\frac{Tot_M}{Tot_{LR}+Tot_{HR}}$	7-19
Mean length of sexual activity as FSW	μ_F^{-1}	45-54 months
Mean length of sexual activity as client	μ_M^{-1}	154-191 months
Mean life expectancy after infection with HIV	α^{-1}	87-138.5 months
Initial proportion of infected FSWsin 1985	$\frac{HIV_{LR}+HIV_{HR}}{Tot_{LR}+Tot_{HR}}$	0%-5%
Initial proportion of infected clients in 1985	HIV_M/Tot_M	0%-5%
Condom trajectory priors parameters definition	**Notation**	**Prior**
Allometric parameters (dBR)	m	$\mathcal{N}(1, 10^6) \times \mathbb{I}_{]1,+\infty[}$
Growth rate (dSigm)	r	$\mathcal{N}(0, 10^6) \times \mathbb{I}_{]1,+\infty[}$
Asymptote (dBR, dSigm)	η	$Unif(0, 1)$
Initial Value (all trajectory priors)	CU_{t_0}	$Unif(0, 1)$
Time of inflection (dBR, dSigm)	t_{in}	$Unif(1985, 2009)$
Allometric parameters, initial conditions and asymptote (dBR)	(CU_{t_0}, η, m)	0 if $CU_{t_0} \geq m^{\frac{1}{1-m}}\eta$
Volatility (BM)	σ	$Unif(0, 0.5)$

Table 3.1: Table of priors for the different components of $\{\theta_{i.c.}, \theta_{tr.}, \theta_{CU}\}$

due to the high dimension of the discretised representation of $CU_{0:n}^{dis}$ and its strong correlation with the vector of constant parameters, θ. This correlation imposes problems to Gibbs schemes on θ and $CU_{0:n}^{dis}$, leading to extremely poor mixing and convergence properties. As motivated in Chapter 2, the particle MCMC algorithm (Andrieu et al., 2010) offers a solution by updating the two components jointly, thus reducing the problem to a small-dimensional MCMC on θ based on the estimates of the likelihood $\hat{p}_{pf}(y_{1:n}|\theta)$ that are provided by a particle filter.

Each iteration of the random walk Metropolis Hastings algorithm, operated in a transformed parameter space (log or logit) to ensure positivity or boundedness constraints on each parameter, requires an execution of the particle filter. Consequently, substantial computational cost is induced if the importance sampling covariance matrix Σ^q is ill-adapted. Adaptive approaches (Roberts and Rosenthal, 2009) can be used to tune Σ^q but they can require lengthy explorations of the target space. We propose to speed up this process by pre-exploration of a proxy posterior density $p^{EKF}(\theta|y_{1:n})$ relying on a Gaussian approximation of the dynamic system and the Extended Kalman filter methodology as illustrated and motivated in Chapter 2. A simple bootstrap version of the particle filter is used as it is not straightforward to consider data-driven transition proposals given the complex observation regime of our model.

3.3 Evaluation methodology based on ensemble simulations

Given the limited amount of prevalence data available (four or five prevalence observations, including initial conditions), it is very likely that the posterior densities will be influenced substantially by the choice of CU trajectory priors and their parameters. In this section we explore the performance of the proposed inferential mechanism via simulation-based experiments designed to mimic the behaviour of datasets typically encountered in the context of application studied. Clearly, the approach of this Chapter heavily relies on the HIV infection model and the results will be quite sensitive to its specification. We therefore set up the simulation experiments under the assumption that the model of Section 3.2.1, parameterised according to the priors of Section 3.2.3, is correct. Focus is given on quantities related with the CU trajectories that can be estimated from the samples of the posterior distribution provided by the MCMC algorithms of Section 3.2.4, under the different choices presented in Section 3.2.2. We also provide some discussion regarding the static parameters appearing in the CU trajectory priors.

3.3.1 Parameter of interest

By fitting each of the previously introduced models we obtain samples from the marginal posterior density $p^{meth}(CU_{0:n}^{dis}|y_{1:n})$ ($meth \in$ {dBR, dSigm, BM}). However, our interest mainly lies in the amplitude of the shift in CU between 2003 and April 2009 measuring the estimated increase in CU during the intervention, henceforth denoted by ΔCU. The posterior draws of CU trajectories can be transformed to provide samples from the posterior of this parameter of interest. The samples can then be used to form an estimator $\hat{\Delta}CU^{meth}$ of ΔCU, such as the posterior median of $\hat{p}^{meth}(\Delta CU|y_{1:n})$. In what follows we explore the frequentist properties of this estimator, derived from each of the trajectory priors.

It may also be of interest to assess the estimating capabilities, given the limited amount of data, for the hyperparameters of the various CU priors (CU_0, η, r, m, t_{in} and σ). It appears that the limited prevalence data contain information for some of them (CU_0, η, t_{in}), whereas some others are hard to estimate and are determined mostly by their prior (r, m and σ). Nevertheless, from application point of view interest lies mainly on ΔCU while the remaining quantities (in CU priors) can be regarded as nuisance parameters. Another appealing feature of ΔCU is that it appears in all models and therefore provides an omnibus quantity for comparison. Hence, inference properties of these parameters (CU_0, η, r, m, t_{in} and σ) are only studied indirectly through inference properties of ΔCU.

3.3.2 Measures of performance

The performance of each estimator $\hat{\Delta}CU^{meth}$ in estimating ΔCU is evaluated from the following criteria (where $L = 100$, the number of simulations):

$$\begin{aligned} Bias^{meth} &= \tfrac{1}{L}\textstyle\sum_i(\hat{\Delta}CU_t^{meth} - \Delta CU_t) \\ MSE^{meth} &= \tfrac{1}{L}\textstyle\sum_i(\hat{\Delta}CU_t^{meth} - \Delta CU_t)^2 \\ Std^{meth} &= \sqrt{MSE^{meth} - (Bias^{meth})^2} \end{aligned}$$

In addition to the quantities above we are also interested in assessing the discriminative ability of each model in detecting increases in CU. Focus is given to increases in CU that are at least as high as a pre-specified threshold T. When analysing the data, a researcher may decide that CU did increase more than T if the value of the estimator $\hat{\Delta}CU^{meth}$ is higher than a user-specified threshold t. Each decision mechanism may lead to different types of error and is therefore associated with a particular sensitivity and specificity. More specifically we can define the true and false positives in the following way

- Sensitivity (true positives rate) for t: $\frac{\sharp(\hat{\Delta}CU^{meth}>t, \Delta CU>T)}{\sharp(\Delta CU>T)}$
- Specificity (1 - false positives rate) for t: $\frac{\sharp(\hat{\Delta}CU^{meth}<t, \Delta CU<T)}{\sharp(\Delta CU<T)}$

We proceed by first reporting sensitivities and specificities corresponding to the case of $t = T$. This corresponds to saying that ΔCU is higher than T if its estimator is higher than T. We then use a range of different t's and obtain the sensitivity-specificity pair that corresponds to each of them. A lower detection threshold t will increase the sensitivity of the method, but it also increases the risk for false positives, and vice versa. These pairs are combined to form the Receiver Operating Characteristics (ROC) curve by plotting sensitivity versus 1-specificity. The area under the ROC curve (AUC) provides an overall measure of discriminatory power as it reflects the probability of correctly classifying a randomly chosen positive instance as higher than a randomly chosen negative one (Fawcett, 2006). For example, an AUC value of 50% indicates no power (i.e random choice).< This detailed procedure is repeated to assess the ability to detect two different levels of increase in CU, with T set to 20% and 40% respectively.

3.3.3 Simulation procedure

The performance of the estimators derived from the different trajectory priors is measured using a set of simulated experiments where CU trajectories are sampled from a given growth curve model, and parameters from $\theta_{i.c.}$ and $\theta_{tr.}$ are sampled following their prior distributions. To maximise the utility of this test procedure for future application of this methods to help evaluate Avahan in different districts, only plausible and realistic CU trajectories are considered: cases with prevalence in 2010 between 2% and 40% and with CU shifts that occurred after 1995. Furthermore, the test trajectories have been sampled so that ΔCU regularly spans the $[0; 0.6]$ interval.

For each of these experiments, an epidemic is simulated to provide observations (y_i^{sim}) replicating the observation scheme applied in Mysore: three prevalence estimates among female sex workers and one among clients, concentrated during the period of the intervention (step 2 of Figure 3.2). From these observations, the MCMC algorithm is applied to each method to sample from $p(CU_{1:n}^{meth}|y_{1:n}^{sim})$ (step 3 of Figure 3.2). Then, given the posterior CU samples the estimators $\hat{\Delta}CU^{meth}$ can be computed and compared to their true counterparts ΔCU (step 4 of Figure 3.2) by calculating the measures of performance of the previous subsection. Examples of ROC curves obtained in such manner are provided for the Brownian motion trajectory prior in Figure 3.3.

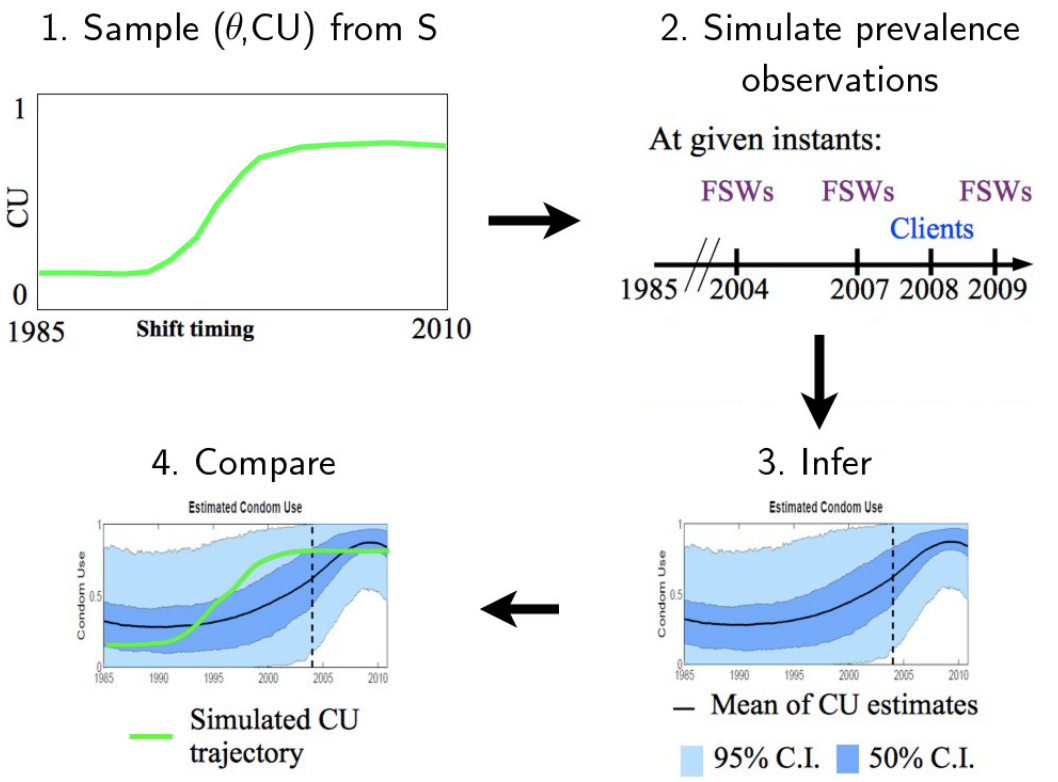

Figure 3.2: Simulation procedure, repeated 100 times for each trajectory prior

3.4 Results

3.4.1 Comparison of the CU trajectory models from ensemble simulations

The results of the simulation experiments are presented in Tables 3.2 and 3.3. Table 3.2 focuses on the frequentist properties of the estimators, derived from the median of the posterior densities provided by each trajectory prior, and reports the bias, the standard deviation and the MSE of each estimator. Table 3.3 concentrates on the ability of the model to classify shift amplitudes of CU from 2003 to 2009 in the right order (AUCs), and more specifically on the risk of overstating versus understating the quantity of interest. In other words, we aim to address questions such as *was the shift in condom use during the intervention over 0.2 (0.4)?*, via the corresponding sensitivity and specificity and the resulting AUC.

The first table of this section (Table 3.2) suggests that no model tends to consistently overstate ΔCU as all biases are negative. More precisely, the dBR model tends to strongly understate the shift in amplitude, by 0.23 in average. The bias of the dSigm model is

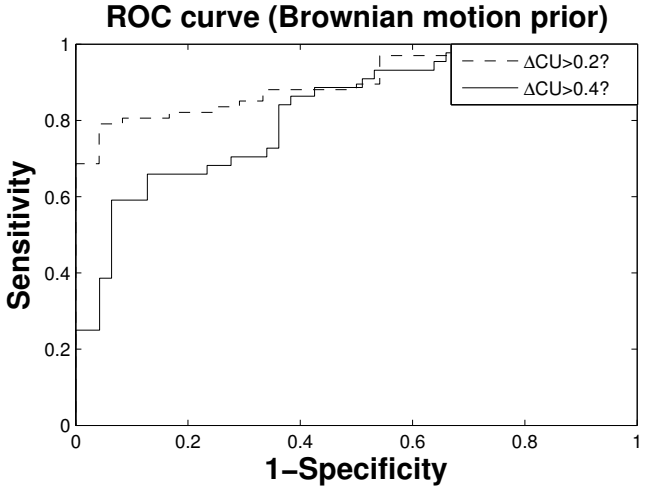

Figure 3.3: ROC curve when testing for $\Delta CU > 0.2$ and $\Delta CU > 0.4$, under Brownian motion trajectory prior. This curve was estimated from 100 simulations. Very similar shapes are obtained for the alternative trajectory priors.

smaller (-0.17), but optimal results are obtained with the Brownian Motion model (-0.13). Similarly, in terms of MSE, the performance of the BM model is better. Figure 3.4 shows the bias, estimated from 100 simulations, of each trajectory prior as a function of the true amplitude of the shift in CU. It suggests that the bias increases as a function of the size of the true amplitude of the shift in CU, and that the ranking of the different models is consistent across different configurations (from no shift in CU to moderate and high shifts in CU). If, for example, the true shift is $+0.5$, it is on average underestimated by 0.15 with the best method (BM) and more than 0.35 points with the BR method.

Table 3.3 and Figure 3.4 suggest that all estimators based on the median of the posterior density of $p(\Delta CU|y_{1:n})$ have good distinguishing power: the AUC is between 0.82 and 0.91 in all cases. In line with the results of Table 3.2, the estimates provided by the BM model achieve better sensitivity (68% and 49%) than the other models (between 5% and 51%), and very good specificity (over 94%). The performance, particularly the sensitivity, decreases as the level of increase in CU that is being tested for increases.

The results presented in these tables provide an informative qualitative assessment for the ability of the different models to capture ΔCU from limited prevalence data on an important and diverse set of likely scenarios (100 experiments). First of all, MSE and AUC figures suggest that although the number of prevalence observations is low and some elements of the transmission process are uncertain, it is still possible to extract information on the time-varying parameter and provide estimates of the amplitude of the shift in CU

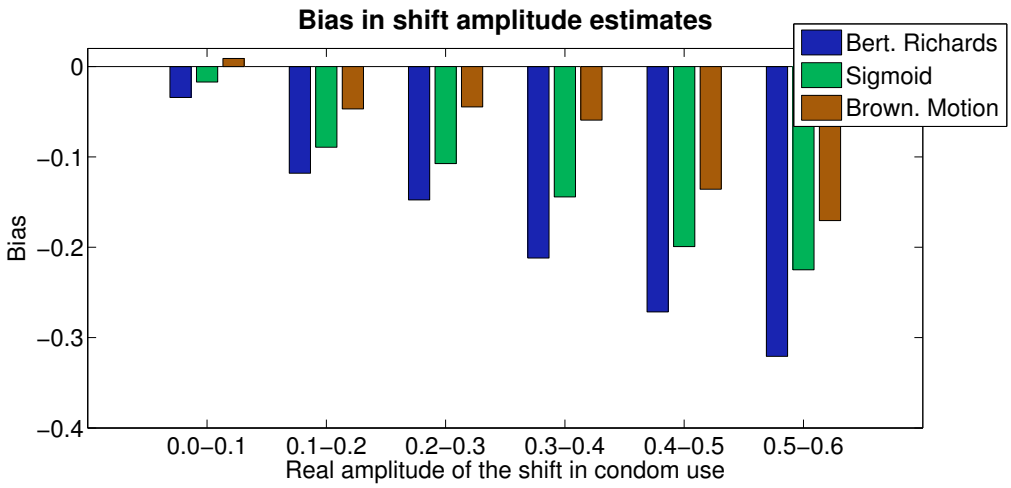

Figure 3.4: Bias of each model as a function of the true amplitude of the shift in condom use, estimated from 100 simulations.

during the intervention. Furthermore, there seems to be a possibility to control the risk of overstating these quantities by analysing the outputs of the three models that offer different levels of compromise between sensitivity and specificity. Thus, although the procedure may fail to identify some shifts in CU, we have provided some evidence suggesting that if a shift is detected it is likely to be true. The latter, in the context of interest, results in conservative estimates of intervention impact on CU trends.

The bias in estimating ΔCU under each of the CU trajectory priors can be attributed to a large extent to the prior implied by each formulation on ΔCU. As mentioned in Section 3.2.3 the BM approach results in a symmetric prior on ΔCU that is centered around 0 with 2.5% and 97.5% points at ± 0.6 respectively. The posterior median is therefore pulled towards 0 resulting in conservative estimates. The amount of shrinkage depends on the upper limit of the Uniform prior on σ. The corresponding priors under the dBR and dSigm formulations result in priors for ΔCU that put more mass around 0, although this heavily depends on the values of r and m that are hard to estimate. The resulting biases are therefore higher but they have been obtained without placing informative priors on their hyperparameters, as was done with σ under the BM formulation.

The two models with the higher overall performance, BM and dSigm, are quite different in nature: the dSigm trajectories are smooth, whereas under the Brownian motion prior they are non-differentiable. Hence, the choice between the two models can also be based on prior beliefs of the researcher regarding the smoothness of the CU trajectories.

	Deterministic Bertalanffy-Richards	Deterministic empirical sigmoid	Brownian motion
Bias	-0.23	-0.17	-0.13
Error standard deviation	0.16	0.17	0.17
Mean Squarred Error (MSE)	0.078	0.0057	0.045

Table 3.2: Frequentist properties of the different estimators of the amplitude of the shift in condom use during the intervention, estimated from 100 simulations

		Deterministic Bertalanffy-Richards	Deterministic Empirical Sigmoid	Brownian motion
$\Delta CU > 0.2$?	AUC	0.91	0.9	0.9
	Sensitivity	46%	51%	68%
	Specificity	100%	100%	96%
$\Delta CU > 0.4$?	AUC	0.85	0.83	0.82
	Sensitivity	5%	38%	49%
	Specificity	100%	95%	94%

Table 3.3: General distinctive power (AUC) of the median estimator of the shift, and specific sensitivity and specificity when answering: is the shift in CU during the intervention stronger than 0.2? than 0.4? These quantities were estimated over 100 simulations.

3.4.2 Application: what can be inferred on the trajectory of CU in Mysore from limited HIV prevalence data?

Mysore is one of the districts targeted by the Avahan intervention, and Avahan was the first HIV prevention intervention in this region. Four HIV prevalence estimates have been obtained between 2003 and 2009: three among female sex workers, and one among clients. Results from the inference procedure using a Brownian motion model are shown in Figure 3.5, suggesting a strong impact of the intervention. The purpose of this Chapter was to assess what level of increase of CU between 2003 and 2009 can be inferred while controlling the risk of overstating it. As it was shown in section 3.4.1, dSigm models could provide a good alternative to the BM formulation. Hence, we also present here results obtained with this model for the Mysore dataset (see figure 3.5, bottom panel). Table 3.4 shows the estimates of ΔCU for each of the three presented trajectory priors. The results indicate a positive increase in all cases. In particular, for the BM and dSigm models the corresponding posterior means are 0.54 and 0.55 while the 95% credible intervals are $[0.04; 0.99]$ and $[0.14; 0.99]$ respectively.

		Posterior mean	Posterior median	95% credible interval
ΔCU	Deterministic Bertalanffy-Richards	0.30	0.28	$[0.11; 0.73]$
	Deterministic Sigmoid	0.53	0.54	$[0.14; 0.99]$
	Brownian motion	0.52	0.55	$[0.04; 0.99]$

Table 3.4: Estimates of the change in CU in Mysore between 2003 and 2009.

A stronger conclusion regarding a lower bound for the CU shift between 2003 and 2009 can be made by comparing the posteriors medians to the results of Table 3.3. If the ensemble of simulations is to be considered realistic, an argument in favour of a CU increase being at least 0.4 can be made. Since the posterior medians are more than 0.4 under both BM and dSigm models (0.54 and 0.55 respectively), Table 3.3 suggests that a statement for $\Delta CU > 0.4$ will be correct with probability given by the specificity of each model (94% BM and 95% dSigm). While being more informative than the credible intervals obtained directly from the posterior densities (over 0.04 and 0.14 respectively with BM and dSigm), these numbers are heavily dependent on the assumption that the simulations of Section 3.3 provided an adequate representation of the reality.

Finally, Figure 3.5 and Table 3.4 show that the results obtained from the deterministic Sigmoid and Brownian motion models strongly coincide: they suggest that CU was stable over the 1985-2003 period, remaining below 0.5, sharply increased between 2003 and 2007, and stabilised between 0.8 and 0.9.

3.5 Generalisation to ten districts targeted by Avahan, and integration of survey-based estimates

3.5.1 Impact estimators derived from limited prevalence data

The approach presented in the previous sections was repeated to analyse data from other districts of southern India targeted by the Avahan intervention. In each district, three or four HIV prevalence estimates have been obtained between 2003 and 2010. The prevalence estimates, along with observation dates, are presented in Table 3.5. The Table illustrates the variety of situations; for example, the level of prevalence among FSWs at the time of the first IBBA was as high as 33.9% in Belgaum and 37.3% in Yevatmal, and below 10% in Shimoga end Chennai. Yet, in all regions the estimated levels of prevalence have decreased between the first and last IBBA. In this section, we apply the methodology introduced earlier in the Chapter to estimate the impact of the intervention during the complete period of the intervention (from 2003 to 2010). Although most of the parameters of the HIV model

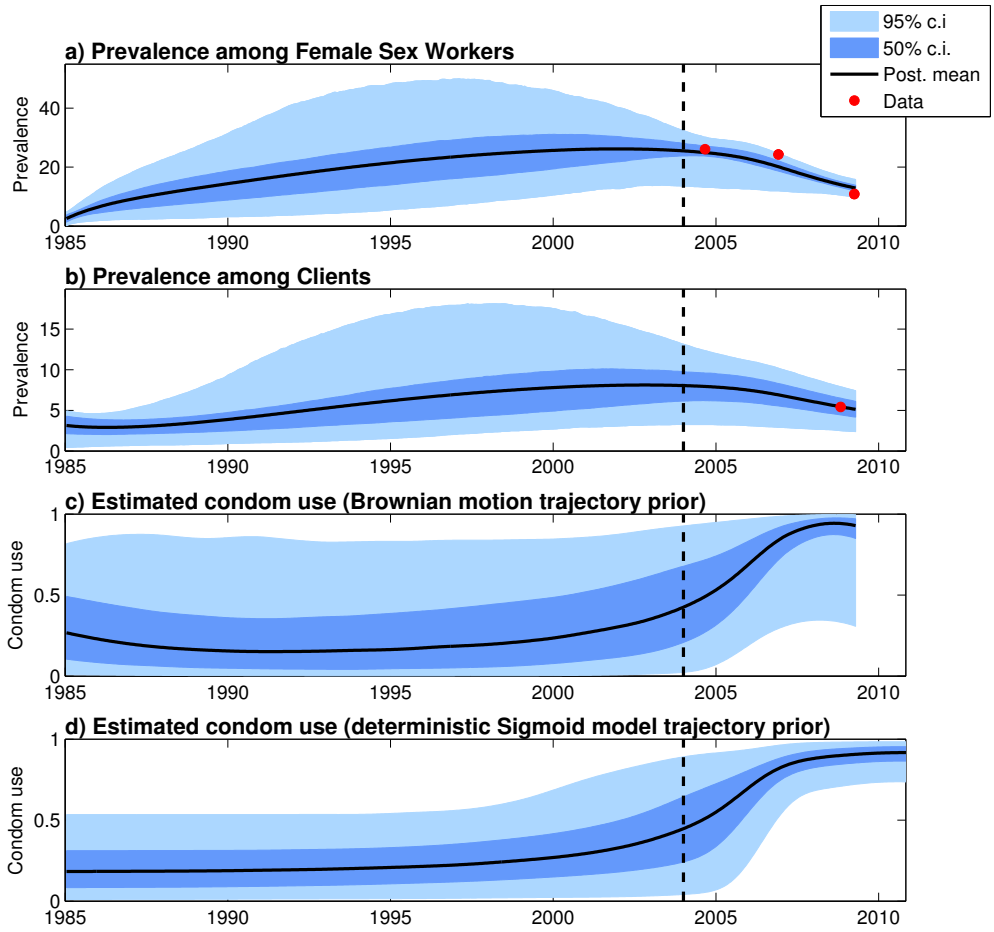

Figure 3.5: Estimates obtained for Mysore district.
a) reconstructed prevalence trajectory among female sex workers when condom use modelled with Brownian motion
b) reconstructed prevalence trajectory among clients when condom use modelled with Brownian motion
c) reconstructed condom use trajectory when modelled with Brownian motion
d) reconstructed condom use trajectory when modelled with deterministic Sigmoid

District	Date	Group	Estimate
Mysore	2004	FSWs	26.11%
	2006	FSWs	24.2%
	2008	Clients	5.4%
	2009	FSWs	11.1%
Belgaum	2005	FSWs	33.9%
	2007	Clients	6.2%
	2008	FSWs	27.3%
	2010	FSWs	22.3%
Bellary	2005	FSWs	15.6%
	2007	Clients	6.0%
	2008	FSWs	14.2%
	2010	FSWs	6.3%
East Godavari	2006	FSWs	26.3%
	2006	Clients	8.3%
	2009	FSWs	23.3%
	2009	Clients	9.6%
Guntur	2006	FSWs	21.3%
	2006	Clients	6.6%
	2009	FSWs	8.4%
	2009	Clients	7.1%

District	Date	Group	Estimate
Hyderabad	2006	FSWs	14.3%
	2006	Clients	2.4%
	2009	FSWs	9.6%
	2009	Clients	3.7%
Yevatmal	2006	FSWs	37.3%
	2006	Clients	10.9%
	2009	FSWs	26.7%
	2009	Clients	11.7%
Salem	2006	FSWs	12.9%
	2006	Clients	3.5%
	2009	FSWs	11.3%
	2009	Clients	1.9%
Shimoga	2005	FSWs	9.7%
	2007	Clients	2.3%
	2008	FSWs	8.9%
Chennai	2006	FSWs	3.2%
	2006	Clients	2.2%
	2009	FSWs	2.0%

Table 3.5: Prevalence estimates in each region with corresponding years.

represent biological quantities that are considered to be constant over the different districts, prostitution practices significantly vary between regions. Consequently, the prior densities imposed upon the mean number of clients per month per FSW for high and low risk groups is specifically adjusted in each district using IBBA survey data. Similarly, the number of encounters per client per month as well as the mean length of sexual activity as FSW or client are adapted to each district.

The analysis presented in the previous section shows that the BM and Sigm models allow similar levels of sensitivity and specificity. For reasons of clarity, we will only present here the estimates obtained with the BM model in the different districts. The resulting ΔCU estimates are summarised in table 3.6. If the results in table 3.3 were to be trusted, assuming that the ensemble simulations provided an adequate representation of reality, table 3.6 would suggest that the shift in CU in Mysore and Bellary was over 0.4 with probability 94%, and additionally over 0.2 in Belgaum and Guntur with probability 96%. By only relying on the 5% quantiles of the posterior densities of ΔCU, results suggest that the shift in CU is significantly positive in Mysore, Belgaum, Bellary and Guntur.

Region	Posterior median	95% credible interval
Mysore	0.55	[0.04; 0.99]
Belgaum	0.30	[0.00; 0.99]
Bellary	0.50	[0.01; 0.99]
East Godavari	0.02	[-0.29; 0.99]
Guntur	0.26	[0.00; 0.99]
Hyderabad	0.12	[-0.11; 0.99]
Yevatmal	0.19	[-0.22; 0.99]
Salem	0.02	[-0.18; 0.98]
Shimoga	-0.00	[-0.22; 0.93]
Chennai	0.00	[-0.24; 0.98]

Table 3.6: ΔCU indirectly estimated from prevalence data in each region, with associated credible intervals

3.5.2 Contrast of model outputs with survey-based condom use estimates

The results of Table 3.6 are based on prevalence estimates and the HIV transmission model in which some constant parameters are parameterised using the information obtained from cross-sectional surveys (IBBAs). As part of these surveys, FSWs have also been asked for an estimation of the proportion of their last commercial sex acts that had been protected with condoms. From these answers, an estimate of the shift in CU between the first and last IBBA among FSWs can be derived, that is provided for each district in Table 3.7. To allow direct comparison, the median estimator of the posterior density of the shift in CU between the last and first IBBA among FSWs is computed from the outputs of the indirect inference procedure. The estimates for each district are given in Table 3.7, allowing for the direct comparison of the information derived from direct surveys among FSWs, and the information indirectly inferred from prevalence estimates and the HIV transmission model. The difference between both estimates of CU progression between the first and last IBBA are shown in the last column: it is below 0.05 in every district but Mysore (+0.17), Bellary (+0.07), Salem (-0.07) and Shimoga (-0.15).

There is a significant dependence of this difference on the number of FSW prevalence estimates available in each district. This association provides potential explanations for the important discrepancies observed in four cases. In both Salem and Shimoga, only two estimates of prevalence among FSWs are available, which gives more weight to the prior that is centered around zero. Additionally, survey-based CU estimates at first IBBA are consistently significantly higher than model-based ones: under the levels of CU at first IBBA hypothesised by survey-based estimates, as strong progressions as suggested by the model are unlikely. For example, it is impossible to have an 0.46 increase in CU over 5 years

Region	CU estimate at first IBBA			CU progression between first and last IBBA		
	Survey-base	Model-based	Difference	Survey-base	Model-based	Difference
Mysore	0.65	0.42	+0.23	0.29	0.46	-0.17
Belgaum	0.91	0.77	+0.14	0.05	0.07	-0.02
Bellary	0.82	0.69	+0.13	0.08	0.15	-0.07
East Godavari	0.93	0.75	+0.18	0.05	0.00	+0.05
Guntur	0.96	0.98	-0.02	0.03	0.01	+0.02
Hyderabad	0.94	0.49	+0.45	0.02	0.02	+0.00
Yevatmal	0.98	0.37	+0.61	-0.00	0.04	-0.04
Salem	0.91	0.27	+0.64	0.07	0.00	+0.07
Shimoga	0.72	0.08	+0.64	0.15	0.00	+0.15
Chennai	0.96	0.85	+0.11	0.03	0.00	+0.03

Table 3.7: Survey-based estimate at first IBBA and estimated progression untill last IBBA. Model-based estimates of the corresponding quantities indirectly estimated from prevalence data.

in Mysore if its level in 2004 was 0.65. In addition, the results from ensemble simulations in the setting of Mysore indicate that model-based estimates of CU in 2004 (date of the first IBBA in Mysore) already tend to over-estimate reality by 0.05 (with standard deviation 0.2). Hence, this evidence confirms that survey answers may be consistently positively biased, which was already suspected due to social desirability. This bias would explain the discrepancies between model and survey estimates in Mysore and Bellary.

3.5.3 Bayesian synthesis and final estimates

In this final section, we derive estimates of the shift in CU during the period of the intervention by synthetising the information provided by prevalence estimates, the HIV transmission model and cross-sectional behavioural surveys including survey-based condom use estimates. Nonetheless, due to the a priori suspected risk of over-estimation in survey-based estimates of condom use, and to the consistently higher estimates provided by surveys when compared to model-based estimates, we incorporate in the model the risk for survey answers to be overestimated by up to 25%. The observation function for survey-based estimates of CU is the following:

$$y \sim Bin(n, \rho CU_t), \tag{3.13}$$

Here, n corresponds to the number of FSWs that have answered the survey (generally around 400), CU_t is the true value of condom use at time t, and ρ is the overestimation parameter which prior is set to be uniform between 1 and 1.25. The resulting estimates of CU

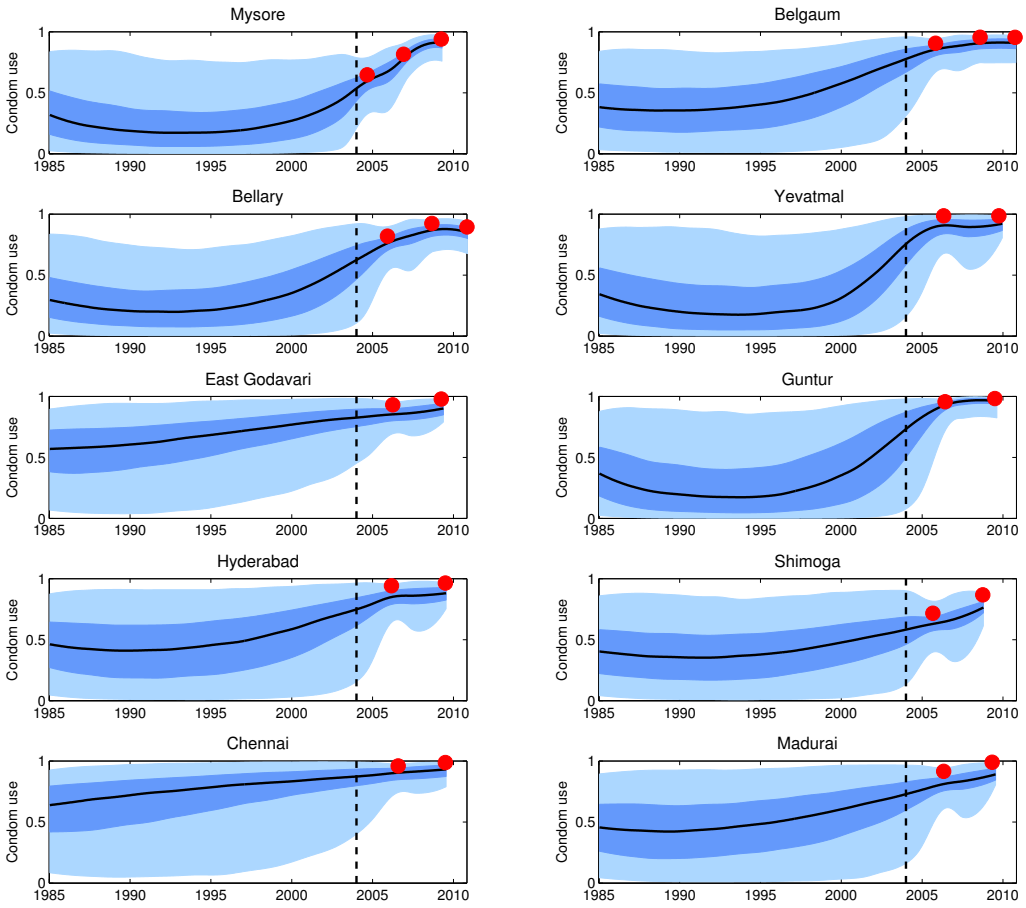

Figure 3.6: Condom use estimated trajectories resulting from the Bayesian synthesis of prevalence estimates and cross-sectional behavioural surveys including survey-based condom use estimates (red dots)

are presented in Table 3.8, with associated trajectories provided in Figure 3.6. The incorporation of survey-based condom use estimates allows for narrower posterior estimates. In particular, the lower bound of the 95% credible interval in Yevatmal is brought over zero, providing evidence for an increase of condom use over the period of the intervention.

Region	Posterior median	95% credible interval
Mysore	0.50	[0.15; 0.96]
Belgaum	0.19	[0.00; 0.94]
Bellary	0.36	[0.02; 0.97]
Yevatmal	0.33	[0.00; 0.97]
Shimoga	0.22	[-0.08; 0.87]
Guntur	0.39	[0.03; 0.99]
Hyderabad	0.18	[-0.04; 0.96]
East Godavari	0.07	[-0.05; 0.94]
Chennai	0.05	[-0.06; 0.95]
Salem	0.20	[-0.24; 0.98]

Table 3.8: ΔCU estimates in each region with associated credible intervals resulting from the Bayesian synthesis of prevalence data and biased survey estimates

3.6 Discussion

In this Chapter, we presented a Bayesian approach to draw conclusions regarding the evolution of time-varying behavioural parameters in the context of HIV such as condom use among female sex workers. Inference can be based on prevalence estimates while a substantial amount of information from additional sources can be incorporated via prior distributions. In order to describe CU trajectories we introduced three different formulations based on Brownian motion and growth curves such as the generalised Bertalanffy Richards and empirical sigmoid models. To our knowledge, these formulations are new in this context. The presented computational framework allows estimation of CU trajectories as well as functionals thereof, using the methodology introduced in Chapter 2. Nevertheless, in comparison to the situations considered in Chapter 2, the problem of evaluating the Avahan intervention by estimating its impact on CU from prevalence estimates is of additional difficulty due to the limited amount of information; the application to Mysore district was based on three observations of prevalence among FSWs and one among clients, plus hypothesis on the initial value of prevalence in 1985. A series of simulation experiments were conducted in order to assess the validity of the procedure, examining the frequentist properties of the underlying estimators and the ability of the model to avoid overestimation via conducting ROC analysis. The evidence from the simulation experiments is encouraging, suggesting that the approach can be used in this context for making conservative estimates of changes in CU both with the Brownian motion and the deterministic sigmoid trajectory priors. However, the overall performance is bound to depend on the deterministic HIV infection model which was parameterised based on a substantial amount of prior information, as in Pickles et al. (2010), as well as on assumptions such as the very low HIV

prevalence in 1985. Most of the prior information utilised in this study was obtained from additional data sources (IBBAs).

While the representation of HIV transmissions in this Chapter is simpler in behavioural terms in comparison with the model presented in Pickles et al. (2010), the model is enriched as it explores the CU trajectories space rather than working with three pre-determined scenarios. Nevertheless, there are reasons for a potential overestimation of the shift amplitude in this simpler model as coinfection with other sexually transmitted diseases were ignored (although higher transmission probability per unprotected act were allowed to compensate for the latter), and no acute phase was considered. However, diffusion driven models aim at capturing and compensating for structural mis-specifications while capturing the main dynamics of the system and have been shown here to provide conservative estimates. Overall it may be viewed as a different and complementary choice in the trade-off between richness and tractability of the model compared to Pickles et al. (2010). Lastly, this approach relies on the hypothesis that changes in transmission probabilities are solely related to changes in CU, ignoring for example potential changes in the frequency of commercial sex partnerships. This choice can be motivated by the strong focus of the Avahan intervention on prevention measures and the relative stability in the frequency of commercial sex exhibited by the series of cross-sectional bio-behavioural surveys that were conducted during the period of the intervention.

CHAPTER 4

Bayesian inference with the advanced Hybrid Monte Carlo algorithm

4.1 Introduction

Chapters 2 and 3 have motivated and illustrated the use of the particle marginal Metropolis Hastings version of the PMCMC algorithm to explore the joint posterior probability density $p(x_{0:n}^{dis}, \theta | y_{0:n})$ in the context of Markovian indirectly observed stochastic processes. Building up on the principles of the PMCMC algorithms, the authors of Chopin et al. (2012) have defined an alternative pseudo-marginal approach: the SMC^2 algorithm. This algorithm allows for an automatic calibration of the number of particles and a progressive adaptation to the structure of the posterior density as observations accumulate. From this perspective it is an interesting alternative to the PMCMC to achieve full Bayesian inference on indirectly observed stochastic processes. Nevertheless, both algorithms require the propagation of J particles for each value of θ. Their computational complexity is consequently of the same order of magnitude: $O(nJN_\theta)$. We explore in this third Chapter the use of the advanced version of the Hybrid Monte Carlo algorithm (HMC), targeted to Hilbert spaces to infer the paths ($x_{0:n}^{dis}$) and parameters (θ) of indirectly observed stochastic processes (Beskos et al., 2011).

Typically, the efficiency of MCMC algorithms deteriorates as the dimension of the target density increases. For example, the authors of Roberts et al. (1997) have shown that the scaling of the importance sampling distribution of a random walk Metropolis algorithm, in cases where the target is a multivariate normal density of dimension d, should decrease linearly in d to ensure convergence. As a consequence, for high values of d the chain can only make small local moves and generate strongly correlated samples. In such cases, the exploration of probability densities becomes computationally intractable as the

dimension of the problem increases. The Hybrid Monte Carlo algorithm provides a solution to efficiently explore complex and high-dimensional target densities (Girolami and Calderhead, 2011). In particular, the advanced version of the HMC algorithm defined in Beskos et al. (2011) allows for large, gradient-driven moves in Hilbert spaces. These algorithms will be presented in Section 4.2. To our knowledge, the work presented in this Chapter is the first attempt to jointly estimate the paths and parameters of a stochastic process with an advanced Hybrid Monte Carlo algorithm: the Hilbert space of trajectory paths is augmented with the parameter space, allowing for joint and simultaneous updates of $x_{0:n}^{dis}$ and θ. As such, the advanced HMC can be seen as an alternative to the pseudo-marginal approaches. Differences between the particle Marginal Metropolis Hastings algorithm utilised in the first Chapters of this thesis and the advanced HMC algorithm mainly lie in the fact that the PMMH can be seen as Markov Chain designed in a small-dimensional space, based on the costly estimation of $\hat{p}_{pf}(y_{1:n}|\theta)$, while the advanced HMC operates in the high-dimensional joint space of $x_{0:n}^{dis}$ and θ based on the direct and exact estimation of the augmented likelihood $\hat{p}_{pf}(y_{1:n}|x_{0:n}^{dis}, \theta)$. One of the first quantitative comparisons between these two algorithms in the context of stochastic volatility models will be presented in Section 4.3 of this Chapter.

Furthermore, we explore the ability of the advanced HMC algorithm to deal with non-Markovian stochastic processes. We consider the possibility for a stochastic volatility model to be driven by a fractional Brownian motion whose increments are correlated in time, which is motivated by recent studies that have questioned the memoryless nature of the volatility of financial time-series (Chronopoulou and Viens, 2012). Difficulties arise when attempting to achieve full Bayesian inference for non-Markovian stochastic processes as particle filters cannot be directly used to explore the filtering density $p(x_{0:n}^{dis}|y_{1:n}, \theta)$. In this Chapter, we adapt the advanced HMC algorithm to the fractional stochastic volatility model, and illustrate with simulations how the level of memory of the system, characterised by the *Hurst* parameter H, can be estimated within a few hours from the equivalent of a year-long time series of daily market prices.

This Chapter presents work being conducted in collaboration with Kostas Kalogeropoulos (London School of Economics) and Alex Beskos (University College London).

4.2 An efficient MCMC sampler

In the absence of tractable formulas for the forward Kolmogorov equations, the dynamic of the system is approximated by discretising time over periods of length δ. Under suitable conditions, the probability density over the trajectories of the discretised system converges in distribution to the one of the original continuous-time system as δ tends to 0. The prin-

cipal problem addressed by the advanced HMC algorithm is the robustness of the mixing properties of inference algorithms as δ becomes increasingly small.

4.2.1 A priori decoupling $x_{0:n}$ and θ

Our general definition of indirectly observed stochastic processes allows for the dependency of the density of the underlying stochastic process $x_{0:n}$ on components of the parameter θ. In some cases, a sample trajectory of the stochastic process $x_{0:n}$ contains infinite information on its driving parameters. For example, the posterior probability density $p(\sigma|\{x_0, x_\delta, x_{2\delta}, \ldots, x_N\})$ of the volatility of a discretised trajectory of a simple Brownian motion $dx_t = \sigma dB_t$ tends to a Dirac distribution as δ tends to 0. When possible, it is profitable to change the formulation of the model so as to a priori decouple the paths $x_{0:n}$ and parameter vector θ. This reparameterisation is particularly necessary when the joint space of $x_{0:n}$ and θ is explored with a joint scheme that updates the paths and parameters of the system alternatively.

In the simple case of a uni-dimensional Brownian motion driven by a constant volatility, for example, an auxiliary variable $\dot{x}_{0:n} = \sigma^{-1}x_{0:n}$ can be defined which has a priori unit volatility: the prior distribution of $\dot{x}_{0:n}$ is independent of the parameters of the model. In the general case, ad hoc solutions need to be found that have been explored in Golightly and Wilkinson (2008) and Kalogeropoulos et al. (2011).

4.2.2 Classic version of the Hybrid Monte Carlo algorithm

The HMC algorithm was first introduced in Duane et al. (1987), and originally called Hamiltonian Monte Carlo algorithm as a reference to the physical interpretation of its principles. The rationale for the HMC algorithm is an analogy with classic dynamics, where probability $\pi(q)$ is assimilated to energy $E(q)$ through the formula $\pi(q) = \frac{1}{Z}e^{-E(q)/T}$ (Neal, 2010), where q is the variable of interest, T the temperature (generally fixed to 1), and Z the normalising constant. The current position of the chain, q, is augmented with a momentum vector p, thus defining a total energy H:

$$H(q,p) = -\log \pi(q,p) = -\log \pi(q) + \frac{1}{2}p^t p \tag{4.1}$$

The vector p can be seen as the initial momentum of a point object located in q, moving without friction along a hilly surface where the height of the surface at a given point is the opposite of its potential energy $\phi(q) = -\log \pi(q)$, and the kinetic energy is $\frac{1}{2}p^t p$. The trajectory of the object under these hypotheses can be obtained by integrating the energy conservation equations resulting into the system of Hamiltonian equations 4.2. To avoid confusion, the time variable introduced in the Hamiltonian equations will be denoted with

a subscript H. A consequence of the conservation of energy is that when the trajectory of q approaches regions with higher potential, the kinetic energy decreases. On the contrary, the speed increases as q moves towards regions of lower potential:

$$\begin{cases} \dfrac{dq}{dt_H} = \dfrac{\partial H}{\partial p} \\ \dfrac{dp}{dt_H} = -\dfrac{\partial H}{\partial q} \end{cases} \tag{4.2}$$

The trajectory resulting from this physical analogy is used to define a mapping $\psi : \{q_i, p_i\} \to \{q_{i+1}, p_{i+1}\}$ in the joint $\{q, p\}$ space. From this mapping, a theoretical two-step MCMC algorithm can be constructed. First, a new value of the momentum p is sampled from the density imposed upon p, which is classically a standard multivariate normal density. Then, the mapping generates a new couple of values for the position and momentum, of equal energy hence equal probability. It can be shown that this scheme respects the detailed balance condition, and thus converges to the target density π (Neal, 2010). Nevertheless, the system of Hamiltonian equations can only be solved by discretising the time t_H. The number of integration steps conducted, leading to a proposed couple $\{x^*, p^*\}$, is denoted L, and their size is noted h. The properties of the continuous-time mapping ψ are not necessarily preserved by its discretised version. For example, the following Euler integration scheme loses the volume-preservation property of the Hamiltonian:

Euler Hamiltonian mapping
Set $q^* = q$ and $p^* = p$
for $i = 1$ to $L - 1$ **do**
  Set $q^* = q^* + hp^*$ and $p^* = p^* - h\nabla\phi(q^*)$
end for

Importantly, the integration scheme does not guaranty the conservation of the total energy. However, this issue can be addressed by a Metropolis-Hastings step based on the ratio $\frac{\pi(q^*,p^*)}{\pi(q,p)} = \exp[-H(q^*, p^*)+H(q, p)]$ (see algorithm 11). When the number of leapfrogs is one, the HMC is equivalent to the Metropolis-Adjusted Langevin Algorithm (MALA), introduced in Roberts and Stramer (2002) and mentioned in the Introduction of this thesis. The leapfrog integration method is generally preferred over the Euler scheme as it allows for an exact conservation of volume:

Leapfrog Hamiltonian mapping
Set $q^* = q$ and $p^* = p - \frac{h}{2}\nabla\phi(q)$

for $i = 1$ to $L - 1$ **do**
  Set $q^* = q^* + hp^*$ and $p^* = p^* - h\nabla\phi(q^*)$
end for
Set $q^* = q^* + hp^*$ and $p^* = p^* - \frac{h}{2}\nabla\phi(q^*)$

Algorithm 11 HMC algorithm

Initialise $q^(0)$
for $i = 0$ to N **do**
  Sample $p \sim \mathcal{N}(0, Id)$
  Apply a Hamiltonian mapping $\psi_{h,L} : \{q, p\} \rightarrow \{q^*, p^*\}$
  Accept q^* with probability $1 \wedge \frac{\pi(q^*)}{\pi(q^{(i)})} \exp(-\frac{1}{2}(p^t p - p^{*t} p^*))$
end for

In particular, the integration error induced by the leapfrog scheme is of order h^3, while the rate of convergence of the Euler scheme is one order of magnitude lower (Neal, 2010). The exploration of a target density with the classic HMC algorithm is expected to require far less iterations than when using a random walk Metropolis scheme. To begin with the optimal acceptance rate, that can be monitored through the number of leapfrogs L or by their size h, is considered to be around 80%, whereas the optimal acceptance rate of a random walk exploration of the target is 23% (Roberts et al., 1997). Most importantly, the moves proposed by the HMC algorithm will be less local, which can significantly diminish the correlation between successively generated samples. However, the authors of Beskos et al. (2011) have shown that the size h of the jumps need to decrease as $O(d^{-1/4})$, with d being the dimension of the target density. For this reason, the mixing performance of the classic HMC will not be robust to the increase of the dimension of $x_{0:n}^{dis}$ as the discrete approximation of the modelled dynamic is refined.

4.2.3 Advanced Hybrid Monte Carlo algorithm

The authors of Beskos et al. (2011) have proposed a discretisation scheme for the Hamiltonian equations that is well-defined in the Hilbert space of the continuous path $x_{0:n}$. It is based on the decomposition of the Hamiltonian equations into two systems:

$$\begin{cases} \dfrac{dq}{dt_H} = 0, \quad \dfrac{dp}{dt_H} = -\nabla\phi(q) \\ \dfrac{dq}{dt_H} = p, \quad \dfrac{dp}{dt_H} = -q \end{cases} \tag{4.3}$$

The solution operators of 4.3 can be respectively defined as Ξ_t and $\tilde{\Xi}_t$:

$$\begin{aligned}\Xi_{t_H}(q,p) &= (q, p - t\nabla\phi(q)) \\ \tilde{\Xi}_{t_H}(q,p) &= (\cos(t_H)q + \sin(t_H)p, -\sin(t_H)q + \cos(t_H)p)\end{aligned} \tag{4.4}$$

By defining the integration time step h such as $\cos(t_H) = \frac{1-h^2/4}{1+h^2/4}$, an elementary integration leapfrog can be alternatively expressed as:

$$\begin{aligned}p_{h/2} &= p_0 - \frac{h}{2}\frac{q_0 + q_h}{2} - \frac{h}{2}\nabla\phi(q_0) \\ q_h &= q_0 + hp_{h/2} \\ p_h &= p_{h/2} - \frac{h}{2}\frac{q_0 + q_h}{2} - \frac{h}{2}\nabla\phi(q_h)\end{aligned} \tag{4.5}$$

From this novel mapping, the advanced HMC algorithm can be derived on the same principle as in its classic version (Algorithm 11). The mathematical proofs for the validity of the algorithm in the Hilbert space can be found in Beskos et al. (2011) and Beskos et al. (2012). It has been applied to the estimation of diffusion paths, and positively compared to the advanced MALA algorithm introduced in Beskos et al. (2008). We propose to extend its use to the joint estimation of the path and parameters of a potentially non-Markovian stochastic process, and compare its sampling performance and computational burden to the PMMH algorithm in the Markovian case. In this perspective, the target density is augmented to encompass the parameter vector θ and its prior density:

$$\phi(q) = \phi(x_{0:n}, \theta) = -\log p(y_{0:n}|x_{0:n}, \theta) - p(\theta) \tag{4.6}$$

And the Hamiltonian equations are decomposed in the following manner:

$$\begin{aligned}\frac{dq}{dt_H} &= 0, \quad \frac{dp}{dt_H} = -\nabla\phi(x_{0:n}, 0) \\ \frac{dq}{dt_H} &= p, \quad \frac{dp}{dt_H} = -(x_{0:n}, 0)^t\end{aligned} \tag{4.7}$$

4.2.4 Scaling mass matrix

Both the classic and advanced versions of the HMC algorithm can be further modified by introducing a scaling matrix M reflecting the shape of the target density and potential correlations between the components of the position vector q. For illustration, we provide with the modified classic Leapfrop Hamiltonian mapping and corresponding HMC algorithm (Algorithm 12):

Leapfrog Hamiltonian mapping

Set $q^* = q$ and $p^* = p + \frac{h}{2}M^{-1}\nabla\pi(q)$

for $i = 1$ to $L - 1$ **do**
  Set $q^* = q^* + hp^*$ and $p^* = p^* + hM^{-1}\nabla\pi(q^*)$
end for
Set $q^* = q^* + hp^*$ and $p^* = p^* + \frac{h}{2}M^{-1}\nabla\pi(q^*)$

Algorithm 12 HMC algorithm with scaling mass matrix M

Initialise $q^(0)$
for $i = 0$ to N **do**
  Sample $p \sim \mathcal{N}(0, M^{-1})$
  Apply a Hamiltonian mapping $\psi_{h,L} : \{q, p\} \to \{q^*, p^*\}$
  Accept q^* with probability $1 \wedge \frac{\pi(q^*)}{\pi(q^{(i)})} \exp(-\frac{1}{2}(p^t M p - p^{*t} M p^*))$
end for

4.3 Application: full Bayesian inference for a fractional stochastic volatility model

4.3.1 A long-memory stochastic volatility model

We begin by motivating the use of stochastic volatility models, and the central role they play in the definition and valuation of financial products called derivatives. The value of derivative products depends on the evolution of the price of one or several auxiliary goods that are referred to as *underlying assets*. They are materialised by a contract between two parties: one of the most classic types of derivatives are options, that provide the buyer the right (but not the obligation) to buy or sell an underlying asset on a given date at a price specified by the initial contract (the strike price). In the cases where the option provides a right to buy at maturity (European Call option), the buyer will naturally be willing to exercise the option at a time t if the price of the underlying asset P_t at this instant is higher than the strike price P_K. If there was no uncertainty in the evolution of the price of the underlying asset, the associated derivative products would have no value: the probability for the asset price P_t to ever go beyond P_K is either 0 or 1. In both cases the buyer or the seller has no interest in committing to the contract. Following this reasoning, it appears clearly that the higher the uncertainty regarding the evolution of P_t, the higher the probability for it to exceed the strike price P_K and generate a benefit to the buyer. Hence, if the dynamic of the asset price is modelled with a stochastic differential equation (Equation 4.8), where the drift function is usually determined by the risk-free interest rate, the value of a derivative product is generally an increasing function of the volatility σ of the underlying price, aside with a series of other factors.

$$dP_t = \mu(P_t, \theta)dt + \sigma(P_t, \theta)dB_t \tag{4.8}$$

Consequently, it is important for both parties to have a sound estimation of this volatility, in order to determine the price they are willing to pay for the derivative product. It is also a mean to estimate the risk that is induced by the contract: given this estimation, each party takes positions to hedge against this risk (cancel or reduce the risk of loss). A first solution to this problem was provided by Fischer Black and Myron Scholes in 1973, through the derivation of a differential equation that governs the price of an option over time (Black and Scholes, 1973). This equation is meant to be used to quantify the risk taken by both parties at each instant, and theoretically allows them to perfectly hedge against it. The impact of this model has been considerable, and it is still an ubiquitous tool on trading floors. Nevertheless, the Black & Scholes model relies on assumptions that have been long questioned. Importantly, they imply that the volatility of the underlying price of an option is constant, which has been constantly negated by observations. The latter has motivated extensions to the classic formulation: stochastic volatility models are an example of these extensions, in which the volatility v_t itself also follows a differential equation:

$$\begin{cases} dP_t = \mu_P(P_t, vt, \theta)dt + \sigma_P(v_t, \theta)dB_t^{(1)} \\ dv_t = \mu_v(v_t, \theta)dt + \sigma_v(v_t, \theta)dB_t^{(2)} \end{cases} \tag{4.9}$$

Stochastic volatility models are a canonic example of indirectly observed stochastic processes. Inference of the driving parameter vector θ has been the subject of active research, posing similar problems to the ones encountered with population-level compartmental epidemic models. Hence, efficient and asymptotically exact inference has been made possible by the recent development of the PMCMC and SMC^2 algorithms. However, empirical evidence shows that the independence of the stochastic increments driving the dynamic of the volatility is a questionable hypothesis. It has been observed that the autocorrelation function of the squared high frequency returns often only slowly decays to zero, suggesting a long memory for the dynamic of the stochastic volatility (Ding et al., 1993; Lobato and Savin, 1998; Chronopoulou and Viens, 2012). The authors of Comte and Renault (1998) have introduced the use of a fractional Brownian motion in a stochastic volatility model to account for long memory of the process. As in Chronopoulou and Viens (2012), we model log-returns ($u_t = \log P_t$) as a mean-reverting Ornstein-Uhlenbeck process coupled to a Scott volatility model ($\sigma(v_t) = \exp(v_t/2)$) driven by a fractional ex-

tension dB_t^H of the classic Brownian motion motion:

$$\begin{cases} du_t &= (\mu_u - \sigma_v(v_t)^2/2)dt + \sigma_u(v_t)dB_t \\ dv_t &= \kappa(\mu_v - v_t)dt + \sigma_v dB_t^H. \end{cases} \tag{4.10}$$

A fractional Brownian motion (fBM) is a centered Gaussian process $\{B_t^H, t \in \mathbb{R}_+\}$ which paths are continuous with probability 1, that is characterised by its Hurst parameter $H \in (0, 1]$ and its covariance:

$$Cov(B_t^H, B_s^H) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}) \tag{4.11}$$

The classic Brownian motion corresponds to the case $H = \frac{1}{2}$. For higher values, increments of the Brownian motion are positively correlated. For lower values, they are negatively correlated. Figure 4.2 provides with three realisations of the fractional Brownian motion for different values of the Hurst parameter H. This framework allows for the incorporation of long memory in the dynamic of the volatility, with a parameter H that characterises and quantifies the type and level of memory involved. The calibration of this type of models, however, remains an open challenge from a statistical methodology perspective due to the lack of Markovianity of the implied stochastic volatility model. In this case, the particle filter cannot be directly applied to sequentially explore the filtering density $p(x_{0:n}|y_{1:n}, \theta)$. We propose here to apply the advanced Hybrid Monte Carlo Algorithm to achieve efficient and asymptotically exact inference on this example of non-Markovian indirectly observed stochastic process. As in previous examples, in practice the algorithm will consider the driving noise B_t^H on a regular grid of N points of length $\delta = T/N$. The increments are defined as:

$$g_k = B_k^H - B_{k-1}^H, \quad 1 \leq k \leq N \tag{4.12}$$

These increments determine the N-dimensional fractional Gaussian noise vector g. The direct way to generate realisations of the fBM is to multiply a vector of N i.i.d. realisations of $\mathcal{N}(0, 1)$ and multiply it by the Cholesky decomposition of Γ, the covariance matrix of g:

$$\Gamma := \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(N-2) & \gamma(N-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(N-1) & \gamma(N-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix} \tag{4.13}$$

Here, $\gamma(k) = \mathrm{E}\,[\, g_{k_0+k}\, g_{k_0}\,] = \frac{\delta^{2H}}{2}|k+1|^{2H} + \frac{\delta^{2H}}{2}|k-1|^{2H} - \delta^{2H}|k|^{2H}$ for $k \geq 1$, and $\gamma(k) = \delta^{2H}$ if $k = 0$. This process allows to a priori decouple the path of the noise driving the

system from the parameters of the model, following Subsection 4.2.1. However, calculating the Cholesky decomposition of Γ for each new value of H requires $O(N^2)$ operations. We rely on the Davies and Harte method for sampling g, presented in Wood and Chan (1994), to implement an $O(N \log N)$ implementation of the advanced HMC algorithm for the fractional stochastic volatility model.

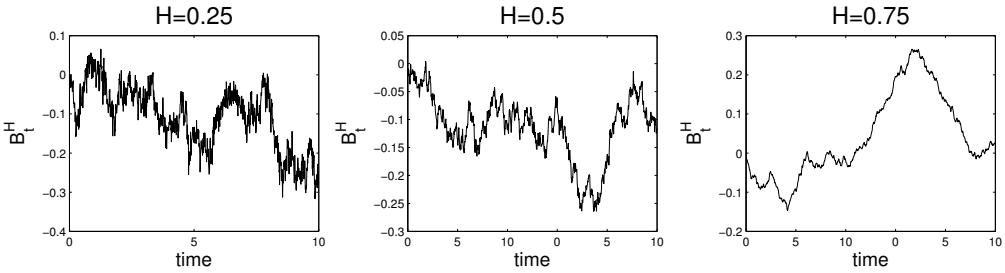

Figure 4.1: Realisations of the fractional brownian motion, with volatility 0.08 and respective Hurst parameters 0.25, 0.5 and 0.75

4.3.2 $O(N \log N)$ implementation of the advanced HMC on a fractional stochastic volatility model

The Davies and Harte method for sampling g relies on the *Toeplitz* structure of Γ: each descending diagonal from left to right is constant. By introducing the auxiliary matrix

$$\Gamma^f = \begin{pmatrix} 0 & \gamma(N-1) & \cdots & \gamma(2) & \gamma(1) \\ \gamma(N-1) & 0 & \cdots & \gamma(3) & \gamma(2) \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(1) & \gamma(2) & \cdots & \gamma(N-1) & 0 \end{pmatrix}, \tag{4.14}$$

a circulant covariance matrix can be constructed which rows can be obtained by shifting the previous one by one index to the right:

$$C = \begin{pmatrix} \Gamma & \Gamma^f \\ \Gamma^f & \Gamma \end{pmatrix}. \tag{4.15}$$

The circulant property allows for a fast eigen-expansion based on the fast Fourier transform methodology, providing the following decomposition:

$$C = P \Lambda_H P^\star \tag{4.16}$$

Where P is the $2N \times 2N$ transformation matrix with constituent elements

$$P_{jk} = \tfrac{1}{\sqrt{2N}} \exp\big(-2\pi i \tfrac{jk}{2N}\big) \tag{4.17}$$

The matrix P corresponds to a projection from the natural to the frequency domain, and P^* is its complex transform. The diagonal elements of $\Lambda_H = diag\{\lambda_0, \lambda_1, .., \lambda_{2N-1}\}$ are determined by the following, where $(c_{0,j})$ denotes the elements of the first row of matrix C:

$$\lambda_k = \sum_{j=0}^{2N-1} c_{0,j} \exp\big(-2\pi i \tfrac{jk}{2N}\big) . \tag{4.18}$$

Using the fast Fourier transform, these elements can be obtained in $O(N \log N)$ operations (Cooley and Tukey, 1965). With this decomposition, the square root of C can be extracted in $O(N)$ operations:

$$C^{1/2} = P \Lambda_H^{1/2} P^\star . \tag{4.19}$$

Furthermore, the $O(N \log N)$ cost of the multiplication of a vector of N i.i.d. realisations of $\mathcal{N}(0,1)$ by $C^{1/2}$ can be further reduced to a linear cost, as suggested in Wood and Chan (1994), by sampling a vector $x \sim \mathcal{N}(0, I_{2N})$ and multiplying it by the sparse matrix Q constructed in the following way:

- $Q^{11} = \text{diag}\{1, 1/\sqrt{2}, 1/\sqrt{2}, \ldots, 1/\sqrt{2}\}$
- $Q^{12} = \text{diag}\{0, i/\sqrt{2}, i/\sqrt{2}, \ldots, i/\sqrt{2}\}$
- $Q^{21} : Q^{21}_{N-i+1,i} = 1/\sqrt{2}$ for $2 \le i \le N-1$
- $Q^{22} : Q^{22}_{N-i+1,ii} = i/\sqrt{2}$ for $2 \le i \le N-1$; $Q^{22}_{0,0} = 1$

The sampling process of fractional Brownian motion increments $g_{0:n}$ is summarised in algorithm 13.

Algorithm 13 Simulation of a fractional Brownian motion realisation $g_{0:n}$

Sample $x_{0:2N} \sim \mathcal{N}(0, I_{2N})$ by $C^{1/2}$
Calculate $x'_{0:2N} = \Lambda^{1/2} Q x_{0:2N}$. [cost $\mathcal{O}(N)$]
Calculate $x''_{0:2N} = P x'_{0:2N}$ and return $g_j = x''_{0:N}$. [cost $\mathcal{O}(N \log N)$]

Additionally, the implementation of the advanced HMC algorithm requires the calculation of gradients $\nabla\phi(q) = \nabla\phi(x_{0:2N}, \theta)$. We will now focus on the corresponding formulas, and specify how they can be calculated while preserving the $O(N \log N)$ complexity of the algorithm. The available observations provide the values of the log-price u_t of the underlying asset at a finite set of points:

$$y_{1:n} = \{u_{t_k} : k = 0, 1, \ldots, n\}.$$

Conditional on the path $v_{0:n}$, u_t satisfies the Markov property. Hence, without loss of generality, we focus on the pair of observations (y_{k-1}, y_k). The model defined by the set of equations 4.10 determines the distribution of y_k given y_{k-1} and the path $v_{k-1:k}$:

$$y_k|y_{k-1}, v_{k-1:k} \sim \mathcal{N}\left\{y_{k-1} + \int_{t_{k-1}}^{t_k} (\mu_u - \sigma_u(v_s)^2/2)ds,\ \int_{t_{k-1}}^{t_k} \sigma_u(v_s)^2 ds\right\} \tag{4.20}$$

Multiplying the above pair of successive observations determines $p(y_{1:n}|v_{0:n}, \theta)$:

$$p(y_{1:n}|v_{0:n}, \theta) = \prod_{k=1}^{n} p(y_k|y_{k-1}, v_{k-1:k}, \theta) \tag{4.21}$$

Hence,

$$\log[p(y_{1:n}|v_{0:n}, \theta)] = C + \sum_{k=1}^{n} -\frac{1}{2}\log(\int_{t_{k-1}}^{t_k} \sigma_u(v_s)^2 ds) - \frac{[y_k - y_{k-1} - \int_{t_{k-1}}^{t_k}(\mu_u - \sigma_u(v_s)^2/2)ds]^2}{2 \times \int_{t_{k-1}}^{t_k} \sigma_u(v_s)^2 ds} \tag{4.22}$$

And accounting for the fact that integrals are discretised:

$$\log[p(y_{1:n}|v_{0:n}, \theta)] = C + \sum_{k=1}^{n} -\frac{1}{2}\log(\sum_{i=j_{t_{k-1}}}^{j_{t_k}-1} \sigma_u(v_i)^2\delta) - \frac{[y_k - y_{k-1} - \sum_{i=j_{t_{k-1}}}^{j_{t_k}-1}(\mu - \sigma_u(v_i)^2/2)\delta_t]^2}{2 \times \sum_{i=j_{t_{k-1}}}^{j_{t_k}} \sigma_u(v_i)^2\delta_t} \tag{4.23}$$

From this formula, the expressions for $\nabla_q \log[p(y_{1:n}|v_{0:n}, \theta)]$ can be derived using the chain rule:

$$\begin{aligned}
\nabla_{x_{0:2N}} \log p(y_{1:n}|v_{0:n}, \theta) &= \left(\frac{dg_{0:n}}{dx_{0:2N}}\right)^\top \left(\frac{dv_{0:n}}{dg_{0:n}}\right)^\top \nabla_v \log p(y_{1:n}|v_{0:n}, \theta_y) \\
\nabla_{\theta_g} \log p(y_{1:n}|v_{0:n}, \theta_y) &= \left(\frac{dg_{0:n}}{d\theta_v}\right)^\top \left(\frac{dv_{0:n}}{dg_{0:n}}\right)^\top \nabla_v \log p(y_{1:n}|v_{0:n}, \theta_y) \\
\nabla_{\theta_v} \log p(y_{1:n}|v_{0:n}, \theta) &= \left(\frac{dv_{0:n}}{d\theta_v}\right)^\top \nabla_v \log p(y_{1:n}|v_{0:n}, \theta_y) \\
\nabla_{\theta_y} \log p(y_{1:n}|v_{0:n}, \theta) &= \nabla_{\theta_y} \log p(y_{1:n}|v_{0:n}, \theta_y)
\end{aligned} \tag{4.24}$$

The direct implementation of these formulas leads to a quadratic cost in N. We will present here how the complexity of these calculations can be reduced to $O(N)$ or $O(N \log N)$, focusing on the illustrative cases of $\left(\frac{dv_{0:n}}{dg_{0:n}}\right)^\top$, $\nabla_v \log p_N(y_{0:n}|v_{0:n}, \theta_y)$ and $\left(\frac{\partial g_{0:n}}{\partial H}\right)$. First, it appears clearly that

$$\frac{\partial v_i}{\partial g_j} = 0 \quad if\ i \leq j \tag{4.25}$$

and

$$\frac{\partial v_{i+1}}{\partial g_j} = \frac{\partial v_i}{\partial g_j}(1 + \kappa\delta) + \sigma_v \delta_{ij} \tag{4.26}$$

Thus,

$$\frac{\partial v_i}{\partial g_j} = \sigma_v(1-\kappa\delta_t)^{i-j} \times \delta_{i>j} \tag{4.27}$$

This can be used to iterative calculate $\nabla_v \log p(y_{1:n}|v_{0:n}, \theta_y)$ and $\left(\frac{dv_{0:n}}{dg_{0:n}}\right)^\top$:

$$\begin{aligned}\frac{\partial \log[p(y_{1:n}|v_{0:n},\theta)]}{\partial g_{0:n}} &= \sum_i \left(\frac{dv_i}{dg_j}\right)^\top \frac{\partial \log p(y_{1:n}|v_{0:n},\theta_y)}{\partial v_i} \\ &= (1+\kappa\delta)\frac{\partial \log[p(y_{1:n}|v_{0:n},\theta)]}{\partial g_{j+1}} + \sigma_v \frac{\partial \log p(y_{1:n}|v_{0:n},\theta_y)}{\partial v_{j+1}}\end{aligned} \tag{4.28}$$

Similarly, the calculation of $\left(\frac{\partial g_{0:n}}{\partial H}\right)$ can be achieved in $O(N \log N)$ operations through the eigen decomposition allowed by the fast Fourier transform:

$$\frac{\partial g_{0:n}}{\partial H} = P\frac{\partial \Lambda^{1/2}}{\partial H} Q x_{0:2N} \tag{4.29}$$

Similarly, all the necessary gradients of $\phi(x_{0:2N}, \theta)$ can be computed with $O(N)$ or $O(N \log N)$ operations. Exhaustive details for the different calculations involved can be found in the appendix (Appendix B).

4.3.3 Simulation and Results

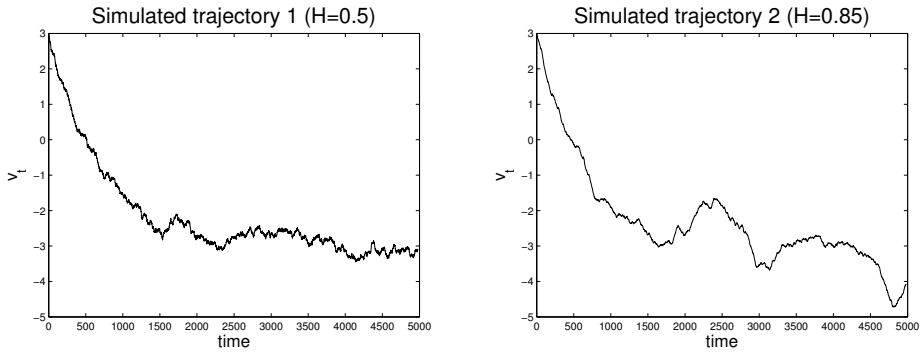

Figure 4.2: Simulated paths of fractional stochastic volatility with respective Hurst parameters 0.5 and 0.85. Additionally, we set $\sigma_v = 0.08$, $\kappa = 0.03$, $\mu_v = -3$, and $v_0 = 3$.

This section focuses on two simulated realisations of the fractional stochastic volatility model introduced in 4.10. The choice of parameter values for these two simulations are based on the choices and findings made in Chronopoulou and Viens (2012): in both datasets, the volatility of v_t is set to 0.08, κ is equal to 0.03, μ_v and μ_u are respectively set to -3 and -0.002, and the initial value v_0 is set to 3. The only parameter that differs between the two datasets is the Hurst parameter H, that is set equal to 0.5 in Dataset 1, and 0.85 in Dataset 2

in order to compare the mixing performance of the advanced HMC algorithm both in the absence and presence of memory. We will also assess on this example the identifiability of the Hurst parameter H. Lastly, Dataset 1 will be used to compare the mixing performance of the advanced HMC and PMMH algorithms in the Markovian case (H fixed to 0.5).

Low-informative uniform priors are defined over the parameters σ_v ($Unif[0; 0.5]$), κ ($Unif[0; 20]$), μ_u ($Unif[-20; 20]$), μ_v ($Unif[-20; 20]$) and , v_0 ($Unif[-20; 20]$). Furthermore, H is constrained between 0.4 and 1, as evidence from the literature indicate that the increments of stochastic volatility models tend to be positively correlated (Chronopoulou and Viens, 2012). The length of the simulated time series is 250, which corresponds to a year of daily asset prices. The discretisation time step δ is set to 0.05, leading to an underlying vector $x_{0:2N}$ of dimension 5,000.

Sampling performance

We start by comparing the performance of different versions of the advanced HMC algorithm on Datasets 1 and 2. In particular, we explore the possibility of updating $x_{0:2N}$ and θ simultaneously (this will be indicated with the superscript *Joint*), or in an alternate manner (superscript *Gibbs*). For both versions, three values of L are considered: 1, 10, 20 (the value of L is indicated by the subscript). When $L = 1$, the algorithm is equivalent to the MALA. The leapfrog size h is manually calibrated each time to set the acceptance rate between 60% and 90%, and the algorithm is run for 20,000 iterations. The computation time, that is accounted for to compare the efficiency of the different algorithms, corresponds to a Matlab implementation run on a Macbook Pro with 2.3 GHz Intel Core i7 processor.

Results of these experiments are provided in Tables 4.1 and 4.2. For each case, the minimum ESS (in %) over θ and $x_{0:2N}$ is given, along with the CPU time taken by each iteration of the algorithm. From these numbers, the average number of independent samples generated over 100 seconds is computed, which allows a comparison of the relative efficiencies of the different algorithms. First of all, the results suggest that efficient exploration of the target density $p(x_{0:2N}, \theta|y_{0:n})$ can be be achieved by using the advanced HMC algorithm, irrespectively of the absence ($H = 0.5$) or presence ($H = 0.85$) of underlying memory. The most time-efficient implementation is HMC_{20}^{Joint}, where 20 joint leapfrog steps are made at each iteration to update the paths and parameters simultaneously. In average, this algorithm allows to generate the equivalent of 1.9 independent samples every 100 seconds: reliable estimates of the paths and parameters can be obtained within several hours from a year-long dataset of asset prices. Besides, the Tables indicate that updating the paths and parameters jointly is uniformly better than performing the updates alternatively. This result suggests that the paths and parameters of the system are strongly correlated, and that updating them together is beneficial despite the increased dimension of the target space.

Sampler	$min_\theta(ESS)$	$min_x(ESS)$	time	$100 \times \frac{min_{\theta,x}(ESS)}{time}$	relative $\frac{min_{\theta,x}(ESS)}{time}$
HMC_1^{Gibbs}	0.1%	2.7%	0.22	0.45	1.00
HMC_{10}^{Gibbs}	0.5%	2.2%	0.78	0.64	1.42
HMC_{20}^{Gibbs}	0.9%	1.9%	1.44	0.62	1.38
HMC_1^{Joint}	0.1%	0.5%	0.10	1.00	2.22
HMC_{10}^{Joint}	0.4%	1%	0.53	0.75	1.67
HMC_{20}^{Joint}	1.8%	3%	0.97	1.90	4.22

Table 4.1: Relative efficiency of the different implementations of the advanced HMC algorithm on Dataset 1 ($H = 0.5$), via the minimum ESS (%) and CPU times (seconds).

Sampler	$min_\theta(ESS)$	$min_x(ESS)$	time	$100 \times \frac{min_{\theta,x}(ESS)}{time}$	relative $\frac{min_{\theta,x}(ESS)}{time}$
HMC_1^{Gibbs}	0.1%	0.7%	0.22	0.45	1.18
HMC_{10}^{Gibbs}	0.3%	3.8%	0.78	0.38	1.00
HMC_{20}^{Gibbs}	1.0%	6.0%	1.44	0.69	1.82
HMC_{10}^{Joint}	0.1%	0.5%	0.10	1.00	2.63
HMC_{10}^{Joint}	0.5%	1.0%	0.53	0.94	2.47
HMC_{20}^{Joint}	2.2%	3.0%	0.97	2.26	5.95

Table 4.2: Relative efficiency of the different implementations of the advanced HMC algorithm on Dataset 2 ($H = 0.85$), via the minimum ESS (%) and CPU times (seconds).

Identifiability of the Hurst parameter

After validating the mixing performance of the advanced HMC algorithm, we assess the identifiability of the different parameters involved in the model. The parameter H is of particular interest, as it characterises the level and type of memory of the system. Figures 4.3 and 4.4 suggest that under the vague uniform priors imposed in these experiments, all six parameters are identifiable. The posterior density $p(H|y_{1:n})$ is only weakly informative as it spans all over the range of values allowed (0.4 to 1), but its mode provides a good estimate of H on these two examples. Indeed, its value is 0.47 for Dataset 1, and 0.91 for Dataset 2. Additionally, Figure 4.5 illustrates that the indirectly observed path of the fractional stochastic volatility model can be efficiently reconstructed from daily observed asset prices.

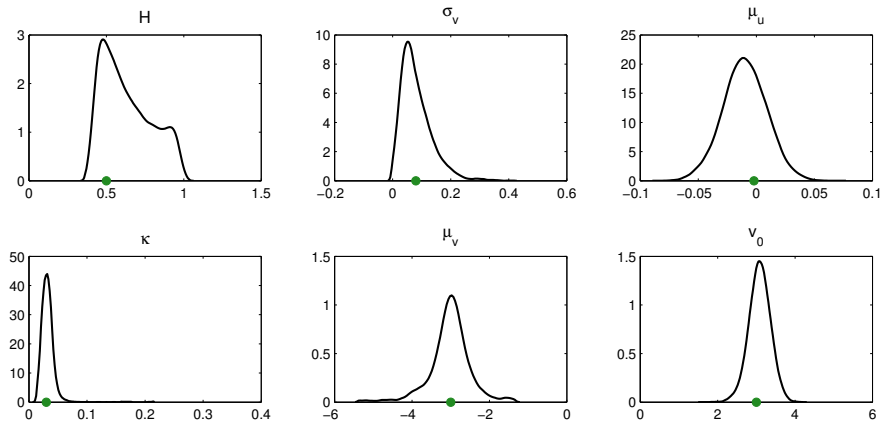

Figure 4.3: Posterior marginal densities for the different parameters of the model with Dataset 1 (H=0.5). True values are indicated with a green dot.

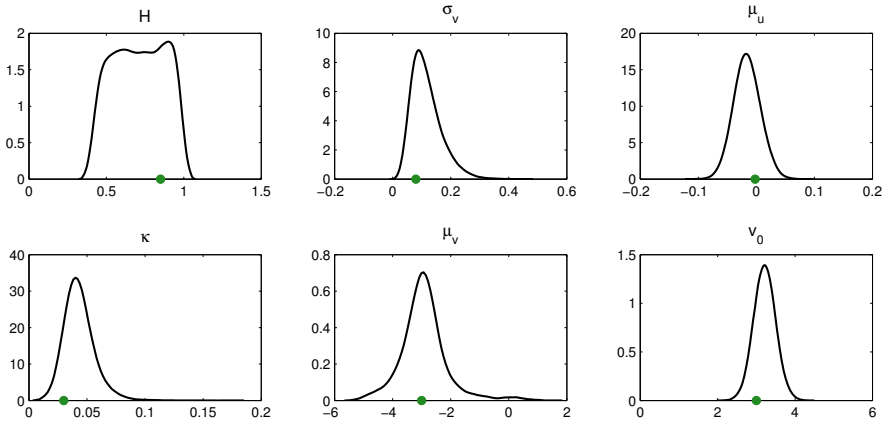

Figure 4.4: Posterior marginal densities for the different parameters of the model with Dataset 2 (H=0.85). True values are indicated with a green dot.

Comparison with the particle MCMC algorithm in the Markovian setting

When the Hurst parameter H is fixed to 0.5, the model we are exploring collapses to a classic Markovian stochastic volatility model to which the PMMH algorithm can be applied. We now compare the performance of the advanced HMC and PMMH algorithms on Dataset 1 in sampling from $p(x_{0:2N}, \theta | y_{1:n})$, where θ is a five-dimensional parameter $\{\sigma_v, \kappa, \mu_u, \mu_v, v_0\}$. As in the previous section, the advanced HMC algorithms are imple-

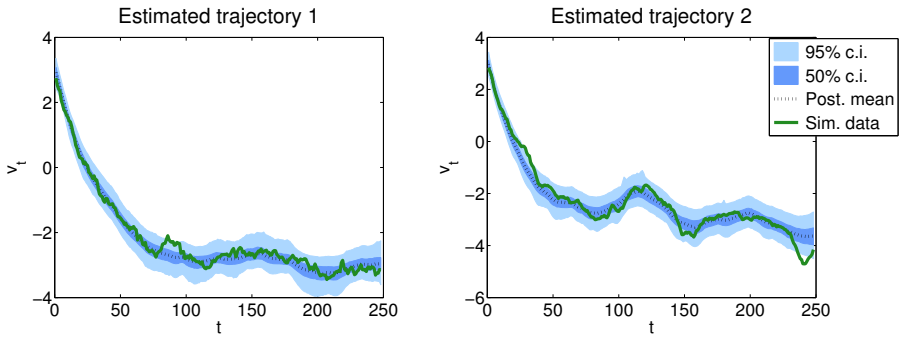

Figure 4.5: Simulated paths of fractional stochastic volatility with respective Hurst parameters 0.5 and 0.85. Additionally, we set $\sigma_v = 0.08$, $\kappa = 0.03$, $\mu_v = -3$, and $v_0 = 3$.

mented in two versions specified by their superscripts $Gibbs$ and $Joint$, indicating whether $x_{0:2N}$ and θ are updated alternatively or jointly. The PMMH is implemented in its simplest version, relying on a default bootstrap sampler and a random walk exploration of the parameter space. A basic adaptive scheme is used where the multiplicative factor of the sampling covariance Σ^q is tuned to achieve a 23% acceptance rate. The samples generated during the transient adaptation phase are discarded in the performance analysis; results are presented in Tables 4.3. In Table 4.4, we consider the case where the scaling mass matrix M of the HMC and the importance sampling covariance matrix Σ^q of the PMMH are no longer set to Id but to the covariance of the posterior density $p(\theta|y_{1:n})$. The dependency of the acceptance rate as a function of the number of particles is illustrated in Figure 4.6. In order to disentangle the impact of the Monte Carlo variability of the likelihood estimate on the PMMH mixing efficiency and the intrinsic performance of the PMMH scheme, two versions of the algorithm are tested with respectively 100 and 2000 particles.

Tables 4.3 and 4.4 suggest that on this example the advanced HMC algorithm compares favourably with the PMMH based on a random walk exploration of the parameters space. To begin with, the comparison of the two series of simulations show that adapting the mass matrix is highly beneficial, as it multiplies by 9 and 14 the mixing efficiency of PMMH algorithms, and by 6 the one of the HMC_{20}^{Joint} algorithm. However, we note that it has no significant impact on the performance of the HMC_{20}^{Gibbs} algorithm. This result indicates, as in the analysis conducted in the Non-Markovian setting, that correlation between the paths and parameters of the system is the limiting factor of the Gibbs implementation of the HMC algorithm. As a consequence, when using the optimal mass matrix $M = Cov[p(\theta|y_{1:n})]$, the PMMH_{100} algorithm mixes better than HMC_{20}^{Gibbs}. In addition, we note that increasing the number of particles of the PMMH algorithm from 100 to 2000 does

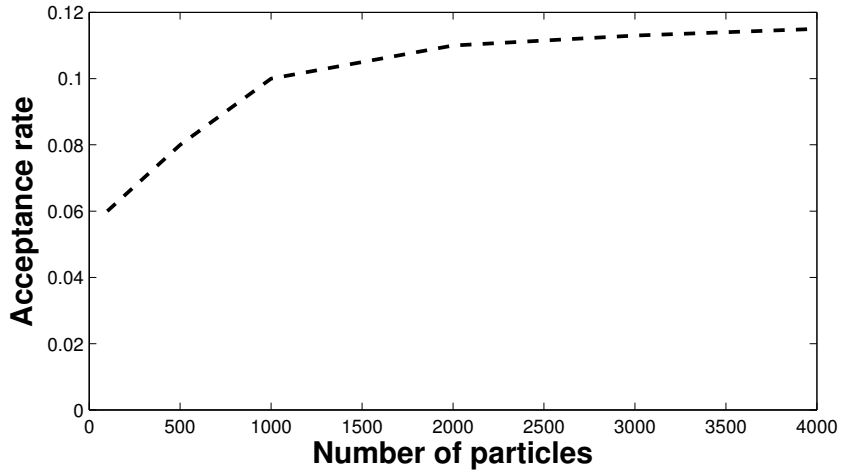

Figure 4.6: Acceptance rate of the PMMH on a stochastic volatility model (H=0.5) as a function of the number of particles.

Sampler	$min_\theta(ESS)$	$min_{\tilde{x}}(ESS)$	time	$100 \times \frac{min_{\theta,\tilde{x}}(ESS)}{time}$	relative $\frac{min_{\theta,\tilde{x}}(ESS)}{time}$
PMMH_{100}	0.1%	1.0%	0.32	0.31	1.0
PMMH_{2000}	0.1%	1.2%	1.02	0.1	0.3
HMC_1^{Gibbs}	0.1%	1.8%	0.17	0.58	1.9
HMC_{10}^{Gibbs}	1.0%	4.5%	0.62	1.61	5.2
HMC_{20}^{Gibbs}	2.4%	4.2%	1.17	2.05	6.6
HMC_1^{Joint}	0.1%	0.5%	0.09	1.11	3.6
HMC_{10}^{Joint}	0.6%	1.0%	0.40	1.50	4.8
HMC_{20}^{Joint}	2.4%	2.9%	0.76	3.16	10.2

Table 4.3: Relative efficiency, via the minimum ESS (%) and CPU times (seconds). In this case, the mass (M) and sampling covariance (Σ^q) matrices are set to the Identity matrix.

not contribute significantly to the mixing efficiency, but multiplies the computational burden by three. For this reason, the version with 100 particles is taken as a reference for the performance of the PMMH algorithm: on this example, the *Joint* version of the advanced HMC algorithm with 20 leapfrog steps is 6 to 10 times more efficient than the PMMH algorithm. These results may vary depending on the number of observations being considered, as the reparameterisation used in the HMC requires $O(N \log N)$ operations, and the number of particles used in the PMMH would need to be adapted. The particle Gibbs algorithm version of the PMCMC introduced in Andrieu et al. (2010) has not be tested in our analysis, although it could have taken direct advantage of the formulas derived for

Sampler	$min_\theta(ESS)$	$min_{\tilde{x}}(ESS)$	time	$100 \times \frac{min_{\theta,\tilde{x}}(ESS)}{time}$	relative $\frac{min_{\theta,\tilde{x}}(ESS)}{time}$
PMMH_{100}	0.9%	1.7%	0.32	2.8	1.0
PMMH_{2000}	1.4%	2.0%	1.02	1.37	0.5
HMC_{20}^{Gibbs}	2.6%	2.7%	1.17	2.22	0.8
HMC_{20}^{Joint}	13.0%	40.3%	0.76	17.1	6.1

Table 4.4: Relative efficiency, via the minimum ESS (%) and CPU times (seconds). In this case, the mass (M) and sampling covariance (Σ^q) matrices are set to the posterior density Covariance matrix.

$\nabla_\theta p(\theta|x_{0:n}^{dis}, y_{1:n})$. However, it is likely that its mixing performance would have been similar to the ones of the HMC_{20}^{Gibbs} algorithm, as the latter seem to have been directly limited by the strong correlation between the path and parameters of the system.

Limitations

Although the results obtained with the fractional stochastic volatility model 4.10 are encouraging, difficulties are encountered when incorporating a leverage effect in the stochastic volatility model:

$$\begin{cases} du_t &= (\mu_u - \sigma_u(v_t)^2/2)dt + \sqrt{1-\rho^2}\sigma_u(v_t)dW_t + \rho\sigma_u(v_t)dB_t^H, \\ dv_t &= \kappa(\mu_v - v_t)dt + \sigma_v dB_t^H. \end{cases} \tag{4.30}$$

A traceplot of the parameter ρ generated from a run of the HMC_{10}^{Joint} algorithm is given in Figure 4.7, along with the associated joint likelihood values. This plot illustrates a limitation of the advanced HMC algorithm, as poor mixing performances of the algorithm are observed when the chain explores high values of ρ in absolute value. These difficulties can be understood by looking at the formula of the augmented log-likelihood in Equation 4.31. Conditioned on $x_{0:2N}$, the amplitude of the observation noise is proportional to $\sqrt{1-\rho^2}$ which corresponds to the term $-\frac{1}{2}\log(1-\rho^2)$. Consequently, the likelihood density $p(y_{1:n}|x_{0:2N}, \theta)$ as a function of $x_{0:2N}$, conditioned on high values of $|\rho|$, is more peaked: it reaches higher values (as illustrated in Figure 4.7), and induces smaller moves. Yet, the latter does not imply that the marginal posterior $p(\rho|y_{1:n})$ is necessarily higher for high values of $|\rho|$, as it is equal to the integral of the joint posterior $p(x_{0:2N}, \theta|y_{1:n})$ over all trajectories of $x_{0:2N}$.

$$\log[p(y_{1:n}|x_{0:2N},\theta)] = C + \sum_{k=1}^{n} -\frac{1}{2}\log((1-\rho^2)\sum_{i=j_{t_{k-1}}}^{j_{t_k}-1}\sigma_u(v_i)^2\delta_t) \tag{4.31}$$
$$- \frac{[y_k - y_{k-1} - \sum_{i=j_{t_{k-1}}}^{j_{t_k}-1}(\mu_u - \sigma_u(v_i)^2/2)\delta_t - \rho\sum_{i=j_{t_{k-1}}}^{j_{t_k}-1}\sigma_u(v_i)\Delta B_i^H]^2}{2\times(1-\rho^2)\sum_{i=j_{t_{k-1}}}^{j_{t_k}}\sigma_u(v_s)^2\delta_t}$$

The issue illustrated by this example can be imputed to the divergence of the joint likelihood function as ρ tends to ± 1 for volatility trajectories that closely match the observations. This problem may be tempered by the introduction of additional data, or by using more informative priors for the leverage parameter ρ. However, robust and general solutions to this problem may require the development of pseudo-marginal methods adapted to the non-Markovian setting, that would prevent the chain from being trapped in peaks of the joint likelihood. The latter is a subject for further research.

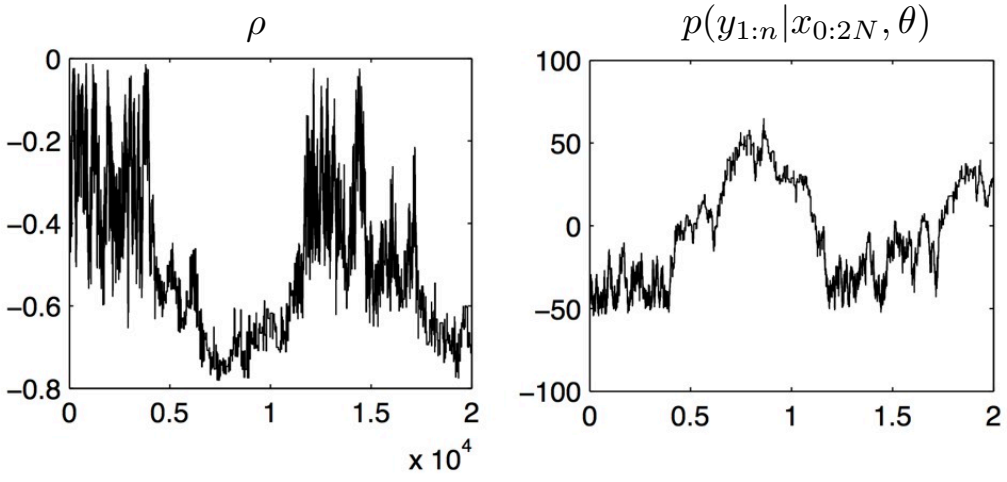

Figure 4.7: Traceplots of values taken by ρ and the augmented likelihood $p(y_{1:n}|\dot{x}_{0:n},\theta)$ during a problematic realisation of the HMC_{10}^{Joint} algorithm.

4.4 Discussion

We have illustrated in this Chapter the potential of advanced Hybrid Monte Carlo methods to achieve efficient exploration of the joint posterior density $p(x_{0:n},\theta|y_{1:n})$ for indirectly observed stochastic processes. Simulations have shown on two examples that efficient

inference for model 4.10 can be performed from a year-long time series of daily asset prices. These examples suggest that it is possible to infer the type and level of memory of the fractional Brownian motion driving a stochastic volatility model. The extension to model 4.30, however, indicates that difficulties can be encountered when strong dependencies between the paths and parameters appear a posteriori in some regions of the explored space.

In addition, we have shown on two simulated datasets that updating simultaneously the paths and parameters of the system with the advanced HMC algorithm could lead to better performance than a simple adaptive implementation of the PMMH algorithm. These results motivate further empirical and theoretical comparisons with the PMMH and PGibbs implementations of the PMCMC, as well as with the SMC^2 algorithm. For example, it will be interesting to compare the dependency of the cost induced by the different algorithms as a function of the number of observations n, the length of the parameter vector θ, and the dimension of the driving stochastic process x_t. In particular, if this dimension is denoted k, the authors of Daum and Huang (2003) have shown that in the general case the number of particles required to control the variance of the likelihood computed with a particle filter increases exponentially as k grows. On the other hand, if the advanced HMC algorithm was applied to a system with multi-dimensional driving stochastic process x_t, the number of gradient calculations would increases linearly with k. Nonetheless, the mixing efficiency may deteriorate with increasing values of k: this open question remains to be investigated. From the perspective of epidemiology, the possibility of achieving efficient and exact inference for non-Markovian systems may allow interesting extensions to the actual modelling framework. For example, non-exponential waiting times could be introduced in the models to offer a better representation of disease transmission dynamics. Additionally, if the linear complexity of the method as a function of k is confirmed, spatially-structured compartmental epidemic models in continuous time could become computationally tractable.

At last, the advanced HMC algorithm appears to imply cumbersome calculations to derive the analytic formulas of the gradient $\nabla\phi(q)$. However, the calculations presented in this Chapter and in the corresponding Appendix suggest that a systematic recursive formulation could be generalised for systems that are fully Markovian or Markovian conditionally on $x_{0:n}$. For such classes of problems that can be expressed in a generic manner, the use of gradient-driven inference methods as the advanced HMC could be facilitated by symbolic calculation libraries, thus extending the paradigm of plug-and-play algorithms.

CHAPTER 5

Future Research

The natural extensions of the work presented in Chapters 2, 3 and 4 have been discussed at the end of each Chapter. The theoretical foundations of a general framework reflecting environmental stochasticity through stochastic rates and time-varying parameters, both under the infinite population and multinomial approximations, will be explored further. The robustness of the Gaussian approximation utilised to extend the use of the Extended Kalman Filter to this general framework will also be the subject to further examination. This generic plug-and-play toolbox for inference on epidemic models, allowing for the automatic calibration of the particle MCMC, shall permit diverse applications in epidemiology and public health as illustrated in Chapters 2 and 3. Additionally, the novel methodology introduced in Chapter 4 offers promising perspectives: it permits exact and efficient exploration of the posterior marginal density $p(\theta, x_{0:n}^{dis}|y_{1:n})$ in the Non-Markovian setting, while being robust to refinements in the discretisation of time. Furthermore, the ability of the advanced HMC to jointly update θ and $x_{0:n}^{dis}$ at each iteration with global, gradient-driven moves makes it an interesting alternative to pseudo-marginal approaches. Its extension to more general settings with multi-dimensional underlying processes $x_{0:n}$ is a natural step for future research, along with a more thorough comparison with the PMMH, PGibbs and SMC^2 algorithms.

Lastly, we provide here with three related subjects of further research.

5.1 Bayesian inference for sparse high-dimensional systems

The epidemic models that have been utilised in this thesis did not account for the geographic distribution of individuals: all cases were aggregated over a given city (London, in Chapter 2), or over a given district (Southern Indian distrcts in Chapter 3). However, there is increasing evidence that the spatial dimension plays an important role in epidemic

dynamics. The authors of Grenfell et al. (2001), for example, show how measles stays endemic in large cities between two epidemics, and how smaller cities of the periphery are randomly reignited at each epidemic burst. More recent studies have identified similar traveling waves for influenza (Stark et al., 2012), and stressed the need for adapted tools to fully apprehend the epidemiological and evolutionary dynamics of viruses over space and time (Viboud et al., 2013). The general framework of population-level compartmental epidemic models explored in this thesis does allow for the definition of geographical patches (cities, districts, regions, etc). For example, a two-cities SI model could take the following form:

Reaction	**Effect**	**Rate**
Infection city 1	$(S_t^{(1)}, I_t^{(1)}, S_t^{(2)}, I_t^{(2)}) \to (S_t^{(1)} - 1, I_t^{(1)} + 1, S_t^{(2)}, I_t^{(2)})$	$\beta \frac{S_t^{(1)}}{N} I_t^{(1)}$
Infection city 2	$(S_t^{(1)}, I_t^{(1)}, S_t^{(2)}, I_t^{(2)}) \to (S_t^{(1)}, I_t^{(1)}, S_t^{(2)} - 1, I_t^{(2)} + 1)$	$\beta \frac{S_t^{(2)}}{N} I_t^{(2)}$
S migration, city 1 to city 2	$(S_t^{(1)}, I_t^{(1)}, S_t^{(2)}, I_t^{(2)}) \to (S_t^{(1)} - 1, I_t^{(1)}, S_t^{(2)} + 1, I_t^{(2)})$	$\alpha S_t^{(1)}$
I migration, city 1 to city 2	$(S_t^{(1)}, I_t^{(1)}, S_t^{(2)}, I_t^{(2)}) \to (S_t^{(1)}, I_t^{(1)} - 1, S_t^{(2)}, I_t^{(2)} + 1)$	$\alpha I_t^{(1)}$
S migration, city 2 to city 1	$(S_t^{(1)}, I_t^{(1)}, S_t^{(2)}, I_t^{(2)}) \to (S_t^{(1)} + 1, I_t^{(1)}, S_t^{(2)} - 1, I_t^{(2)})$	$\alpha S_t^{(2)}$
I migration, city 2 to city 1	$(S_t^{(1)}, I_t^{(1)}, S_t^{(2)}, I_t^{(2)}) \to (S_t^{(1)}, I_t^{(1)} + 1, S_t^{(2)}, I_t^{(2)} - 1)$	$\alpha I_t^{(2)}$

Naturally, the size of the state vector z_t containing the number of individuals in each compartment increases linearly with the number of geographical patches. Similarly, the number of reactions that correspond to infections also increase linearly, as infections occur independently in each patch. If migration is allowed between any couple of cities, the number of reactions that correspond to migrations increase quadratically. We have seen in the Introduction that the dimension of the driving stochastic process x_t was equal to the number of sources of noise. For the simple SI model proposed here, the dimension of x_t will increase linearly with the number of patches n_p if only independent environmental noise over the transmission rates β is considered. If all reactions are stochastic, which is the case when demographic stochasticity is accounted for, the dimension of x_t will increase quadratically with the number of patches n_p.

In the general case, the number of particles required to control the level of noise in the estimation of the marginal likelihood $p(\theta|y_{1:n})$ with a particle filter increases exponentially with the dimension of the driving stochastic process x_t (Daum and Huang, 2003). For this reason, exploring the spatial and temporal dynamics of an epidemic model accounting for the different sources of stochasticity with the PMCMC approach utilised in this thesis quickly becomes intractable. However, the correlation structure of the vector x_t will be sparse, with most information concentrated in diagonal blocks corresponding to each patch, in addition to some terms outside of these blocks corresponding to correlation in-

duced by migration. These terms shall be of much lower magnitude than the correlation observed within a same patch, except in the case of small cities in which epidemics are driven by bigger cities. In this setting, could the specific sparse structure of the correlation matrix of x_t be exploited to avoid the exponential cost of the estimation of $p(\theta|y_{1:n})$ with a particle filter? In particular, is it possible to achieve a linear computational complexity in the number of patches n_p by assuming conditional independence? Under this assumption, the model could be explored as n_p independent patches submitted to an external forcing from the other patches, with patch-specific independent particles. The exact definition and properties of this approximate filtering algorithm for sparse high-dimensional stochastic processes will be the subject of further research. Alternatively, if a high-dimensional MCMC solution is adapted to the exploration the joint posterior density $p(\theta, x_{0:n}|y_{1:n})$, the partial independence between patches may be exploited in an exact manner by running a Gibbs algorithm in which city-specific components of $\{\theta, x_{0:n}\}$ are updated alternatively.

5.2 Sequential version of the advanced Hybrid Monte Carlo Algorithm

Monte Carlo Markov Chains rely on an iterative exploration of the target space. Asymptotic convergence results ensure that the successive positions of the chain provide in the long run equivalent information than independent and identically distributed samples from the target density. One limitation of the advanced HMC algorithm, as any MCMC algorithm, is its intrinsic sequential structure that prevents direct parallelisation. A classic alternative to MCMC algorithms, as illustrated by the PMCMC and SMC^2 couple of methods frequently evoked in this thesis, are Sequential Monte Carlo algorithms.

Algorithm 14 Sequential version of the advanced HMC algorithm

Initialise J particles $\{(\theta^{(1)}, x_{0:n}^{(1)}), .., (\theta^{(J)}, x_{0:n}^{(J)})\}$ from the prior distribution $p(x_{0:n}|\theta)p(\theta)$
Compute the weights $\omega_1(\theta^{(j)}, x_{0:n}^{(j)}) = p(y_1|\theta^{(j)}, x_{0:n}^{(j)})$ and $W_1^i = \frac{\omega_1(\theta^{(j)}, x_{0:n}^{(j)})}{\sum_k \omega_1(\theta^{(k)}, x_{0:n}^{(k)})}$
Resample $\{W_1^j, (\theta^{(j)}, x_{0:n}^{(j)})\}$ to obtain J equally weighted particles $\{\frac{1}{N}, (\overline{\theta}^{(j)}, \overline{x}_{0:n}^{(j)})\}$
Sample $(\theta'^{(j)}, x_{0:n}'^{(j)}) \sim K_1(.|(\overline{\theta}^{(j)}, \overline{x}_{0:n}^{(j)}))$
for $i = 2$ to n **do**
  Compute the weights $\omega_i(\theta^{(j)}, x_{0:n}^{(j)}) = \frac{p(y_{1:i}|\theta^{(j)}, x_{0:n}^{(j)})}{p(y_{1:i-1}|\theta^{(j)}, x_{0:n}^{(j)})}$ and $W_i^j = \frac{\omega_i(\theta^{(i)}, x_{0:n}^{(i)})}{\sum_k \omega_i(\theta^{(j)}, x_{0:n}^{(j)})}$
  Resample $\{W_i^j, (\theta^{(j)}, x_{0:n}^{(j)})\}$ to obtain J equally weighted particles $\{\frac{1}{N}, (\overline{\theta}^{(j)}, \overline{x}_{0:n}^{(j)})\}$
  Sample $(\theta'^{(j)}, x_{0:n}'^{(j)}) \sim K_j(.|(\overline{\theta}^{(j)}, \overline{x}_{0:n}^{(j)}))$
end for

Hence, a sequential version of the advanced Hybrid Monte Carlo algorithm can be derived, as illustrated in Algorithm 14, based on the example of the SMC filter with MCMC

moves presented in Doucet and Johansen (2009). For examples as the fractional stochastic volatility model, the dimension of the target space is not augmented at each iteration due to the non-Markovian structure of the problem. The notation $K_j(.|\theta, x_{0:n})$ denotes the transition kernel of the advanced HMC algorithm conditioned on the j first observations, where a sample $(\theta^*, x^*_{0:n})$ is proposed from a Hamiltonian mapping based on $p(y_{1:j}|\theta^*, x^*_{0:n})$ and accepted or rejected through a Metropolis Hastings step. This SMC implementation of the advanced HMC algorithm allows a direct parallelisation of the calculations by computing the weights and updating the samples independently for each particle over multiple processors. Additionally, it has been argued that SMC algorithms could less easily be trapped in local modes (Del Moral et al., 2006), which may be a solution to the problems encountered when estimating the leverage parameter ρ in the fractional stochastic volatility model of Chapter 4. Lastly, improvements of this algorithm may be achieved by performing the rejuvenation step based on the transition kernel before the computing the weights and resampling the particles. Indeed, as suggested in Doucet and Johansen (2009), the distribution of the samples would be closer to the target density $p(\theta, x_{0:n}|y_{1:j})$.

5.3 Epidemic dynamics and climate: a mechanistic exploration

In Chapters 2 and 3 of this thesis, we have proposed a mean to capture the time-varying drivers of epidemics, and we have explored the applications of this approach in a public health setting focusing on evolving contact patterns and behaviour. We have shown how hypothesis on the drivers of epidemics could be tested in real time during an epidemic, and how unobserved and evolving behaviour could be indirectly estimated from epidemiological data. An alternative time-varying driver of epidemics is climate, that can play both a direct and indirect role. It has been proved by experimentation in laboratories that absolute humidity had a significant impact on the influenza virus survival and transmission (Shaman and Kohn, 2009). Furthermore, climate seasonality influences the time spent indoors and the nature of contacts (Lipsitch and Viboud, 2009). The significant influence of climate is also a likely hypothesis for diseases that are transmitted through mosquitoes: the density of *Aedes Aegypti* and *Aedes Albopictus* has been shown to be strongly seasonal, with peaks in rainy periods (Rao, 1967; Schultz, 1993). However, the precise impact of the forcing of climate on the inter-annual variability of infections cases has been hardly quantified, and is a recurrent point of debate. Indeed, the variability in the yearly number of cases of some seasonal diseases as Dengue or Influenza can also be linked to the antigenic drift, or to complex interactions between strains and with other pathogens. An example of irregular dynamics of yearly epidemics is given in Figure 5.1, that shows the number of secondary cases of Dengue along time in the district of Chiang Mai, Thailand, and ex-

hibits a strong inter-annual variability. The authors of Aguiar et al. (2011), for example, have proposed a parsimonious multi-strain model of Dengue that produces similar types of variability solely through the chaotic properties of the system, with the influence of climate being restricted to a simple sinusoidal multiplicative factor of the effective contact rate: $\beta_t = \beta_0 \times (1 + e\sin(\omega t + \phi))$. More direct confrontation of models to data is required; in Cazelles et al. (2005), climatic time-series are confronted to time-series of Dengue in Thailand and analised through wavelet analysis. By comparing not only the type of dynamics but rather by exploring in details the synchrony of both signals, further evidence is provided for the significant role of climate in explaining the inter-annual variability of Dengue epidemics.

This problem offers a natural application to the inference framework proposed in this thesis to capture the influence of time-varying drivers of epidemics. The parsimonious multi-strain model introduced in Aguiar et al. (2011) can be used as a basis to indirectly infer the dynamic of the effective contact rate β_t from the available incidence data. Different models could be used to model the dynamics of this parameter: it could for example be simply modelled as a postivively constrained Brownian motion ($d\log\beta_t = \sigma dB_t$), or a periodic component could be explicitly added, restricting the analysis to the anomaly term $\tilde{\beta}_t$ modelled as a randomly time-varying quantity:

$$\begin{aligned} \beta_t &= \beta_0 \times (1 + e\sin(\omega t + \phi) + \tilde{\beta}_t) \\ d\tilde{\beta}_t &= dB_t \end{aligned} \tag{5.1}$$

Estimating the trajectories $p(\beta_{0:n}|y_{1:n})$ or $p(\tilde{\beta}_{0:n}|y_{1:n})$ may be a mean to start disentangling the role of climate from the consequences of evolutive competition between the four co-existing Dengue strains. Preliminary results on this problem are encouraging: a strongly significant correlation (0.28, with p-value below 10^{-7}) has been identified between $p(\tilde{\beta}_{0:n}|y_{1:n})$ and the anomalies of the El Niño 1-2 index. The problem will be further explored, in the perspective of building an explicit forcing model $p(\tilde{\beta}_{0:n}|y^{clim}_{1:n})$. From such a model, Bayesian synthesis of climate and epidemic data could be achieved, leading to the exploration of $p(\tilde{\beta}_{0:n}, \theta|y^{epi}_{1:n}, y^{clim}_{1:n})$. It is possible that the integration of both sources of data may increase our understanding of Dengue dynamics, and extend the predictability horizon of Dengue epidemics. Furthermore, it might also contribute to the understanding and response to the re-emergence of Dengue in regions like Europe, from which it had disappeared since 1927.

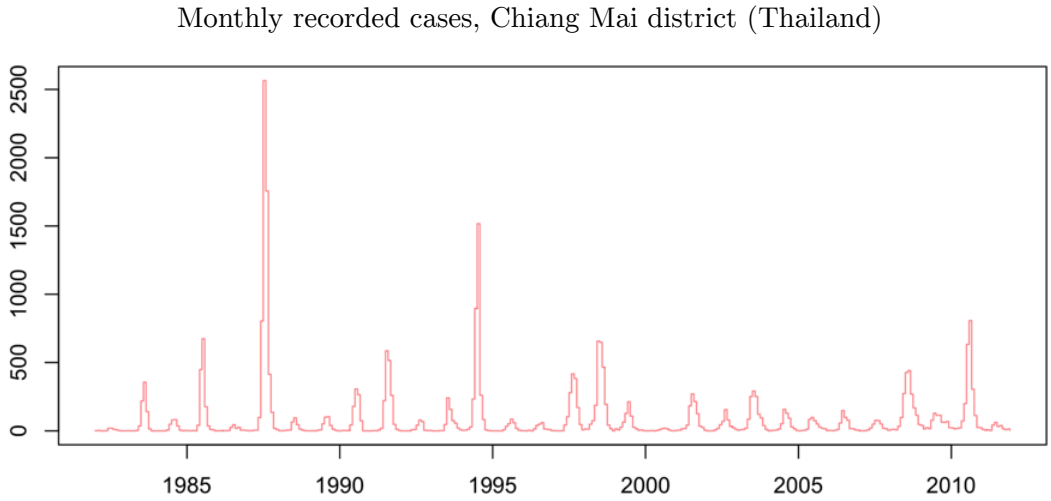

Figure 5.1: Monthly recorded secondary Dengue cases in the Chiang Mai district, Thailand

APPENDIX **A**

Supplementary material for Chapter 2

Assessing the validity and limitations of the Extended Kalman approximation

In this appendix we compare the particle filter with the Extended Kalman Filter, to quantify the impact of Gaussian and Taylor approximations on the path estimates of the system. A set of 100, 7-month long time-series of weekly influenza cases were drawn from the BM model. In order to simulate realistic epidemic datasets, we randomly selected influenza time-series $(y_{1:n}^{Goog,j})_{j=1}^{100}$ from the freely available Google FluTrend data (Ginsberg et al., 2008). For each of the datasets, we obtained estimates of $(\beta_{0:n}^{Sim,j})_{j=1}^{100}$ and corresponding parameters $(\theta_j)_{j=1}^{100}$. These were then used to generate time-series $(y_{1:n}^{Sim,j})_{j=1}^{100}$. The static parameters of the model (initial conditions, k, γ, σ and τ) were assumed to be known, in order to isolate the problem of estimating β_t. We compare the following two estimators of β_i^{Sim}:

- $\hat{\beta}_i^{pf} = E^{pf}(\beta_i|y_{1:i}^j, \theta^*)$ obtained from the filtering distribution
- $\hat{\beta}_i^{EKF} = E^{EKF}(\beta_i|y_{1:i}, \theta^*)$ where $E^{EKF}(.)$ denotes expectation under the approximate Extended Kalman Filtering density.

The performance of the estimators is measured through their bias and Mean Squared Error (MSE). The bias of the estimates provided by the EKF is 0.0285 while use of $\hat{\beta}_i^{Filt,j}$ reduces the bias by about 78% (0.0063). The corresponding relative reduction in MSE is 10% (0.0270 to 0.0242). Use of the smoothing distribution estimator is associated with a further 87% (0.0032) reduction in the MSE. In conclusion, the bias introduced by the Extended Kalman approximation is non-negligible with regards to the level of accuracy that can be obtained with exact particle methods on this type of datasets. Nevertheless,

this study has shown the approximation to be reasonable and motivates the use of the approximated model as a proxy for the exact one.

	MSE	**Bias**
EKF	0.0269	0.0285
Particle filtering	0.0242	0.0064
Improvement with regards to EKF	-10%	-77%
Particle smoothing	0.0032	0.0027
Improvement with regards to P. filtering	-87%	-64%

Table A.1: Mean Squarred Error and Bias of β_t estimates provided by the EKF, particle filter and particle smoother

Details on Gibbs data-augmentation schemes

In this section we provide more details on the Gibbs schemes discussed in Chapter 2. Stochastic epidemic models presented in this Chapter can be written as

$$\begin{cases} dx_t^{\theta_t} & = \mu^{\theta_t}(x_t^{\theta_t}, \theta)dt + \sigma^{\theta_t}(x_t^{\theta_t}, \theta)dB_t, \;\; 0 < t < t_n \\ y_{1:n}|z_{0:n}, \theta & \sim \mathbb{P}_y(y_{1:n}|z_{0:n}, \theta) \\ z_{0:n} & = f(x_{0:n}^{\theta_t}, \theta) \end{cases} \tag{A.1}$$

where z_t represents the ODE states vector observed trough partial and noisy data $y_{1:n}$. The rest of the model is defined in section 3.1 of Chapter 2. Since it contains intractable densities we work with the time discretised versions $x_{0:n}^{dis}$ and $z_{0:n}^{dis}$ and proceed using the Euler approximating scheme. A Gibbs algorithm alternates between updating the trajectories of $x_{0:n}^{dis}$ (and consequently $z_{0:n}^{dis}$) given θ, and vice versa. Nevertheless, as the Euler time step δ goes to 0, the quadratic variation process of $x_t^{\theta_t}$ uniquely determines the value of θ in $\sigma^{\theta_t}(.)$ and the algorithm degenerates (Roberts and Stramer, 2001). In practice this translates into a mixing time of $O(\delta^{-1})$.

In the context of diffusion-driven epidemic models, this problem was dealt with suitable reparametrisations such as the ones in Chib et al. (2006) or Kalogeropoulos (2007). The latter uses the Lamperti transform, i.e. $x_t^{\theta_t} \to H(x_t^{\theta_t}, \theta) = \eta(x_t^{\theta_t}, \theta) - \eta(x_0^{\theta_t}, \theta) =: u_t$ where $\eta(\cdot;\theta)$ is an antiderivative of $(\sigma^{\theta_t})^{-1}(\cdot;\theta)$. Assuming that $\sigma^{\theta_t}(\cdot;\theta)$ is continuously differentiable, an application of Ito's lemma provides the SDE of the transformed diffusion u_t as:

$$du_t = \nu(u_t;\theta)dt + dB_t \;, \quad u_0 = 0 \;, \tag{A.2}$$

where

$$\nu(u_t;\theta_x) = \frac{\mu_x\big(H^{-1}(u_t,\theta),\theta_x\big)}{\sigma^{\theta_t}\big(H^{-1}(u_t,\theta),\theta\big)} - \frac{1}{2}\sigma^{\theta_t\prime}\big(H^{-1}(u_t,\theta),\theta\big)\ .$$

Let $\mathbb{P}^u$ denote the distribution of u_t. Girsanov formula provide its density with respect to that of a standard Brownian motion, denoted by $\mathbb{W}$,

$$\frac{d\mathbb{P}^u}{d\mathbb{W}} = \exp\left\{\int_0^{t_n}\nu(u_s;\theta)du_s - \frac{1}{2}\int_0^{t_n}\nu(u_s;\theta)^2 ds\right\}, \tag{A.3}$$

and the state vector can be written as

$$z_{0:n} = h_u(u_{0:n}, x_0^{\theta_t}, \theta, \theta_v). \tag{A.4}$$

The model can be defined from (A.3), (A.6) and (A.1). It contains intractable quantities but can be accurately approximated given the time discretisation of the diffusion path. An alternative reparametrisation, defined in discrete time, was suggested in Chib et al. (2006). It uses the transformation below

$$w_t = \frac{x_t - \big(x_{t-\delta}^{\theta_t} - \delta\mu^{\theta_t}(x_t^{\theta_t},\theta)\big)}{\sigma^{\theta_t}(x_{t-\delta}^{\theta_t},\theta)}, \forall t. \tag{A.5}$$

In our setting the driving Brownian motion of $x_{0:n}^{dis}$, denoted by $w_{0:n}^{dis}$ and provided by (A.5), can be used to provide a discrete skeleton of the state vector

$$z_{0:n}^{dis} = h_w(w_{0:n}^{dis}, x_0^{\theta_t}, \theta, \theta_v). \tag{A.6}$$

The model is now given by $w_{0:n}^{dis}$, that can be transformed to $x_t^{\theta_t}$ for which the Euler-Maruyama approximation can be used, and A.1 which can be approximated using the discretised state vector $z_{0:n}^{dis}$.

Data augmentations schemes can be used for the models above. Gibbs versions of such schemes will alternate between updating $u_{0:n}^{dis}$ (or $w_{0:n}^{dis}$) and consequently $z_{0:n}^{dis}$ given θ, and θ conditional on either $u_{0:n}^{dis}$ (or $w_{0:n}^{dis}$). The first step can be done either by the overlapping block strategies in Chib et al. (2006) and Kalogeropoulos (2007) or with a particle filter in the context of a particle Gibbs algorithm. The second step of updating θ given $u_{0:n}^{dis}$ or $w_{0:n}^{dis}$ is usually implemented through a random walk Metropolis algorithm:

- Let θ^c and $z_{0:n}^{dis}$ be the current values of θ and $z_{0:n}^{dis}$ respectively. Propose θ^* from $q(\theta_x^*|\theta^c)$.
- Compute $z_{0:n}^{dis*} = h_u(u_{0:n}, x_0^{\theta_t}, \theta^*, \theta_v)$
- Accept with probability

$$1 \wedge \frac{\pi(\theta^*, z_{0:n}^{dis*}|y_{1:n}, u_{0:n}^{dis*}\theta_v)q(\theta^c|\theta^*)}{\pi(\theta^c, z_{0:n}^{disc}|y_{1:n}, u_{0:n}^{disc}\theta_v)q(\theta^*|\theta^c)}$$

For the Chib et al. (2006) formulation, $u_{0:n}^{dis}$ can simply be replaced with $w_{0:n}^{dis}$ in the algorithm above.

Unfortunately both of the above algorithms may perform poorly. The problem is that every proposed value of θ implies a proposed trajectory of the ODE states vector z_t. As parts or functionals of this trajectory are observed with error, the proposed value of θ will not be accepted unless its associated z_t trajectory is close to these observations. Consequently, only small steps can be made on the θ space and the algorithm mixes very slowly. The problem intensifies as the noise variance becomes smaller and as the time horizon of the epidemic increases. Implementations of such algorithms in the simulated and real data of this Chapter are in line with this argument. Figure A displays the posterior draws for σ in the dataset of the simulation experiment 1 of Section 4 in Chapter 2. The posterior draws of σ were obtained from a particle Gibbs algorithm combined with the algorithm above; note that in this model $u_{0:n}^{dis}$ and $w_{0:n}^{dis}$ are equal. In order to isolate the problem, the algorithm was run on σ and β_t only, and all the other parameters were held fixed at the values they where simulated from (a value of $\tau = 0.1$ was used). The *true value* of σ was 0.07 and we used a $\delta = 0.1$. As clearly shown in the traceplot the mixing of the chain is quite poor, thus casting doubts on the reliability of its output. The difference in mixing quality with the corresponding traceplot generated with the PMCMC algorithm is substantial.

To sum up, both formulations of Gibbs data augmentation schemes (with or without reparametrisation) are very likely to lead to inaccurate and inefficient MCMC algorithms. The use of particle MMH algorithms is therefore essential and Chapter 2 focuses on its implementation on diffusion driven epidemic models.

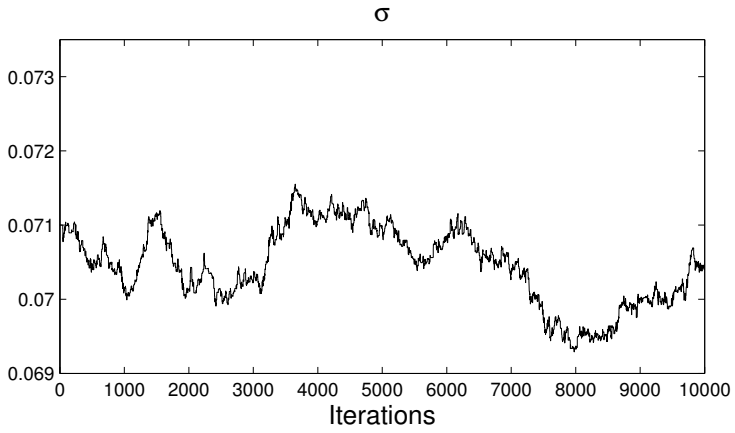

Figure A: MCMC traceplot for σ when using a Particle Gibbs scheme with reparametrization

APPENDIX **B**

Supplementary material for Chapter 4

We provide here with the detailed formula used to compute the different gradients involved in the application of the advanced Hybrid Monte Carlo algorithm to the fractional stochastic volatility model:

$$\frac{\partial \log p(y_{1:n}|v_{0:n},\theta_y)}{v_i},\ \frac{\partial g_{0:n}}{\partial \tilde{x}},\ \frac{\partial g_{0:n}}{\partial H},\ \frac{\partial v_i}{\partial \sigma_v},\ \frac{\partial v_i}{\partial \kappa},\ \frac{\partial v_i}{\partial \mu_v},\ \frac{\partial v_i}{\partial v_0},\ \frac{\partial \log p(y_{1:n}|v_{0:n},\theta_y)}{\partial \mu_u},$$

First, if k is defined as $t_{k-1} \leq j < t_k$:

$$\begin{aligned}\frac{\partial \log[p(y_{1:n}|v_{0:n},\theta)]}{\partial v_j} =& -\frac{\sigma_u'(v_j)\sigma_u(v_j)\delta_t}{\sum_{i=j_{t_{k-1}}}^{j_{t_k}-1}\sigma_u(v_i)^2\delta_t} \\ &- \frac{\sigma_u'(v_j)\sigma_u(v_j)\delta_t[u_k - y_{k-1} - \sum_{i=j_{t_{k-1}}}^{j_{t_k}-1}(\mu_u - \sigma_u(v_i)^2/2)\delta_t]}{\sum_{i=j_{t_{k-1}}}^{j_{t_k}}\sigma_u(v_i)^2\delta_t} \\ &+ \frac{[y_k - y_{k-1} - \sum_{i=j_{t_{k-1}}}^{j_{t_k}-1}(\mu_u - \sigma_u(v_i)^2/2)\delta_t]^2 \times \sigma_u'(v_j)\sigma_u(v_j)\delta_t}{(\sum_{i=j_{t_{k-1}}}^{j_{t_k}}\sigma_u(v_i)^2\delta_t)^2}\end{aligned} \tag{B.1}$$

By direct application of matrix calculus:

$$\frac{\partial g_{0:n}}{\partial \tilde{x}} = P\Lambda^{1/2}Q \tag{B.2}$$

As mentioned in Chapter 3:

$$\frac{\partial g_{0:n}}{\partial H} = P\frac{\partial \Lambda^{1/2}}{\partial H}QZ, \tag{B.3}$$

where

$$\frac{\partial \Lambda_{i,i}^{1/2}}{\partial H} = \sum_{j=0}^{2N-1} \frac{\partial c_{0,j}}{\partial H} \exp\big(-2\pi i \, \tfrac{jk}{2N}\big) \tag{B.4}$$

Which can be directly derived from the fact that for $k \geq 2$,

$$\begin{aligned}\frac{\gamma(k)}{\partial H} =& 2\log(\delta)\delta^{2H}(0.5|k+1|^{2H} + 0.5|k-1|^{2H} - |k|^{2H}) && \text{(B.5)}\\ &+ \delta^{2H}(\log(|k+1|)|k+1|^{2H} + \log(|k-1|)|k-1|^{2H} - 2\log(|k|)|k|^{2H}) && \text{(B.6)}\end{aligned}$$

Otherwise:
if $k = 0$, $\partial\gamma(k)/\partial H = 2\log(\delta)\delta^{2H}$
if $k = 1$, $\partial\gamma(k)/\partial H = 2\log(\delta)\delta^{2H}(0.5 \times 2^{2H} - 1) + \delta^{2H}\log(2)2^{2H}$

$\frac{\partial v_i}{\partial \sigma_v}$ can be computed recursively using the following formula:

$$\frac{\partial v_{i+1}}{\partial \sigma_v} = \frac{\partial v_i}{\partial \sigma_v}(1 - \kappa\delta) + g_i \tag{B.7}$$

Similarly,

$$\frac{\partial v_{i+1}}{\partial \kappa} = \frac{\partial v_i}{\partial \kappa}(1 - \kappa\delta) + \mu_v\delta - u_i\delta \tag{B.8}$$

Again,

$$\frac{\partial v_{i+1}}{\partial v_0} = \frac{\partial v_i}{\partial v_0}(1 - \kappa\delta) \tag{B.9}$$

$$\frac{\partial v_i}{\partial \mu_v} = i\kappa\delta \tag{B.10}$$

At last,

$$\frac{\partial \log p(y_{1:n}|v_{0:n}, \theta_y)}{\partial \mu_u} = \sum_{k=1}^{n} \frac{[y_k - y_{k-1} - \sum_{i=j_{t_{k-1}}}^{j_{t_k}-1} (\mu_u - \sigma_u(v_i)^2/2)\delta_t]}{\sum_{i=j_{t_{k-1}}}^{j_{t_k}} \sigma_v(v_i)^2 \delta_t} \tag{B.11}$$

Bibliography

Aguiar, M., Ballesteros, S., Kooi, B. W., and Stollenwerk, N. (2011). The role of seasonality and import in a minimalistic multi-strain dengue model capturing differences between primary and secondary infections: complex dynamics and its implications for data analysis. *Journal of Theoretical Biology*, 289:181–196.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR*, pages 267–281.

Anderson, R., May, R., and Anderson, B. (1992). *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library.

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.

Arita, I., Nakane, M., and Fenner, F. (2006). Is polio eradication realistic? *Science*, 312(5775):852–854.

Baguelin, M., Hoek, A., Jit, M., Flasche, S., White, P., and Edmunds, W. (2010). Vaccination against pandemic influenza A/H1N1v in england: a real-time economic evaluation. *Vaccine*, 28(12):2370–2384.

Baguelin, M., Jit, M., Miller, E., and Edmunds, W. (2012). Health and economic impact of the seasonal influenza vaccination programme in england. *Vaccine*.

Ballegooijen, W. V. and Boerlijst, M. (2004). Emergent trade-offs and selection for outbreak frequency in spatial epidemics. *Proceedings of the National Academy of Sciences*, 101(52):18246–18250.

Banerjee, A. V. and Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics*, 200:151–178.

Beskos, A., Kalogeropoulos, K., and Pazos, E. (2012). Advanced mcmc methods for sampling on diffusion pathspace. *arXiv preprint arXiv:1203.6216*.

Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.

Beskos, A., Pinski, F., Sanz-Serna, J., and Stuart, A. (2011). Hybrid monte carlo on hilbert spaces. *Stochastic Processes and their Applications*, 121(10):2201–2230.

Beskos, A., Roberts, G., Stuart, A., and Voss, J. (2008). Mcmc methods for diffusion bridges. *Stochastics and Dynamics*, 8(03):319–350.

Beutels, P., Edmunds, W., Antonanzas, F., Wit, G. D., Evans, D., Feilden, R., Fendrick, A., Ginsberg, G., Glick, H., Mast, E., et al. (2002). Economic evaluation of vaccination programmes: a consensus statement focusing on viral hepatitis. *Pharmacoeconomics*, 20(1):1–7.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654.

Boelle, P., Ansart, S., Cori, A., and Valleron, A. (2011). Transmission parameters of the a/h1n1 (2009) influenza virus pandemic: a review. *Influenza and other respiratory viruses*, 5(5):306–316.

Boily, M., Abu-Raddad, L., Desai, K., Masse, B., Self, S., and Anderson, R. (2008). Measuring the public-health impact of candidate hiv vaccines as part of the licensing process. *The Lancet Infectious Diseases*, 8(3):200–207.

Boily, M., Lowndes, C., Vickerman, P., Kumaranayake, L., Blanchard, J., Moses, S., Ramesh, B., Pickles, M., Watts, C., Washington, R., et al. (2007). Evaluating large-scale hiv prevention interventions: study design for an integrated mathematical modelling approach. *Sexually transmitted infections*, 83(7):582.

Boily, M., Mâsse, B., Alsallaq, R., Padian, N., Eaton, J., Vesga, J., and Hallett, T. (2012). Hiv treatment as prevention: considerations in the design, conduct, and analysis of cluster randomized controlled trials of combination hiv prevention. *PLoS Medicine*, 9(7):e1001250.

Boily, M.-C., Lowndes, C., and Alary, M. (2002). The impact of hiv epidemic phases on the effectiveness of core group interventions: insights from mathematical models. *Sexually transmitted infections*, 78(suppl 1):i78.

Bolker, B. and Grenfell, B. (1993). Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 251(1330):75–81.

Bonds, M., Keenan, D., Rohani, P., and Sachs, J. (2010). Poverty trap formed by the ecology of infectious diseases. *Proceedings of the Royal Society B: Biological Sciences*, 277(1685):1185–1192.

Bradley, J., Moses, S., Blanchard, J., Rajaram, S., Ramesh, B., Verma, S., and Alary, M. (2010). Assessing reported condom use among female sex workers in southern india through examination of condom availability. *Sexually transmitted infections*, 86(Suppl 1):i44–i48.

Breman, J. and Arita, I. (1980). The confirmation and maintenance of smallpox eradication. *New England Journal of Medicine*, 303(22):1263–1273.

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.

Brisson, M. and Edmunds, W. (2006). Impact of model, methodological, and parameter uncertainty in the economic analysis of vaccination programs. *Medical decision making*, 26(5):434–446.

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

Carrat, F., Vergu, E., Ferguson, N., Lemaitre, M., Cauchemez, S., Leach, S., and Valleron, A. (2008). Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American journal of epidemiology*, 167(7):775.

Cauchemez, S., Bhattarai, A., Marchbanks, T., Fagan, R., Ostroff, S., Ferguson, N., Swerdlow, D., Sodha, S., Moll, M., Angulo, F., et al. (2011). Role of social networks in shaping disease transmission during a community outbreak of 2009 h1n1 pandemic influenza. *Proceedings of the National Academy of Sciences*, 108(7):2825.

Cauchemez, S., Boelle, P., Thomas, G., and Valleron, A. (2006). Estimating in real time the efficacy of measures to control emerging communicable diseases. *American journal of epidemiology*, 164(6):591.

Cauchemez, S. and Ferguson, N. (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in london. *Journal of the Royal Society Interface*, 5(25):885.

Cauchemez, S., Valleron., A., Boëlle, P., Flahault, A., and Ferguson, N. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452(7188):750–754.

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.

Cazelles, B., Chavez, M., McMichael, A., Hales, S., et al. (2005). Nonstationary influence of ei nino on the synchronous dengue epidemics in thailand. *PLoS Medicine*, 2(4):313.

Chib, S., Pitt, M., and Shephard, N. (2006). Likelihood based inference for diffusion driven state space models. *Working Paper*, 4(12):155.

Choi, Y., Jit, M., Gay, N., Cox, A., Garnett, G., and Edmunds, W. (2010). Transmission dynamic modelling of the impact of human papillomavirus vaccination in the united kingdom. *Vaccine*, 28(24):4091–4102.

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

Chronopoulou, A. and Viens, F. (2012). Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8(2):379–403.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323.

Cooley, J. and Tukey, J. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, pages 297–301.

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.

Dargatz, C. (2007). A diffusion approximation for an epidemic model. Technical report, Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München.

Daum, F. and Huang, J. (2003). Curse of dimensionality and particle filters. *Working paper*.

Davies, R. and Harte, D. (1987). Tests for hurst effect. *Biometrika*, 74(1):95–101.

De Angelis, D., Gilks, W., and Day, N. (1998). Bayesian projection of the acquired immune deficiency syndrome epidemic. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(4):449–498.

Deering, K., Vickerman, P., Moses, S., Ramesh, B., Blanchard, J., and Boily, M. (2008). The impact of out-migrants and out-migration on the hiv/aids epidemic: a case study from south-west india. *AIDS*, 22:S165.

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

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

Demiris, N. and O'Neill, P. (2005). Bayesian inference for stochastic multitype epidemics in structured populations via random graphs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):731–745.

Ding, Z., Granger, C., and Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106.

Doob, J. (1942). Topics in the theory of markoff chains. *Transactions of the American Mathematical Society*, pages 37–64.

Doob, J. (1945). Markoff chainsÑdenumerable case. *Trans. Amer. Math. Soc*, 58(3):455–473.

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

Drummond, M. and Jefferson, T. (1996). Guidelines for authors and peer reviewers of economic submissions to the bmj. *Bmj*, 313(7052):275–283.

Duane, S., Kennedy, A., Pendleton, B., and Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, 195(2):216–222.

Dukic, V., Lopes, H., and Polson, N. (2009). Tracking flu epidemics using google flu trends and particle learning. *Bla*, Bla.

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

Fenner, F., Henderson, D., Arita, I., Jezek, Z., Ladnyi, I., et al. (1988). Smallpox and its eradication. *technical report*.

Ferguson, N. (2007). Capturing human behaviour. *Nature*, 446(7137):733.

Fine, P. and Clarkson, J. (1982). Measles in england and wales–i: An analysis of factors underlying seasonal patterns. *International Journal of Epidemiology*, 11(1):5.

Finkenstadt, B., Morton, A., and Rand, D. (2005). Modelling antigenic drift in weekly flu incidence. *Statistics in medicine*, 24(22):3447–3461.

Funk, S., Salathe, M., and Jansen, V. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of The Royal Society Interface*, 7(50):1247–1256.

Garcia, O. (1983). A stochastic differential equation model for the height growth of forest stands. *Biometrics*, 39(4):1059–1072.

Garnett, G. (2002). An introduction to mathematical models in sexually transmitted disease epidemiology. *Sexually transmitted infections*, 78(1):7–12.

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

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

Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361.

Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M., and Brilliant, L. (2008). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014.

Girolami, M. and Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214.

Golightly, A. and Wilkinson, D. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, 52(3):1674–1693.

Granich, R., Gilks, C., Dye, C., Cock, K. D., and Williams, B. (2009). Universal voluntary hiv testing with immediate antiretroviral therapy as a strategy for elimination of hiv transmission: a mathematical model. *The Lancet*, 373(9657):48–57.

Grenfell, B., Bjornstad, O., and Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723.

Griffin, J., Garske, T., Ghani, A., and Clarke, P. (2011). Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks. *Biostatistics*, 12(2):303.

Gupta, S., Ferguson, N., and Anderson, R. (1998). Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science*, 280(5365):912–915.

Hanck, S., Blankenship, K., Irwin, K., West, B., and Kershaw, T. (2008). Assessment of self-reported sexual behavior and condom use among female sex workers in india using a polling box approach: a preliminary report. *Sexually transmitted diseases*, 35(5):489.

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

Hay, A., Gregory, V., Douglas, A., and Lin, Y. (2001). The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci*, 356(1416):1861–70.

He, D., Dushoff, J., Day, T., Ma., J., and Earn, D. (2011). Mechanistic modelling of the three waves of the 1918 influenza pandemic. *Theoretical Ecology*, 1:1–6.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343.

Hilleman, M. (2002). Realities and enigmas of human viral influenza: pathogenesis, epidemiology and control. *Vaccine*, 20(25):3068–3087.

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

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

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

Jones, K., Patel, N., Levy, M., Storeygard, A., Balk, D., Gittleman, J., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993.

Kalogeropoulos, K. (2007). Likelihood-based inference for a class of multivariate diffusions with unobserved paths. *Journal of Statistical Planning and Inference*, 137(10):3092 – 3102. Special Issue: Bayesian Inference for Stochastic Processes.

Kalogeropoulos, K., Dellaportas, P., and Roberts, G. (2011). Likelihood-based inference for correlated diffusions. *Canadian journal of statistics*, 39(1):52–72.

Kew, O. (2012). Reaching the last one per cent: progress and challenges in global polio eradication. *Current Opinion in Virology*.

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.

Kumar, R., Mehendale, S., Panda, S., Venkatesh, S., Lakshmi, P., Kaur, M., Prinja, S., Singh, T., Virdi, N., Bahuguna, P., et al. (2011). Impact of targeted interventions on heterosexual transmission of hiv in india. *BMC public health*, 11(1):549.

Kurtz, T. (1981). *Approximation of population processes*, volume 1. Society for Industrial Mathematics.

Lei, Y. and Zhang, S. (2004). Features and partial derivatives of bertalanffy-richards growth model in forestry. *Nonlinear Analysis: Modelling and Control*, 9(1):65–73.

Lessne, G. and Hanumara, R. (1988). Growth curve analysis in marketing: Method and application. *Journal of Marketing Research*, 25(4):391–396.

Lipsitch, M. and Viboud, C. (2009). Influenza seasonality: lifting the fog. *Proceedings of the National Academy of Sciences*, 106(10):3645–3646.

Lobato, I. and Savin, N. (1998). Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics*, 16(3):261–268.

Lowndes, C., Alary, M., Verma, S., Demers, E., Bradley, J., Jayachandran, A., Ramesh, B., Moses, S., Adhikary, R., and Mainkar, M. (2010). Assessment of intervention outcome in the absence of baseline data: "reconstruction" of condom use time trends using retrospective analysis of survey data. *Sexually transmitted infections*, 86(Suppl 1):i49.

Lunn, D. (2004). WinBUGS Differential Interface - Worked Examples.

McMichael, A., Woodruff, R., and Hales, S. (2006). Climate change and human health: present and future risks. *The Lancet*, 367(9513):859–869.

Meehl, G. (1987). The annual cycle and interannual variability in the tropical pacific and indian ocean regions. *Monthly Weather Review*, 115(1):27–50.

Melegaro, A., Jit, M., Gay, N., Zagheni, E., and Edmunds, W. (2011). What types of contacts are important for the spread of infections? using contact survey data to explore european mixing patterns. *Epidemics*, 3(3):143–151.

Merwe, R. V. D., Doucet, A., Freitas, N. D., and Wan, E. (2001). The unscented particle filter. *Advances in Neural Information Processing Systems*, pages 584–590.

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

Miller, E., Hoschler, K., Hardelid, P., Stanford, E., Andrews, N., and Zambon, M. (2010). Incidence of 2009 pandemic influenza a h1n1 infection in england: a cross-sectional serological study. *The Lancet*, 375(9720):1100–1108.

Neal, R. (2010). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 54:113–162.

Øksendal, B. (2003). *Stochastic differential equations: an introduction with applications*. Springer Verlag.

Omran, A. (1971). The epidemiologic transition: a theory of the epidemiology of population change. *The Milbank Memorial Fund Quarterly*, 49(4):509–538.

Pickles, M., Foss, A., Vickerman, P., Deering, K., Verma, S., Demers, E., Washington, R., andS. Moses, B. R., Blanchard, J., et al. (2010). Interim modelling analysis to validate reported increases in condom use and assess hiv infections averted among female sex workers and clients in southern india following a targeted hiv prevention programme. *British Medical Journal*, 86(Suppl 1):i33.

Pokern, Y., Stuart, A., and Wiberg, P. (2009). Parameter estimation for partially observed hypoelliptic diffusions. *Journal of the Royal Statistical Society: Series B*, 71(1):49.

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

Raghavendra, K., Barik, T., Reddy, B., Sharma, P., and Dash, A. (2011). Malaria vector control: from past to future. *Parasitology research*, 108(4):757–779.

Rao, T. R. (1967). Distribution, density and seasonal prevalence of aedes aegypti in the indian subcontinent and south-east asia. *Bulletin of the World Health Organization*, 36(4):547.

Rasmussen, D., Ratmann, O., and Koelle, K. (2011). Inference for nonlinear epidemiological models using genealogies and time series. *PLoS computational biology*, 7(8):e1002136.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.

Read, J. and Keeling, M. (2006). Disease evolution across a range of spatio-temporal scales. *Theoretical population biology*, 70(2):201–213.

Richards, F. (1959). A flexible growth function for empirical use. *Journal of experimental Botany*, 10(2):290–301.

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

Roberts, G. and Rosenthal, J. (1998). Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268.

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

Roberts, G. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603.

Roberts, G. and Stramer, O. (2002). Langevin diffusions and metropolis-hastings algorithms. *Methodology and computing in applied probability*, 4(4):337–357.

Rota, P. and Bellini, W. (2003). Update on the global distribution of genotypes of wild type measles viruses. *Journal of Infectious Diseases*, 187(Supplement 1):S270–S276.

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

Särkkä, S. (2007). On unscented kalman filtering for state estimation of continuous-time nonlinear systems. *IEEE Transactions on Automatic Control*, 52(9):1631–1641.

Särkkä, S. and Sottinen, T. (2008). Application of girsanov theorem to particle filtering of discretely observed continuous-time non-linear systems. *Bayesian Analysis*, 3(3):555–584.

Schultz, G. (1993). Seasonal abundance of dengue vectors in manila, republic of the philippines. *Southeast Asian Journal of Tropical Medicine and Public Health*, 24(2):369–375.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Shah, N., Wright, A., Bai, G., Barrera, L., Boulahbal, F., Martin-Casabona, N., Drobniewski, F., Gilpin, C., Havelkova, M., Lepe, R., et al. (2007). Worldwide emergence of extensively drug-resistant tuberculosis. *Emerging infectious diseases*, 13(3):380.

Shaman, J. and Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences*, 106(9):3243.

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

Stark, J., Cummings, D., Ermentrout, B., Ostroff, S., Sharma, R., Stebbins, S., Burke, D., and Wisniewski, S. (2012). Local variations in spatial synchrony of influenza epidemics. *PloS one*, 7(8):e43528.

Stollenwerk, N., Aguiar, M., Ballesteros, S., Boto, J., Kooi, B., and Mateus, L. (2012). Dynamic noise, chaos and parameter estimation in population biology. *Interface Focus*, 2(2):156–169.

Suhrcke, M., Stuckler, D., Suk, J., Desai, M., Senek, M., McKee, M., Tsolova, S., Basu, S., Abubakar, I., Hunter, P., et al. (2011). The impact of economic crises on communicable disease transmission and control: a systematic review of the evidence. *PloS one*, 6(6):e20724.

Turner, C. and Miller, H. (1997). Zenilman's anomaly reconsidered: Fallible reports, ceteris paribus, and other hypotheses. *Sexually transmitted diseases*, 24(9):522.

UNAIDS (2000). Costing guidelines for hiv prevention strategies. *official communication*.

UNAIDS (2009). *AIDS epidemic update*. WHO Regional Office Europe.

Unkel, S., Farrington, C., Garthwaite, P., Robertson, C., and Andrews, N. (2012). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A*, 175(1):49–82.

Van Der Merwe, R. and Wan, E. A. (2001). The square-root unscented kalman filter for state and parameter-estimation. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 6, pages 3461–3464. IEEE.

Viboud, C., Nelson, M., Tan, Y., and Holmes, E. (2013). Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1614).

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

Vickerman, P., Foss, A., Pickles, M., Deering, K., Verma, S., Demers, E., Lowndes, C., Moses, S., Alary, M., and Boily, M. (2010). To what extent is the hiv epidemic in southern india driven by commercial sex? a modelling analysis. *AIDS*, 24(16):2563.

Wahba, G. (1990). *Spline models for observational data*. Society for Industrial Mathematics.

Wallinga, J., Teunis, P., and Kretzschmar, M. (2006). Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American journal of epidemiology*, 164(10):936–944.

WHO (2012a). Polio case count.

WHO (2012b). World health statisics.

Wood, A. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Comput. Graph. Statist.*, 3(4):409–432.

Yuancai, L., Marques, C., and Macedo, F. (1997). Comparison of schnute's and bertalanffy-richards' growth functions. *Forest Ecology and management*, 96(3):283–288.

Zenilman, J., Weisman, C., Rompalo, A., Ellish, N., Upchurch, D., Hook 3rd, E., Celentano, D., et al. (1995). Condom use to prevent incident stds: the validity of self-reported condom use. *Sexually transmitted diseases*, 22(1):15.

Zwietering, M., Jongenburger, I., Rombouts, F., and Van't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6):1875–1881.