Bayesian Inference for Stochastic Differential Mixed-effects Models

Gavin Andrew Whitaker

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics & Statistics
Newcastle University
Newcastle upon Tyne
United Kingdom

July 2016

Abstract

Stochastic differential equations (SDEs) provide a natural framework for modelling intrinsic stochasticity inherent in many continuous-time physical processes. When such processes are observed in multiple individuals or experimental units, SDE driven mixed-effects models allow the quantification of both between and within individual variation. Performing Bayesian inference for such models, using discrete-time data that may be incomplete and subject to measurement error, is a challenging problem and is the focus of this thesis.

Since, in general, no closed form expression exists for the transition densities of the SDE of interest, a widely adopted solution works with the Euler-Maruyama approximation, by replacing the intractable transition densities with Gaussian approximations. These approximations can be made arbitrarily accurate by introducing intermediate time-points between observations. Integrating over the uncertainty associated with the process at these time-points necessitates the use of computationally intensive algorithms such as Markov chain Monte Carlo (MCMC).

We extend a recently proposed MCMC scheme to include the SDE driven mixed-effects framework. Key to the development of an efficient inference scheme is the ability to generate discrete-time realisations of the latent process between observation times. Such realisations are typically termed diffusion bridges. By partitioning the SDE into two parts, one that accounts for nonlinear dynamics in a deterministic way, and another as a residual stochastic process, we develop a class of novel constructs that bridge the residual process via a linear approximation. In addition, we adapt a recently proposed construct to a partial and noisy observation regime. We compare the performance of each new construct with a number of existing approaches, using three applications: a simple birth-death process, a Lotka-Volterra model and a model for aphid growth.

We incorporate the best performing bridge construct within an MCMC scheme to determine the posterior distribution of the model parameters. This methodology is then applied to synthetic data generated from a simple SDE model of orange tree growth, and real data consisting of observations on aphid numbers recorded under a variety of different treatment regimes. Finally, we provide a systematic comparison of our approach with an inference scheme based on a tractable approximation of the SDE, that is, the linear noise approximation.

Acknowledgements

I am extremely grateful to my supervisors, Andrew Golightly and Richard Boys, for their generous advice, unwavering support and unbounded enthusiasm, without which this thesis would not have come into fruition. For their fondness of drunken charades, impromptu Jägerbombs and for teaching me the most important lesson: that there is no H in the word 'assume', but there are 8 R's in 'bastard'. Thank you.

I owe an immeasurable debt to my mum, dad, grandad and the rest of my family for all the love, kindness and support (both emotionally and financially) they have given me. For the numerous times I have bored them as they enthusiastically ask what I'm doing, and for convincing me to pursue this deranged venture in the first place.

Special thanks must go to Jack, Gemma, Ben Ostell, Dan, Hep, Stevie, Ben Schwarz, Sam, Chief, Duffy, Carl, Ash, Scott, Lord Tegglington III, Becca and the rest of the NUTS lot for all the drama, laughs, pub times, unpronounceable play titles, terrible karaoke and multiple hangings of Nige. Bully!

To Patrick, Sam, Nadia, Cheryl, Rachael, Billy, Gabby, Charlie, Chris and Martin for constantly making me laugh, helping to brighten the darker times and for making it clear, that whatever I do in life, I won't escape a euphemistic song about my 'corn on the cob'.

My thanks go to Adam, Jay, Craig, Dave, Norm, Gimp, Dowie, Phant, Mowgli and Dan for providing the laughs whenever I went back home. To Stevie - who ensured I looked a much better golfer than I actually am. To Hossack, with whom I created a comedy group that neither of us can remember the name of, and a big bold wave to little Laura for her friendship, and for the kitchen incident of 67 - tremendous, terrific, unbelievable.

Finally, to those friends and colleagues in the School of Mathematics and Statistics who have made my time there so enjoyable - especially Ste, Jowett, David 'I drink gin now' Robertson, Ashleigh, Keith, Rob, Matt and Fred. To Cushing - for having the mind of a six year old, and George - for making me appreciate the weight of water.

If you do not feature in this list, you may still have played an important role - you just should have done more to leave a lasting impression.

Declaration

Parts of this thesis have been published by the author:

- Chapter 3 has previously been published as: Whitaker, G. A., Golightly, A., Boys, R. J. and Sherlock, C. 'Improved bridge constructs for stochastic differential equations', *Statistics and Computing*. To appear, (2016).
- Parts of Chapter 4 and the whole of Chapter 5 has previously been published as: Whitaker, G. A., Golightly, A., Boys, R. J. and Sherlock, C. 'Bayesian inference for diffusion driven mixed-effects models', *Bayesian Analysis*. To appear, (2016).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Throughout history, interest has lain in understanding and modelling the dynamics of systems evolving through time. Instances include (but are not limited to) the growth of populations, the interactions between certain species, the spread of epidemics and more recently, intra-cellular processes. Initially the dynamics of these systems were captured through the use of ordinary differential equations (ODEs); for example, Kermack and McKendrick (1927) describe the spread of a disease through a population using three ODEs. These three ODEs model the changes in the number of individuals who are Susceptible (those who could catch the disease), Infectious (those who have the disease) and Recovered (those who no longer have the disease). This model is known as the SIR model. However, the evolution of these systems is not entirely predictable and is subject to random variation. The deterministic nature of the ODE description is unable to capture this random variation and so has proved to be an unsatisfactory means through which to capture the true dynamics of such systems. Hence an alternative modelling framework is required, which can account for random behaviour.

A system where the introduction of randomness appears fundamental is the stock market, specifically the pricing of options and shares. Black and Scholes (1973) and Merton (1973) developed a framework for the fair pricing of options. Integral to their work was the idea of stochastic differential equations (SDEs). SDEs consist of both a deterministic and stochastic part, and capture the dynamics of a system through a solution which fluctuates around the deterministic solution. However, it should be noted that the mean of the stochastic solution is not the ODE solution. Some application areas and indicative references where SDEs have been used include finance (Cox et al., 1985; Bibby and Sørensen, 2001; Chiarella et al., 2009; Kalogeropoulos et al., 2010; Stramer et al., 2010), systems biology (Golightly and Wilkinson, 2005, 2006, 2008; Finkenstädt et al., 2008; Komorowski et al., 2009, 2010; Fuchs, 2013; Fearnhead et al., 2014; Golightly et al., 2015), population

dynamics (Gilioli et al., 2008; Heydari et al., 2014), physics (van Kampen, 1981; Ramshaw, 1985; Tuckwell, 1987), medicine (Walsh, 1981; Fogelson, 1984; Capasso and Morale, 2009), epidemics (Dargatz, 2007; Allen, 2008; Gray et al., 2011), biology (Leung, 1985), epidemiology (Barbour, 1974; Chen and Bokka, 2005; Alonso et al., 2007), genetics (Fearnhead, 2006; Tian et al., 2007) and traffic control (McNeil, 1973). The solution of an SDE gives a continuous-time, continuous-valued stochastic process typically referred to as a diffusion process.

Unfortunately, analytic intractability of SDEs governing most nonlinear multivariate diffusion processes precludes a closed-form expression for the transition densities. Consequently, inferring the parameters of the SDE using observations taken at discrete times is a challenging problem. Methods to overcome this difficulty include simulated maximum likelihood estimation (Pedersen, 1995; Durham and Gallant, 2002), closed-form expansion of the transition density (Aït-Sahalia, 2002, 2008; Picchini et al., 2010; Stramer et al., 2010; Picchini and Ditlevsen, 2011), exact simulation approaches (Beskos et al., 2006, 2009, 2013; Sermaidis et al., 2013) and Bayesian imputation approaches (Elerian et al., 2001; Eraker, 2001; Roberts and Stramer, 2001; Golightly and Wilkinson, 2008; Stramer and Bognar, 2011; Kou et al., 2012; Schauer et al., 2016). The latter method replaces an intractable transition density with a first order Euler-Maruyama approximation, and uses data augmentation to limit the discretisation error incurred by the approximation. Whilst exact algorithms that avoid discretisation error are appealing, they are limited to diffusions which can be transformed to have unit diffusion coefficient, known as reducible diffusions. On the other hand, the Bayesian imputation approach has received much attention in the recent literature due to its wide applicability.

The essential idea of the Bayesian imputation approach is to augment low frequency data by introducing intermediate time-points between observation times. An Euler-Maruyama scheme is then applied by approximating the transition densities over the induced discretisation as Gaussian. Computationally intensive algorithms, such as Markov chain Monte Carlo (MCMC), are then used to integrate over the uncertainty associated with the missing/unobserved data. Care must be taken in the design of such schemes due to

1. dependence between the parameters and the latent process;

2. dependence between values of the latent process itself.

The former was first highlighted as a problem by Roberts and Stramer (2001). Techniques to overcome this issue include the use of a reparameterisation (Roberts and Stramer, 2001; Golightly and Wilkinson, 2008, 2010) and a particle MCMC method which permits a joint update of the parameters and latent values (Golightly and Wilkinson, 2011; Picchini, 2014). Overcoming dependence between values of the latent process can be accomplished

by jointly updating latent values in blocks. This requires the ability to sample the diffusion process (or an approximation thereof) at intermediate times between two fixed values. The resulting realisation is typically referred to as a *diffusion bridge*.

Designing *diffusion bridge* constructs that can be applied in general multivariate settings is a challenging problem and has received much attention in recent literature. The simplest approach (see, for example, Pedersen (1995)) is based on the forward dynamics of the diffusion process and generates a bridge by sampling recursively from the Euler-Maruyama approximation of the unconditioned SDE. This myopic approach induces a discontinuity as the discretisation gets finer, and is well known to lead to low Metropolis-Hastings acceptance rates. The modified diffusion bridge (MDB) construct of Durham and Gallant (2002) (and the extensions to the partial and noisy observation case in Golightly and Wilkinson (2008)) pushes the bridge process towards the observation in a linear way and provides the optimal sampling method when the drift and diffusion coefficients of the SDE are constant (Stramer and Yan, 2006). However, this construct does not produce efficient proposals when the process exhibits nonlinear dynamics. Several approaches have been proposed to overcome this problem. For example, Lindström (2012) (see also Fearnhead (2008) for a similar approach) combines the myopic and MDB approaches, with a tuning parameter governing the precise dynamics of the resulting sampler. Del Moral and Murray (2015) (see also Lin et al. (2010)) use a sequential Monte Carlo scheme to generate realisations according to the forward dynamics, pushing the resulting trajectories towards the observation using a sequence of reweighting steps. Schauer et al. (2016) combine the ideas of Delyon and Hu (2006) and Clark (1990) to obtain a bridge based on the addition of a guiding term to the drift of the process under consideration. The guiding term is derived using a tractable approximation of the target process. Methods that generate continuous sample paths exactly have been proposed by Beskos et al. (2006) for reducible diffusions. Beskos et al. (2013) use Hybrid Monte Carlo (HMC) on pathspace to generate SDE sample paths under various observation regimes. For the applications considered, the authors found reasonable gains in overall efficiency (as measured by minimum effective sample size per CPU time) over an independence sampler with a Brownian bridge proposal. However, we note that HMC also requires careful choice of the tuning parameters (namely the number of steps (and their size) in the leapfrog integrator) to maximise efficiency.

When repeated measurements on a system of interest are made, differences between individuals or experimental units can be incorporated through random effects. Quantification of both system (intrinsic) variation and variation between units leads to a stochastic differential mixed-effects model (SDMEM), the natural extension of an SDE. Difficulty in performing inference for SDEs has resulted in relatively little work on SDMEMs.

Picchini et al. (2010) propose a procedure for obtaining approximate maximum likelihood estimates for SDMEM parameters based on a two step approach: they use a closed-form Hermite expansion (Aït-Sahalia, 2002, 2008) to approximate the transition density, before using Gaussian quadrature to numerically integrate the conditional likelihood with respect to the random parameters. As noted by Picchini and Ditlevsen (2011), the approach is, in practice, limited to a scalar random effect parameter since Gaussian quadrature is increasingly computationally inefficient as the dimension of the random effect parameter grows. The methodology is extended in Picchini and Ditlevsen (2011) to deal with multiple random effects. A number of limitations remain however. In particular a reducible diffusion process is required. Another drawback is that the method cannot account for measurement error. A promising approach appears to be the use of the extended Kalman filter (EKF) to provide a tractable approximation to the SDMEM. This has been the focus of Overgaard et al. (2005), Tornøe et al. (2005) and Berglund et al. (2011). The `R` package `PSM` (Klim et al., 2009) uses the EKF to estimate SDMEMs. Unfortunately, a quantification of the effect of using these approximate inferential models appears to be missing from the literature. Donnet et al. (2010) discuss inference for SDMEMs in a Bayesian framework, and implement a Gibbs sampler when the SDE (for each experimental unit) has an explicit solution. When no explicit solution exists they propose to approximate the diffusion process using the Euler-Maruyama approximation.

1.1 Thesis aims

The aim of this thesis is to provide a framework that permits (simulation-based) Bayesian inference for a large class of multivariate SDMEMs using discrete-time observations that may be incomplete (so that only a subset of model components are observed) and subject to measurement error. We further require our framework to accommodate processes that exhibit nonlinear dynamics between observation times, as this nonlinearity can be important when, for example, the process is observed sparsely in time.

As a starting point, we consider the Bayesian imputation approach described above. We adapt the reparameterisation technique (known as the modified innovation scheme) of Golightly and Wilkinson (2008, 2010) (see also Stramer and Bognar (2011); Fuchs (2013); Papaspiliopoulos et al. (2013)) to the SDMEM framework. A key requirement of the scheme is the ability to sample the latent process between two fixed values. Previous implementations of the modified innovation scheme have typically focused on the modified diffusion bridge construct of Durham and Gallant (2002). For the SDMEM considered in Section 5.2 we find that this construct fails to capture the nonlinear dynamics exhibited between observation times. We therefore develop a novel class of bridge constructs that

are computationally and statistically efficient, simple to implement, can be applied in scenarios where only partial and noisy measurements of the system are available and can capture nonlinear behaviour.

Finally, we provide a systematic comparison of our approach with an inference scheme based on a linear noise approximation (LNA) of the SDE. The LNA approximates transition densities as Gaussian, and when combined with Gaussian measurement error, allows the latent process to be integrated out analytically. Essentially a forward (Kalman) filter can be implemented to calculate the marginal likelihood of all parameter values of interest, facilitating a marginal Metropolis-Hastings scheme targeting the marginal parameter posterior of interest. It should be noted, however, that evaluation of the Gaussian transition densities under the LNA requires the solution of an ODE system whose order grows quadratically with the number of components (say d) governed by the SDE. The computational efficiency of an LNA based inference scheme will therefore depend on d, and on whether or not the ODE system can be solved analytically.

1.2 Outline of thesis

In the following we outline the subsequent chapters contained within this thesis. Chapter 2 introduces the concept of a diffusion process as the solution to an Itô SDE, including a specific example of Brownian motion. Brief but fundamental concepts of Itô calculus are discussed and generalisations to multivariate processes are considered. A short introduction to Bayesian inference and a review of Markov chain Monte Carlo is given. Such techniques are incredibly useful as they allow for a random sample to be drawn from a density of interest, which needs only be known up to a constant of proportionality. The chapter concludes by examining a tractable approximation to an SDE, that is, the linear noise approximation.

In Chapter 3 we discuss the challenging problem of constructing a *diffusion bridge* for a multivariate diffusion. We initially review bridge constructs from existing literature, before detailing our novel approach which aims to bridge the latent process by bridging a conditioned residual process. We also offer extensions to the recently proposed guided proposal of Schauer et al. (2016). We complete the chapter with three examples of increasing difficulty, designed to gauge the statistical efficiency (and demonstrate the associated properties) of each *diffusion bridge*. The three examples are, a simple birth-death process, a Lotka-Volterra model and a model for aphid growth.

Stochastic differential mixed-effects models are presented in Chapter 4 and a Bayesian inference scheme for SDMEMs is outlined. Each step is outlined in detail and the problems surrounding parameter inference are considered. The primary issue to overcome is

intolerable mixing, due to the dependence between the parameters and latent values. We discuss solutions to this problem, including the modified innovation scheme. Finally within Chapter 4, we extend the LNA to SDMEMs and detail an inference scheme based on this tractable approximation.

Chapter 5 details two numerical examples. First, we consider a synthetic dataset generated from an SDMEM driven by the simple univariate model of orange tree growth presented in Picchini et al. (2010) and Picchini and Ditlevsen (2011). The ODE system governing the LNA solution is tractable in this example. Secondly, we fit a model of aphid growth to both real and synthetic data. The real data are taken from Matis et al. (2008) and consist of cotton aphid (*Aphis gossypii*) counts in the Texas High Plains obtained for three different levels of irrigation water, nitrogen fertiliser and block. This application is particularly challenging due to the nonlinear drift and diffusion coefficients governing the SDMEM, and the ability to only observe one of the model components (with error). Moreover, the ODE system governing the LNA solution is intractable and a numerical solver must be used. Finally, we compare inferences made under the SDMEM and LNA using synthetic data generated under four data-poor scenarios.

Conclusions are drawn in Chapter 6 before some areas of possible future work are discussed.

Chapter 2

Diffusion processes and tractable approximations

In this chapter we discuss diffusion processes, showing that a sample path of the process satisfies an Itô SDE. We look at aspects of stochastic calculus before giving a brief outline of Monte Carlo methods with a view to performing inference for the process governing an SDE. Furthermore we consider a tractable approximation of the SDE, namely, the linear noise approximation (LNA). The details within this chapter provide an introduction to SDEs and stochastic calculus; for a more comprehensive review we refer the reader to Arnold (1974) and Øksendal (2003).

2.1 Diffusion processes

Initially, let us consider a univariate stochastic process $\{X_t, t \geq 0\}$ which is continuous in time. Given a sequence of times $t_0 < t_1 < \cdots < t_n$, if

$$P\left(X_{t_n} \leq x' | X_{t_{n-1}} = x, X_{t_{n-2}} = x_{t_{n-2}}, \ldots, X_{t_0} = x_{t_0}\right) = P\left(X_{t_n} \leq x' | X_{t_{n-1}} = x\right),$$

then the process is a (first order) Markov process, that is, the future states of X_t are independent of the past states given the present state. For all times $0 \leq t < t' < \infty$, let us denote the transition kernel of the process by

$$P\left(t, x; t', x'\right) = P\left(X_{t'} \leq x' | X_t = x\right),$$

with $p(t, x; t', x')$ denoting the associated transition density. As X_t is a Markov process, the transition density satisfies the Chapman-Kolmogorov equation. Specifically, for times

t, t^* and t', such that $t < t^* < t'$,

$$p(t,x;t',x') = \int_{-\infty}^{\infty} p\left(t,x;t^*,x^*\right) p\left(t^*,x^*;t',x'\right) dx^*, \tag{2.1}$$

that is, the transition density at time t' is connected to the density at time t via the intermediate time t^*.

A univariate diffusion process $\{X_t\}$ with infinitesimal mean α, known as the drift, and infinitesimal variance β, known as the diffusion coefficient, is defined as a Markov process satisfying the following for all $x \in \mathbb{R}$:

$$0 = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}} p\left(t,x;t+\delta t,x'\right) \left|x'-x\right|^{\kappa} dx', \qquad \forall\ \kappa > 2, \tag{2.2}$$

$$\alpha(x,t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}} p\left(t,x;t+\delta t,x'\right) \left(x'-x\right) dx', \tag{2.3}$$

$$\beta(x,t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}} p\left(t,x;t+\delta t,x'\right) \left(x'-x\right)^2 dx'. \tag{2.4}$$

From (2.2) we have that a large jump has negligible probability over a small time interval, resulting in a sample path of the process being almost surely continuous.

The above can be represented in terms of the expectation, leading to (2.2)–(2.4) becoming

$$0 = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathrm{E}\big(\left|X_{t+\delta t} - x\right|^{\kappa} | X_t = x\big), \qquad \forall\ \kappa > 2, \tag{2.5}$$

$$\alpha(x,t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathrm{E}\left(X_{t+\delta t} - x \middle| X_t = x\right), \tag{2.6}$$

$$\beta(x,t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathrm{E}\left\{ \left(X_{t+\delta t} - x\right)^2 | X_t = x \right\}. \tag{2.7}$$

By combining (2.2)–(2.4) with the Chapman-Kolmogorov equation (2.1), we can derive the Kolmogorov differential equations for a diffusion process, known as the Kolmogorov forward and backward equations. Let us consider a diffusion process with infinitesimal mean $\alpha(x,t)$, infinitesimal variance $\beta(x,t)$ and initial condition $X_t = x_t = x$. The Kolmogorov backward equation is given by

$$-\frac{\partial p\left(t,x;t',x'\right)}{\partial t} = \alpha(x,t) \frac{\partial p\left(t,x;t',x'\right)}{\partial x} + \frac{1}{2} \beta(x,t) \frac{\partial^2 p\left(t,x;t',x'\right)}{\partial x^2}. \tag{2.8}$$

This equation describes the dynamics of the diffusion process going 'backwards' in time as it incorporates the prior position of the process (x) at time t. Similarly the Kolmogorov forward equation describes the dynamics of the diffusion process going 'forwards' in time

and is given by

$$\frac{\partial p\left(t,x;t',x'\right)}{\partial t'} = -\frac{\partial}{\partial x'}\left\{\alpha\left(x',t'\right)p\left(t,x;t',x'\right)\right\} + \frac{1}{2}\frac{\partial^2}{\partial x'^2}\left\{\beta\left(x',t'\right)p\left(t,x;t',x'\right)\right\}. \quad (2.9)$$

This equation is commonly known as the Fokker-Planck equation. The full derivations of (2.8) and (2.9) can be found in Allen (2010) and Wilkinson (2011). Plainly, for a given $\alpha(\cdot)$ and $\beta(\cdot)$, the Kolmogorov equations can be used to determine the transition density of the diffusion process. Unfortunately, for arbitrary $\alpha(\cdot)$ and $\beta(\cdot)$, this is rarely possible analytically.

2.1.1 Brownian motion

In 1827 the Scottish botanist and palaeobotanist Robert Brown discovered Brownian motion after examining pollen from a plant suspended in water under the lens of a microscope (Brown, 1828). He noted that minute particles ejected from the pollen grain displayed a continuous irregular motion. In 1900 the French mathematician Louis Bachelier considered Brownian motion as a model for stock, mathematically defining Brownian motion in the process; see Bachelier (1964). The governing laws of Brownian motion were established by Albert Einstein (Einstein, 1905). Norbert Wiener (Wiener, 1923) proved the existence (and provided the construction) of Brownian motion, and it is for this reason that Brownian motion is also referred to as the Wiener process.

The univariate stochastic process $\{W_t, t \geq 0\}$ is defined to be a standard Brownian motion if $W_t \in \mathbb{R}$ depends continuously on t and the following assumptions hold:

1. $P(W_0 = 0) = 1$;
2. For all times $0 \leq t_0 < t_1 < t_2$, $W_{t_2} - W_{t_1}$ and $W_{t_1} - W_{t_0}$ are independent;
3. For all times $0 \leq t_0 < t_1$, $W_{t_1} - W_{t_0} \sim N(0, t_1 - t_0)$.

Assumption 2 ensures that standard Brownian motion has independent increments, and so $W_{t_2} - W_{t_1}$ is independent of the past $\{W_t, t < t_1\}$. Assumption 3 establishes that standard Brownian motion has stationary increments with Gaussian distributions. Note that standard Brownian motion is a diffusion process for which $\alpha(x,t) = 0$ and $\beta(x,t) = 1$. Furthermore, using assumption 3 and that for times $0 \leq t < t' < \infty$, $W_{t'} = W_{t'} - W_t + W_t$, gives the conditional distribution $W_{t'}|W_t = x \sim N(x, t' - t)$.

It clearly follows that the transition density is given by

$$p\left(t,x;t',x'\right) = \frac{1}{\sqrt{2\pi\left(t'-t\right)}}\exp\left\{-\frac{1}{2}\frac{\left(x'-x\right)^2}{t'-t}\right\}.$$

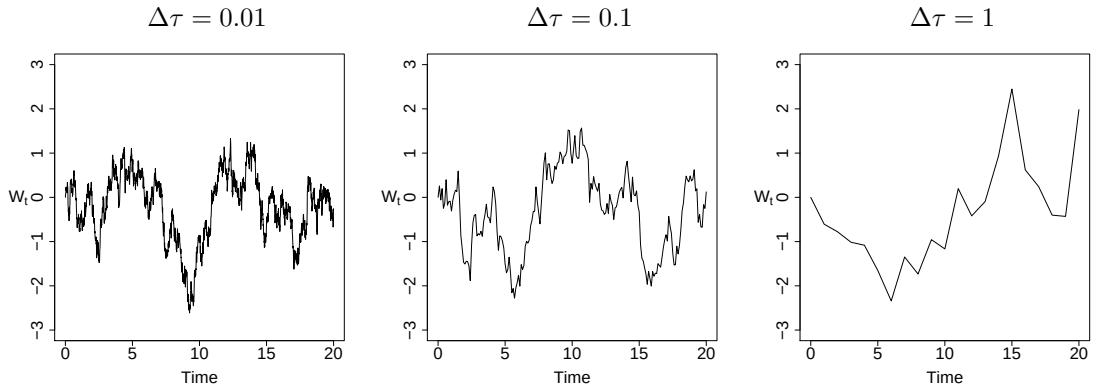

Figure 2.1: A single skeleton path of standard Brownian motion for three different time-steps.

It is straightforward to show that $p(t, x; t', x')$ satisfies the Fokker-Planck equation (2.9). Using this density we can show that standard Brownian motion satisfies (2.2)–(2.4), via (2.5)–(2.7). To see this, we begin by noting that $\Delta W_t = W_{t'} - W_t \sim N(0, t' - t)$. From this we have

$$\mathrm{E}\left(\Delta W_t\right) = 0, \quad \mathrm{E}\left(\Delta W_t^2\right) = t' - t \quad \text{and} \quad \mathrm{E}\left(\Delta W_t^4\right) = 3(t' - t)^2.$$

Dividing by $t' - t$ and taking the limit as $t' \to t$ clearly shows that (2.2), (2.3) and (2.4) are satisfied.

Although generating a full continuous-time realisation is not possible, simulating the process at discrete times is effortless. The resulting trajectory is typically referred to as a skeleton path. For an equally spaced partition of $[0, t]$ given by

$$0 = \tau_0 < \tau_1 < \cdots < \tau_{m-1} < \tau_m = t,$$

with $\Delta\tau = \tau_{i+1} - \tau_i$, $i = 0, \ldots, m - 1$, recursively sampling from

$$W_{\tau_{i+1}} | W_{\tau_i} = x_i \sim N(x_i, \Delta\tau)$$

gives a skeleton path. Figure 2.1 shows a skeleton path of standard Brownian motion for three different time-steps.

Despite the fact that the sample paths of standard Brownian motion are continuous, W_t is not differentiable almost everywhere. Thus, integrals of the form

$$\int_0^t f(s) dW_s = \int_0^t f(s) \frac{dW_s}{ds} ds$$

have no meaning in the Riemann sense. This fact necessitates a definition of a stochastic

integral. The integral we require is known as the Itô integral and is the subject of the next section.

2.2 Itô calculus

Kiyoshi Itô extended the methods of classical calculus to stochastic processes (of which Brownian motion is one), and it is after him that Itô calculus is named.

2.2.1 The Itô integral

Consider a random function $f(X_s, s)$, $s \in [0, t]$, which satisfies

$$\int_0^t \mathrm{E}\left\{f\left(X_s, s\right)^2\right\} ds < \infty.$$

For simplicity we will write the function throughout this section as $f(s)$. The Itô integral can be obtained as follows. First partition $[0, t]$ as

$$0 = \tau_0 < \tau_1 < \cdots < \tau_{m-1} < \tau_m = t,$$

with equidistant time-steps $\Delta\tau = \tau_{i+1} - \tau_i$, $i = 0, \ldots, m-1$. It is clear that $\Delta\tau \to 0$ as $m \to \infty$. Also let $\Delta W_{\tau_i} = W_{\tau_{i+1}} - W_{\tau_i}$, $i = 0, \ldots, m-1$. The Itô stochastic integral of $f(s)$ is then

$$\int_0^t f\left(s\right) dW_s = \underset{m\to\infty}{\text{l.i.m.}} \sum_{i=0}^{m-1} f\left(\tau_i\right) \Delta W_{\tau_i}, \tag{2.10}$$

where l.i.m. is the mean-square limit, that is, the stochastic integral is the mean-square limit of a sequence of partial sums. In contrast to classical calculus the value of the limit depends upon the selection of points within the partition, here the left endpoint of each sub-interval is taken. If we set

$$F_{m-1} = \sum_{i=0}^{m-1} f\left(\tau_i\right) \Delta W_{\tau_i}$$

and

$$L = \int_0^t f\left(s\right) dW_s,$$

then $\underset{m\to\infty}{\text{l.i.m.}} F_{m-1} = L$ implies that

$$\lim_{m\to\infty} \mathrm{E}\left\{(F_{m-1} - L)^2\right\} = 0.$$

The stochastic integral (2.10) is obtained by taking the left endpoint of each sub-interval. If instead we took the midpoint of each sub-interval we would arrive at the Stratonovich stochastic integral. In the context of this thesis, the Itô stochastic integral is more appropriate than the Stratonovich stochastic integral, as Stratonovich calculus does not have the same direct link to the theory of diffusion processes that Itô calculus possesses. Therefore from this point on, any stochastic integral will be assumed to be an Itô integral.

For simple functions $f(X_s)$, the Itô integral can be verified directly using (2.10). As an illustration, consider the case $f(X_s) = 1$. Thus

$$\begin{aligned}
\int_0^t dW_s &= \underset{m\to\infty}{\text{l.i.m.}} \sum_{i=0}^{m-1} \Delta W_{\tau_i} \\
&= \underset{m\to\infty}{\text{l.i.m.}} \left\{(W_{\tau_1} - W_{\tau_0}) + (W_{\tau_2} - W_{\tau_1}) + \cdots + (W_{\tau_m} - W_{\tau_{m-1}})\right\} \\
&= \underset{m\to\infty}{\text{l.i.m.}} (W_{\tau_m} - W_{\tau_0}) \\
&= \underset{m\to\infty}{\text{l.i.m.}} (W_t - W_0) \\
&= W_t - W_0 \\
&= W_t.
\end{aligned}$$

Taking a second function $g(s)$, $s \in [0, t]$, which again satisfies

$$\int_0^t \mathrm{E}\left\{g\,(s)^2\right\} ds < \infty,$$

and the times $0 \leq t^* < t$, then some of the properties of the Itô integral include

1. $\displaystyle\int_0^t f(s)dW_s = \int_0^{t^*} f(s)dW_s + \int_{t^*}^t f(s)dW_s.$
2. $\displaystyle\int_0^t \{Af(s) + g(s)\}\, dW_s = A\int_0^t f(s)dW_s + \int_0^t g(s)dW_s,$ where A is a constant.
3. $\displaystyle\int_0^t \mathrm{E}\{f(s)\}\, dW_s = 0.$
4. $\displaystyle\mathrm{E}\left[\left\{\int_0^t f(s)dW_s\right\}^2\right] = \mathrm{E}\left\{\int_0^t f(s)^2 ds\right\}.$

Property 4 is known as the Itô isometry. A sketch proof of Property 1 can be found in

Allen (2010), whilst proofs of Properties 2–4 are given in Arnold (1974). The Itô integral also has the property that it is a martingale: a process whose future expectation is equal to the current value of the process, regardless of the past.

2.2.2 Stochastic differential equations (SDEs)

An Itô process is a stochastic process $\{X_t, t \geq 0\}$, which satisfies

$$X_t = X_0 + \int_0^t \alpha(X_s, s)ds + \int_0^t \sqrt{\beta(X_s, s)}\, dW_s. \tag{2.11}$$

Here the process can be expressed as a Riemann integral plus an Itô stochastic integral. In differential form, we obtain the Itô SDE

$$dX_t = \alpha(X_t, t)\, dt + \sqrt{\beta(X_t, t)}\, dW_t, \tag{2.12}$$

where $\alpha(X_t, t)$ is the drift and $\beta(X_t, t)$ is the diffusion coefficient. Note that an SDE is reducible if it can be rewritten (via a transformation) to have a unit diffusion coefficient, that is

$$dX_t^* = \alpha^*(X_t, t)\, dt + dW_t.$$

If no such transformation is possible, then the SDE is said to be irreducible.

A pathwise unique solution $\{X_t\}$ to (2.11) exists, if the drift and diffusion coefficients ($\alpha(X_t, t)$ and $\beta(X_t, t)$ respectively) are Lipschitz continuous and the linear growth conditions hold. This requires

$$\left|\alpha\left(x,t\right) - \alpha\left(x',t'\right)\right| + \left|\sqrt{\beta\left(x,t\right)} - \sqrt{\beta\left(x',t'\right)}\right| \leq A\left|x - x'\right| \tag{2.13}$$

and

$$\left|\alpha\left(x,t\right)\right|^2 + \left|\sqrt{\beta\left(x,t\right)}\right|^2 \leq B\Big(1 + |x|^2\Big), \tag{2.14}$$

where A and B are positive constants, $x, x' \in \mathbb{R}$ and $t, t' \in [0, \infty)$ with $t < t'$. Condition (2.14) along with the addition of $\mathrm{E}(X_t)^2 < \infty$ ensures that X_t will not explode. Proofs of the above conditions (along with further details) can be found in Kloeden and Platen (1992) and Øksendal (2003).

Pathwise uniqueness implies that if there are two solutions to (2.11), denoted X_t and X_t', with the same initial condition, then

$$P\left(\sup_{t\in[0,\infty)} \left|\left|X_t - X_t'\right|\right| > 0\right) = 0.$$

The above means X_t and X'_t are equivalent. Such a pathwise unique solution, $\{X_t\}$, to (2.11) is known as a strong solution. If only the drift and diffusion coefficient are specified in advance, and it is possible to find a pair of processes $(\tilde{X}_t, \tilde{W}_t)$ such that (2.11) is satisfied, then $\{\tilde{X}_t\}$ is known as a weak solution. Naturally a strong solution is a weak solution; however the converse is not necessarily true. For further discussion of strong and weak solutions we refer the reader to Øksendal (2003) or Fuchs (2013).

2.2.3 Itô's formula

Take a diffusion process $\{X_t, t \geq 0\}$ which satisfies the SDE (2.12). Let $f(x,t)$ be a real valued function, once differentiable in t and twice differentiable in x. Let

$$f_t = \frac{\partial f}{\partial t}, \quad f_x = \frac{\partial f}{\partial x} \quad \text{and} \quad f_{xx} = \frac{\partial^2 f}{\partial x^2}$$

denote the first partial derivative of f with respect to t, and the first two with respect to x. Itô's formula then gives the SDE satisfied by the process $\{Y_t, t \geq 0\}$, where $Y_t = f(X_t, t)$ as

$$dY_t = f_t\left(X_t, t\right) dt + f_x\left(X_t, t\right) dX_t + \frac{1}{2} f_{xx}\left(X_t, t\right) \left(dX_t\right)^2 . \tag{2.15}$$

Hence, Itô's formula is a method to apply (nonlinear) transformations to SDEs. Equation (2.15) is the Itô calculus counterpart of the classical calculus chain rule. Note that the second derivative term is usually referred to as Itô's correction. When applying Itô's formula the following identities are of use:

$$dt^2 = dt\, dW_t = dW_t\, dt = 0 \quad \text{and} \quad dW_t^2 = dt.$$

2.3 Example: Ornstein-Uhlenbeck process

Consider a homogeneous diffusion process $\{X_t, t \geq 0\}$ satisfying an SDE of the form

$$dX_t = \theta_1(\theta_2 - X_t)\, dt + \theta_3\, dW_t, \quad X_0 = x_0.$$

This SDE can be solved by applying Itô's formula (2.15) with $f(x,t) = xe^{\theta_1 t}$, giving

$$dX_t e^{\theta_1 t} = \theta_1\theta_2 e^{\theta_1 t} dt + \theta_3 e^{\theta_1 t} dW_t. \tag{2.16}$$

Integrating both sides of (2.16) between 0 and t gives

$$\begin{aligned}
&\int_0^t dX_s e^{\theta_1 s} = \int_0^t \theta_1\theta_2 e^{\theta_1 s} ds + \int_0^t \theta_3 e^{\theta_1 s} dW_s \\
\implies \quad &\left[X_s e^{\theta_1 s}\right]_0^t = \left[\theta_2 e^{\theta_1 s}\right]_0^t + \theta_3 \int_0^t e^{\theta_1 s} dW_s \\
\implies \quad &X_t e^{\theta_1 t} - X_0 = \theta_2 e^{\theta_1 t} - \theta_2 + \theta_3 \int_0^t e^{\theta_1 s} dW_s \\
\implies \quad &X_t = x_0 e^{-\theta_1 t} + \theta_2\left(1 - e^{-\theta_1 t}\right) + \theta_3 e^{-\theta_1 t} \int_0^t e^{\theta_1 s} dW_s.
\end{aligned}$$

By linearity and Itô isometry we obtain

$$X_t | X_0 = x_0 \sim N\left\{x_0 e^{-\theta_1 t} + \theta_2\left(1 - e^{-\theta_1 t}\right), \frac{\theta_3^2}{2\theta_1}\left(1 - e^{-2\theta_1 t}\right)\right\}.$$

Note that, taking $t \to \infty$ gives the stationary distribution

$$X_t \sim N\left(\theta_2, \frac{\theta_3^2}{2\theta_1}\right).$$

Figure 2.2 shows ten skeleton paths of the Ornstein-Uhlenbeck process over the interval $[0, 10]$ for the parameter values $\theta_1 = 0.75$, $\theta_2 = 3$ and $\theta_3 = 0.5$. The skeleton paths are simulated with a time-step of $\Delta\tau = 0.1$, where X_0 is a random draw from a $N(0, 25)$ distribution. The 95% central region of the stationary distribution ($X_t \sim N(3, 1/6)$) is also illustrated in the figure. For this specific parameter choice we note that the skeleton paths reach the stationary distribution reasonably quickly, with all being inside the 95% region by time 4.

2.4 Generalisation to multivariate processes

Thus far we have only considered univariate processes. Naturally, many systems of interest cannot be represented by a univariate process, therefore we must extend the discussions above to multivariate processes.

Let us now consider a continuous-time d-dimensional Itô process $\{X_t, t \geq 0\}$ with $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})'$ (where $'$ denotes the transpose) and initial condition $X_0 = x_0$, governed by the SDE

$$dX_t = \alpha(X_t, t)\, dt + \sqrt{\beta(X_t, t)}\, dW_t. \tag{2.17}$$

Here, α is a d-vector of drift functions, the diffusion coefficient β is a $d \times d$ positive definite matrix with a square root representation $\sqrt{\beta}$ such that $\sqrt{\beta}\sqrt{\beta}' = \beta$ and W_t is a d-vector

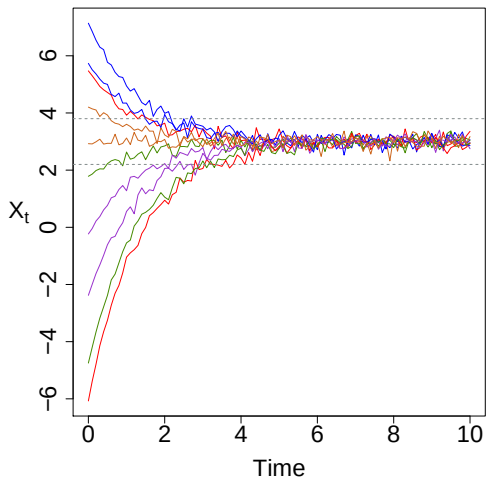

Figure 2.2: Ten skeleton paths of the Ornstein-Uhlenbeck process over $[0, 10]$ with $\Delta\tau = 0.1$. The grey dashed lines indicate the 95% region of the stationary distribution.

of (uncorrelated) standard Brownian motion processes. Equation (2.17) is the natural extension to the univariate SDE given in (2.12).

Generalising (2.6) gives the infinitesimal means

$$\alpha_i(x,t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathrm{E}\left(X_{i,t+\delta t} - x_i \middle| X_t = x\right), \qquad i = 1, \ldots, d. \tag{2.18}$$

Similarly, generalising (2.7) gives the infinitesimal second moments as

$$\beta_{i,j}(x,t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathrm{Cov}(X_{i,t+\delta t} - x_i, X_{j,t+\delta t} - x_j | X_t = x), \qquad i, j = 1, \ldots, d. \tag{2.19}$$

Extending (2.8), we arrive at the multivariate Kolmogorov backward equation

$$-\frac{\partial p\left(t, x; t', x'\right)}{\partial t} = \sum_{i=1}^{d} \alpha_i(x,t) \frac{\partial p\left(t, x; t', x'\right)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \beta_{ij}(x,t) \frac{\partial^2 p\left(t, x; t', x'\right)}{\partial x_i \partial x_j}, \tag{2.20}$$

with $p(t, x; t', x')$ being the multivariate transition density of X_t. The multivariate Kolmogorov forward equation (or multivariate Fokker-Planck equation) is given by

$$\frac{\partial p\left(t, x; t', x'\right)}{\partial t'} = -\sum_{i=1}^{d} \frac{\partial}{\partial x'_i} \left\{\alpha_i\left(x', t'\right) p\left(t, x; t', x'\right)\right\} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x'_i \partial x'_j} \left\{\beta_{ij}\left(x', t'\right) p\left(t, x; t', x'\right)\right\}. \tag{2.21}$$

A (nonlinear) transformation can be applied to (2.17) through the use of the multivariate Itô formula. Again we take $Y_t = f(X_t, t)$, where $f(x, t)$ is a real-valued function, once

differentiable in t and x_i, $i = 1, \ldots, d$. Let

$$f_{k,t} = \frac{\partial f_k}{\partial t}, \quad f_{k,x_i} = \frac{\partial f_k}{\partial x_i} \quad \text{and} \quad f_{k,x_i x_j} = \frac{\partial^2 f_k}{\partial x_i x_j}$$

denote the first partial derivative of the kth element of f with respect to t, the first with respect to x_i and the mixed derivative with respect to x_i and x_j. Thus, the kth component of $\{Y_t, t \geq 0\}$ will satisfy the SDE given by

$$dY_{k,t} = f_{k,t}\, dt + \sum_{i=1}^{d} f_{k,x_i}\, dX_{i,t} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} f_{k,x_i x_j}\, dX_{i,t} dX_{j,t}. \tag{2.22}$$

In the calculation of the above, the following identities are of use:

$$dt^2 = dt\, dW_{i,t} = dW_{i,t}\, dt = 0 \quad \text{and} \quad dW_{i,t}\, dW_{j,t} = \delta_{ij} dt,$$

where δ_{ij} is the Kronecker delta.

2.5 Bayesian inference

Consider a diffusion process $\{X_t, t \geq 0\}$ parameterised by $\theta = (\theta_1, \theta_2, \ldots, \theta_p)'$ and satisfying an SDE of the form

$$dX_t = \alpha(X_t, t, \theta)\, dt + \sqrt{\beta(X_t, t, \theta)}\, dW_t.$$

Given observations at discrete times, resulting in a dataset $D = (x_{t_0}, \ldots, x_{t_n})'$, the likelihood function is

$$L(\theta|D) = \prod_{i=0}^{n-1} p\left(x_{t_{i+1}} | x_{t_i}, \theta\right),$$

where, for notational simplicity, $p(x_{t_{i+1}}|x_{t_i}, \theta)$ denotes the transition density of $X_{t_{i+1}}|X_{t_i} = x_{t_i}$. We let the density $\pi(\theta)$ represent our prior knowledge (or beliefs) about θ. Through the use of Bayes' theorem we may update these beliefs using the data we observe. Thus, the posterior density is

$$\pi(\theta|D) = \frac{\pi(\theta) L(\theta|D)}{\int_\theta \pi(\theta) L(\theta|D) d\theta}, \tag{2.23}$$

which reflects our updated beliefs about θ after observing data D. The denominator of (2.23) can be regarded as a constant of proportionality (as it does not depend upon θ), whence

$$\pi(\theta|D) \propto \pi(\theta) L(\theta|D), \tag{2.24}$$

that is, the posterior is proportional to the product of the prior and the likelihood.

Typically, performing inference for an SDE is complicated by the intractability of the joint posterior density, $\pi(\theta|D)$. To overcome this, we appeal to standard Monte Carlo methods, and in particular Markov chain Monte Carlo, which is the subject of the next section.

2.6 Markov chain Monte Carlo (MCMC)

Markov chain Monte Carlo (MCMC) is an approach used to simulate from a specially constructed Markov chain with stationary distribution $\pi(\cdot)$. Thus, providing that the chain has converged, any value sampled will be from the density of interest $\pi(\cdot)$, here, the joint posterior density. Additionally, for a multidimensional chain, samples of each component will be simulated directly from the marginal density of the respective component. Let us assume that the distribution of interest is the posterior distribution, with density $\pi(\theta|D)$ (known as the target distribution). Here, we discuss two fundamental algorithms to construct these chains, specifically: the Metropolis-Hastings algorithm and the Gibbs sampler.

2.6.1 The Metropolis-Hastings algorithm

Metropolis et al. (1953) introduced the algorithm which was generalised by Hastings (1970), hence the name Metropolis-Hastings. Central to Metropolis-Hastings is the idea of a proposal density, denoted $q(\cdot|\cdot)$, which is some (arbitrary) transition kernel. It can be advantageous to have a proposal density which is easy to simulate from, however it need not (necessarily) have $\pi(\theta|D)$ as its stationary distribution. The Metropolis-Hastings algorithm is then as Algorithm 1.

Step 2 generates a new value of the chain from the proposal density $q(\theta^*|\theta)$, which in step 4 is either accepted (the chain moves) or rejected (the chain remains where it was). Note, $\pi(\theta|D)$ enters the acceptance probability as a ratio, and hence it is only necessary to know $\pi(\theta|D)$ up to a constant of proportionality. Therefore, by (2.24), A (in the acceptance probability of step 3) can be expressed as

$$A = \frac{\pi\left(\theta^*\right) L\left(\theta^*|D\right) q\left(\theta|\theta^*\right)}{\pi\left(\theta\right) L\left(\theta|D\right) q\left(\theta^*|\theta\right)}.$$

Given that we have complete freedom in the choice of the proposal density $q(\cdot|\cdot)$, the natural question is, 'What choices of $q(\cdot|\cdot)$ might be good, or indeed useful?' In particular, a good choice of $q(\cdot|\cdot)$ will lead to a chain which converges rapidly and mixes well; that is, it moves often and well around the support of $\pi(\theta|D)$. Below we discuss some commonly

Algorithm 1 The Metropolis-Hastings Algorithm

1. Initialise the iteration counter $i = 1$ and initialise the chain with $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right)'$, where $\theta^{(0)}$ is chosen from somewhere in the support of $\pi(\theta|D)$.
2. Propose a new value θ^* using the transition kernel $q\left(\theta^*|\theta^{(i-1)}\right)$.
3. Evaluate the acceptance probability $\min(1,A)$, where
$$A = \frac{\pi\left(\theta^*|D\right) q\left(\theta|\theta^*\right)}{\pi\left(\theta|D\right) q\left(\theta^*|\theta\right)}.$$
4. Set $\theta^{(i)} = \theta^*$ with probability $\min(1,A)$, otherwise set $\theta^{(i)} = \theta^{(i-1)}$.
5. Set $i = i + 1$ and return to step 2.

used special cases.

A symmetric proposal

Taking a proposal distribution which is symmetric gives

$$q\left(\theta^*|\theta\right) = q\left(\theta|\theta^*\right), \qquad \forall\ \theta, \theta^*.$$

In this instance, A simplifies to

$$A = \frac{\pi\left(\theta^*|D\right)}{\pi\left(\theta|D\right)},$$

that is, the acceptance probability does not depend on the proposal density.

Random walk Metropolis

It is possible to use a random walk as the proposal distribution $q(\cdot|\cdot)$ in step 2 of Algorithm 1. In this instance $q(\cdot|\cdot)$ takes the form

$$\theta^* = \theta + \omega,$$

where ω are independent identically distributed random variates known as innovations. Typically, ω has a Gaussian distribution with zero mean vector. In this instance the Metropolis-Hastings algorithm is known as a random walk sampler (or random walk Metropolis).

The variance of the random variates ω will determine the mixing of the chain; too low

a variance and the chain will explore the space slowly, as many proposed values will be accepted. Too large a variance and few proposed values will be accepted. Reflecting the correlation within θ in the covariance structure of ω is an important aspect in ensuring the chain efficiently explores the space.

If the target distribution is Gaussian, Roberts and Rosenthal (2001) suggest that the optimal acceptance probability is 0.234. Sherlock and Roberts (2009) extend this result to elliptically symmetric targets and subsequently Sherlock (2013) gives a general set of sufficient conditions for which the optimal acceptance probability is 0.234. Gelman et al. (1996), Roberts et al. (1997) and Roberts and Rosenthal (2001) suggest that the variance of ω should be given by

$$\frac{2.38^2 \ \mathrm{Var}\left(\theta|D\right)}{p},$$

where $\mathrm{Var}\left(\theta|D\right)$ is the variance matrix of the target distribution $\pi(\theta|D)$. Typically though, $\mathrm{Var}\left(\theta|D\right)$ will not be available and hence an estimate from one or more pilot runs should be used.

We note that, for large p, sampling θ^* from a multivariate Normal distribution may be expensive. In these instances, an alternative approach is to take the components of $\omega = (\omega_1, \ldots, \omega_p)'$ as independent identically distributed (univariate) Normal random variates. For example, $\omega_i \sim N(0, \sigma_i^2)$, where

$$\sigma_i^2 = \frac{2.38^2}{p}\mathrm{Var}\left(\theta_i|D\right).$$

The independence sampler

As the name suggests, an independence sampler (or independence chain) proposes a new value θ^* independently of the current value θ. Hence, $q(\theta^*|\theta) = g(\theta^*)$ for some density $g(\cdot)$. Whilst the form of such a proposal may appear to disagree with the Markovian structure of the chain, both θ and θ^* feature in the acceptance probability, meaning a proposal still depends upon the current state, and thus, the Markov property is preserved. Using such a proposal distribution leads to an acceptance probability $\min(1, A)$, where

$$A = \frac{\pi\left(\theta^*|D\right)}{\pi\left(\theta|D\right)} \Bigg/ \frac{g\left(\theta^*\right)}{g\left(\theta\right)}.$$

Clearly, we can increase the acceptance probability by making $g(\cdot)$ and $\pi(\cdot|D)$ as similar as possible. It is worth noting that in the context of an independence sampler (and in contradiction to the above on random walk Metropolis), the higher the acceptance probability, the better. Tierney (1994) suggests the avoidance of densities $g(\cdot)$ with thin tails.

Algorithm 2 Metropolis-Hastings: Componentwise Transitions

1. Initialise the iteration counter $i = 1$ and initialise the chain with $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right)'$.

2. Gain a new value $\theta^{(i)} = \left(\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_p^{(i)}\right)'$ from $\theta^{(i-1)}$ using successive generation from distributions

   $\theta_1^{(i)} \sim \pi\left(\theta_1 \middle| \theta_2^{(i-1)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right)$ using a Metropolis-Hastings step with proposal $q_1\left(\theta_1^* \middle| \theta_1^{(i-1)}\right)$

   $\theta_2^{(i)} \sim \pi\left(\theta_2 \middle| \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right)$ using a Metropolis-Hastings step with proposal $q_2\left(\theta_2^* \middle| \theta_2^{(i-1)}\right)$

   $\vdots$

   $\theta_p^{(i)} \sim \pi\left(\theta_p \middle| \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{p-1}^{(i)}, D\right)$ using a Metropolis-Hastings step with proposal $q_p\left(\theta_p^* \middle| \theta_p^{(i-1)}\right)$.

3. Set $i = i + 1$ and return to step 2.

Componentwise transitions

In practice, the construction of a suitable proposal density could be difficult. However, for many problems of interest, it may be possible to sample from the full conditional distributions for a subset of θ. Let the full conditional distribution for the ith component of θ be denoted by

$$\pi\left(\theta_i | \theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p, D\right) = \pi\left(\theta_i | \theta_{-i}, D\right), \qquad i = 1, \ldots, p.$$

The algorithm for componentwise transitions is given by Algorithm 2. Note that Algorithm 1 is, in fact, just a special case of Algorithm 2.

If the full conditional distribution for the ith component of θ is available to sample from directly, the resulting acceptance probability is one: it is for this reason that this method is also referred to as *Metropolis-within-Gibbs*. If the full conditional distributions are completely known and can be sampled from for all components of θ, we obtain the Gibbs sampler, which is presented in the next section.

Algorithm 3 The Gibbs Sampler

1. Initialise the iteration counter $i = 1$ and initialise the chain with $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right)'$.

2. Gain a new value $\theta^{(i)} = \left(\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_p^{(i)}\right)'$ from $\theta^{(i-1)}$ using successive generation from the full conditional distributions

$$\begin{aligned}
\theta_1^{(i)} &\sim \pi\left(\theta_1 \middle| \theta_2^{(i-1)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right) \\
\theta_2^{(i)} &\sim \pi\left(\theta_2 \middle| \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right) \\
&\vdots \\
\theta_p^{(i)} &\sim \pi\left(\theta_p \middle| \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{p-1}^{(i)}, D\right).
\end{aligned}$$

3. Set $i = i + 1$ and return to step 2.

2.6.2 Gibbs sampling

The Gibbs sampler (or generically Gibbs sampling) originated in the field of image processing. It was introduced by Geman and Geman (1984) before being generalised and brought to the interest of the larger statistical community by Gelfand and Smith (1990). In essence the Gibbs sampler is an MCMC scheme in which the full conditional distributions are used to form the transition kernel.

Assume that, for all components of θ, the full conditional distributions are available and can easily be sampled from. The Gibbs sampler is then given by Algorithm 3.

The chain approaches its equilibrium state as the number of iterations increases, and once the chain has converged, a value of $\theta^{(i)}$ is a sample from $\pi(\theta|D)$. Thus the Gibbs sampler is a way to sample from $\pi(\theta|D)$ when direct sampling is costly, complicated or indeed impossible, but sampling from $\pi(\theta_i|\theta_{-i}, D)$ is possible. Algorithm 3 is known as a fixed sweep Gibbs sampler. Whilst other versions of the Gibbs sampler are available, such as the random scan Gibbs sampler, the fixed sweep is simple to implement, and thus appealing. For details of other versions of the Gibbs sampler see, for example, Chapter 5 of Gamerman and Lopes (2006).

2.6.3 Blocking

Given that the components of θ can take the form of scalars, vectors or matrices, it can be useful to block certain components together in multidimensional problems. Such a strategy

is known as a block update and makes use of multivariate simulation techniques. Blocking is a strategy used to improve the convergence (and indeed mixing) of the chain, although it comes at a higher computational cost. As discussed in Gamerman and Lopes (2006) it is not the case that the larger the block update, the faster the convergence. Indeed for highly multidimensional problems a large block update is likely to be highly detrimental. Instead, components of θ should be blocked together such that the correlations between the blocks is low. Any conditionally independent components should be updated on their own (a single-block update).

2.6.4 Analysing MCMC output

As mentioned above, a Markov chain Monte Carlo scheme will only give samples from the target distribution provided convergence has been reached. It is therefore important to monitor convergence carefully and ensure convergence truly has been reached. As the number of iterations increases the distribution of the chain, $\theta^{(i)}|D$, tends to the posterior distribution $\theta|D$, and convergence is reached. Samples obtained before convergence, when the distribution of the chain is not the posterior are discarded. This number of iterations is known as the burn-in period. Viewing the trajectory of the chain via a trace plot can be used to check convergence informally. In this instance we are looking for the chain to display the same qualitative behaviour after some initial burn-in period. Gelfand and Smith (1990) (amongst others) suggest a number of informal checks for convergence. More formal checks for ensuring convergence has been reached have been proposed by, for example, Heidelberger and Welch (1983), Geweke (1992), Raftery and Lewis (1992, 1996) and Gelman (1996).

Samples of the MCMC scheme will be dependent, meaning successive draws are auto-correlated. Autocorrelation at different lag times can be observed via an autocorrelation plot. If samples are highly correlated then the chains can be thinned; this involves taking every ith iterate to ensure independence, although this comes at the computational cost of having to run the chain for longer.

Once a chain has converged, the (suitably thinned) output can be analysed. It is effortless to compute estimates of summary statistics (or standard statistical measures) such as marginal means and variances. Joint and marginal distributions can be viewed through the use of density plots (or histograms).

Unfortunately, (as mentioned previously) for most problems of interest the form of the SDE will not permit an analytic solution due to the intractability of the transition density, precluding straightforward inference for the unknown parameters. However, it is possible to construct a tractable approximation of the SDE, namely, the linear noise

approximation (LNA), which is the subject of the next section.

2.7 The linear noise approximation (LNA)

The LNA typically refers to an approximation to the solution of the Kolmogorov forward equation governing the transition probability of a Markov jump process. Specifically, the Kolmogorov forward equation is approximated by a Fokker-Planck equation with linear coefficients. The LNA first appeared in Kurtz (1970, 1971), where the technical details of how the LNA can be used as a functional central limit law for density dependent processes was presented. In Elf and Ehrenberg (2003), the LNA is considered for multiple macroscopic examples. Komorowski et al. (2009) discuss the LNA as a method for inferring kinetic parameters in a stochastic biochemical system. The LNA is used to derive a dynamic state space model for molecular populations in Finkenstädt et al. (2013). The accuracy of the LNA is discussed in Ferm et al. (2008) and Wallace et al. (2012). Fearnhead et al. (2014) also examine the accuracy of the LNA and go on to suggest ways to improve the accuracy over larger time frames. Golightly and Gillespie (2013) discuss the LNA as a way to simulate from a stochastic kinetic model and consider a Lotka-Volterra example. Golightly et al. (2015) implement the LNA in a delayed acceptance particle MCMC scheme for the parameters governing a stochastic kinetic model as a way to increase computational efficiency. An alternative derivation for the LNA (to the one given below) is given in Wallace (2010). For further discussion of the LNA we refer the reader to Wilkinson (2011) or van Kampen (2007).

Let us now consider the continuous-time d-dimensional homogeneous Itô process $\{X_t, t \geq 0\}$ satisfying the SDE

$$dX_t = \alpha(X_t, \theta)\, dt + \epsilon\sqrt{\beta(X_t, \theta)}\, dW_t, \qquad X_0 = x_0, \tag{2.25}$$

where $\epsilon \ll 1$. As before, α is a d-vector of drift functions, and the diffusion coefficient β is a $d \times d$ positive definite matrix with a square root representation $\sqrt{\beta}$ such that $\sqrt{\beta}\sqrt{\beta}' = \beta$. However the drift and diffusion coefficient may now depend upon θ as well as X_t, cf. (2.17). Again, W_t is a d-vector of (uncorrelated) standard Brownian motion processes. We now present a derivation of the LNA as a tractable approximation of (2.25).

2.7.1 Derivation of the LNA

As discussed in Fearnhead et al. (2014), the LNA can be derived directly as an approximation to the solution of an SDE. Since, throughout this thesis we take the SDE as the inferential model of interest, our derivation closely follows the approach of Fearnhead et al.

(2014). To begin with, we first partition X_t as

$$X_t = \eta_t + \epsilon R_t, \tag{2.26}$$

where $\{\eta_t, t \geq 0\}$ is a deterministic process satisfying the ODE

$$\frac{d\eta_t}{dt} = \alpha(\eta_t, \theta), \qquad \eta_0 = x_0 \tag{2.27}$$

and $\{R_t, t \geq 0\}$ is a residual stochastic process. Furthermore, we make the assumption that $||X_t - \eta_t||$ is $O(\epsilon)$ over a time interval of interest.

As X_t satisfies the SDE given by (2.25), the residual process (R_t) satisfies

$$dR_t = \frac{1}{\epsilon}\{\alpha(X_t, \theta) - \alpha(\eta_t, \theta)\}\, dt + \sqrt{\beta(X_t, \theta)}\, dW_t. \tag{2.28}$$

This SDE will typically be intractable. However a tractable approximation can be obtained by Taylor expanding $\alpha(X_t, \theta)$ and $\beta(X_t, \theta)$ about η_t. Here we obtain

$$\alpha(\eta_t + \epsilon R_t, \theta) = \alpha(\eta_t, \theta) + \epsilon H_t R_t + \cdots$$

and

$$\beta(\eta_t + \epsilon R_t, \theta) = \beta(\eta_t, \theta) + \cdots,$$

where H_t is the Jacobian matrix with (i,j)th element

$$(H_t)_{i,j} = \frac{\partial \alpha_i(\eta_t, \theta)}{\partial \eta_{j,t}}. \tag{2.29}$$

Collecting terms of $O(\epsilon)$ gives an SDE satisfied by an approximate residual process $\{\hat{R}_t, t \geq 0\}$ of the form

$$d\hat{R}_t = H_t \hat{R}_t\, dt + \sqrt{\beta(\eta_t, \theta)}\, dW_t. \tag{2.30}$$

In the above, we use ϵ to indicate that the stochastic term in (2.25) is small: essentially that, the drift term $\alpha(X_t, \theta)$ dominates the diffusion coefficient $\beta(X_t, \theta)$, or equivalently diffusion $\ll$ drift. However, ϵ does not feature in the evolution of (2.27) or (2.30). Therefore, from here on in, we assume $\epsilon = 1$. Note that, for η_t in equilibrium, (2.30) gives an Ornstein-Uhlenbeck process for $\hat{R}_t$. We consider the solution of (2.30) in the next section.

2.7.2 The LNA solution

Provided the initial condition for (2.30) is a fixed point mass ($\hat{R}_0 = \hat{r}_0$) or follows a Gaussian distribution, $\hat{R}_t$ is Gaussian for all $t > 0$. Let us assume that $\hat{R}_0 \sim N(\hat{r}_0, \hat{V}_0)$. Furthermore let P_t be the $d \times d$ fundamental matrix for the deterministic ODE

$$\frac{d\hat{r}_t}{dt} = H_t \hat{r}_t$$

which satisfies

$$\frac{dP_t}{dt} = H_t P_t, \qquad P_0 = I_d, \tag{2.31}$$

where I_d is the $d \times d$ identity matrix. Now

$$\frac{d}{dt} P_t P_t^{-1} = P_t \frac{dP_t^{-1}}{dt} + \frac{dP_t}{dt} P_t^{-1} = 0.$$

Therefore, using (2.31) it follows that

$$\frac{dP_t^{-1}}{dt} = -P_t^{-1} H_t. \tag{2.32}$$

Set $U_t = P_t^{-1} \hat{R}_t$. It is clear that $U_0 = \hat{R}_0$. We write

$$\begin{aligned} dU_t &= d\left(P_t^{-1} \hat{R}_t\right) \\ &= \left(dP_t^{-1}\right) \hat{R}_t + P_t^{-1} \left(d\hat{R}_t\right). \end{aligned}$$

Using (2.30) and (2.32) gives

$$\begin{aligned} dU_t &= \left(-P_t^{-1} H_t \, dt\right) \hat{R}_t + P_t^{-1} \left(H_t \hat{R}_t \, dt + \sqrt{\beta(\eta_t, \theta)} \, dW_t\right) \\ &= -P_t^{-1} H_t \hat{R}_t \, dt + P_t^{-1} H_t \hat{R}_t \, dt + P_t^{-1} \sqrt{\beta(\eta_t, \theta)} \, dW_t \\ &= P_t^{-1} \sqrt{\beta(\eta_t, \theta)} \, dW_t. \end{aligned}$$

Hence, we can write

$$U_t = U_0 + \int_0^t P_s^{-1} \sqrt{\beta(\eta_s, \theta)} \, dW_s.$$

Appealing to linearity and Itô isometry we obtain

$$U_t | U_0 \sim N\left\{U_0, \int_0^t P_s^{-1} \beta(\eta_s, \theta) \left(P_s^{-1}\right)' ds\right\}. \tag{2.33}$$

Algorithm 4 LNA

1. Set $t = 0$. Initialise θ and X_0. Set $\eta_0 = x_0$, $P_0 = I_d$, $\hat{r}_0 = x_0 - \eta_0$ (i.e. a vector of zeros) and $\psi_0 = 0$ (a $d \times d$ matrix with all elements equal to zero).
2. Solve the system of ODEs ((2.27), (2.35) and (2.36)) over $(t, t + \Delta t]$ to gain values of $\eta_{t+\Delta t}$, $P_{t+\Delta t}$ and $\psi_{t+\Delta t}$.
3. Draw $X_{t+\Delta t}$ from a $N(\eta_{t+\Delta t} + P_{t+\Delta t}\hat{r}_t, P_{t+\Delta t}\psi_{t+\Delta t}P'_{t+\Delta t})$ distribution.
4. Set $t = t + \Delta t$, $P_t = I_d$, $\hat{r}_t = x_t - \eta_t$ and $\psi_t = 0$.
5. Output t and x_t. If $t < T_{\max}$ return to step 2.

Therefore, for the initial assumption above, that is, $\hat{R}_0 (= U_0) \sim N(\hat{r}_0, \hat{V}_0)$, we have that

$$\hat{R}_t | \hat{R}_0 = \hat{r}_0 \sim N\left(P_t\hat{r}_0, P_t\psi_t P_t'\right), \tag{2.34}$$

where

$$\psi_t = \hat{V}_0 + \int_0^t P_s^{-1}\beta(\eta_s, \theta)\left(P_s^{-1}\right)' ds.$$

Thus, the SDE (2.30) satisfied by $\hat{R}_t$ can be solved analytically, where P_t and ψ_t satisfy the ODE system

$$\frac{dP_t}{dt} = H_t P_t, \qquad P_0 = I_d, \tag{2.35}$$

$$\frac{d\psi_t}{dt} = P_t^{-1}\beta\left(\eta_t, \theta\right)\left(P_t^{-1}\right)', \qquad \psi_0 = \hat{V}_0. \tag{2.36}$$

Hence the approximating distribution of X_t is given by

$$X_t \sim N\left(\eta_t + P_t\hat{r}_0, P_t\psi_t P_t'\right). \tag{2.37}$$

In the absence of an analytic solution, the system of coupled ODEs ((2.27), (2.35) and (2.36)) which characterise the LNA, must be solved numerically. Good numerical solvers that use adaptive time-steps are readily available; see, for example Petzold (1983). Throughout this thesis if the system of ODEs is required to be numerically solved in `R` then the `lsoda` package will be used. If the system is to be numerically solved in C then we appeal to the standard ODE solver from the GNU scientific library, namely the explicit embedded Runge-Kutta-Fehlberg $(4, 5)$ method (see Appendix A.1).

Given (2.37) we can obtain a realisation of X_t (at discrete times) using Algorithm 4.

Algorithm 5 LNA (with restart)

1. Set $t = 0$. Initialise θ and X_0. Set $\eta_0 = x_0$, $P_0 = I_d$ and $\psi_0 = 0$ (a $d \times d$ matrix with all elements equal to zero).
2. Solve the system of ODEs ((2.27), (2.35) and (2.36)) over $(t, t + \Delta t]$ to gain values of $\eta_{t+\Delta t}$, $P_{t+\Delta t}$ and $\psi_{t+\Delta t}$.
3. Draw $X_{t+\Delta t}$ from a $N(\eta_{t+\Delta t}, P_{t+\Delta t}\psi_{t+\Delta t}P'_{t+\Delta t})$ distribution.
4. Set $t = t + \Delta t$, $\eta_t = x_t$, $P_t = I_d$ and $\psi_t = 0$.
5. Output t and x_t. If $t < T_{\max}$ return to step 2.

2.7.3 Restarting the LNA

Fearnhead et al. (2014) discuss how the accuracy of the LNA can become poor over time (see also Giagos (2011) for empirical evidence). This is essentially due to the fact that within the approach of Algorithm 4, the ODE satisfied by η_t is integrated for all time. Thus it is possible that a significant difference between η_t and the underlying stochastic process can emerge. It is this difference which causes the accuracy to suffer. As a solution to alleviate this problem, Fearnhead et al. (2014) propose an approach which restarts the LNA at each simulation time. This restart is achieved by setting $\eta_t = x_t$ at each simulation time, and consequently $\hat{r}_t = 0$. Note that $\hat{r}_t$ is now zero for all t and therefore does not feature within the solution. Hence a solution under the LNA is found by solving the system of coupled ODEs ((2.27), (2.35) and (2.36)) over each interval $(t, t + \Delta t]$ where $\eta_t = x_t$, $P_t = I_d$ and $\psi_t = 0$. The approximating distribution of X_t is now

$$X_t \sim N\left(\eta_t, P_t\psi_t P_t'\right). \tag{2.38}$$

The steps to gain a realisation of X_t (at discrete times) incorporating the restart are given in Algorithm 5.

The form of the LNA given above in ((2.27), (2.35) and (2.36)) is (relatively) computationally expensive to implement in large inference schemes (see Chapters 4 and 5). This computational cost is a direct consequence of the number of matrix inverses which need to be calculated, coupled with the amount of matrix multiplication taking place. Typically these operations have complexity of approximately $O(n^3)$. It is however possible to construct an equivalent representation of the LNA, which is less computationally intensive.

Note that (2.34) can be written as

$$\hat{R}_t | \hat{R}_0 = \hat{r}_0 \sim N\left(m_t, V_t\right), \tag{2.39}$$

where it is clear from (2.35) that

$$\frac{dm_t}{dt} = H_t m_t, \qquad m_0 = \hat{r}_0. \tag{2.40}$$

The ODE for $V_t = P_t \psi_t P_t'$ can be obtained as

$$\frac{dV_t}{dt} = \frac{d\left(P_t \psi_t P_t'\right)}{dt},$$

which by the use of the product rule is

$$\begin{aligned}
\frac{dV_t}{dt} &= P_t \frac{d}{dt}\left(\psi_t P_t'\right) + \left(\frac{dP_t}{dt}\right)\psi_t P_t' \\
&= P_t\left\{\psi_t \frac{dP_t'}{dt} + \left(\frac{d\psi_t}{dt}\right)P_t'\right\} + H_t P_t \psi_t P_t' \\
&= P_t\left\{\psi_t P_t' H_t' + P_t^{-1}\beta\left(\eta_t, \theta\right)\left(P_t^{-1}\right)' P_t'\right\} + H_t P_t \psi_t P_t' \\
&= P_t \psi_t P_t' H_t + \beta\left(\eta_t, \theta\right) + H_t P_t \psi_t P_t'. \\
&= V_t H_t' + \beta(\eta_t, \theta, b) + H_t V_t, \qquad V_0 = 0.
\end{aligned} \tag{2.41}$$

Thus, it is possible to obtain a less computationally intensive solution by solving ((2.27), (2.40) and (2.41)) as opposed to ((2.27), (2.35) and (2.36)), where now the approximating distribution of X_t is

$$X_t \sim N\left(\eta_t + m_t, V_t\right). \tag{2.42}$$

Such an idea will play an important role in Chapters 4 and 5.

Unfortunately, as written, (2.42) will suffer from the same issues surrounding accuracy that were discussed above. As a solution, we again restart the LNA at each simulation time. Therefore, we set $\eta_t = x_t$ at each simulation time, resulting in $m_t = 0$ for all time. This renders the ODE (2.40) redundant, and thus no longer needs to be solved. Hence, implementing the LNA in this way reduces the dimension of the system of coupled ODEs, improving computational efficiency further still. Whence, a solution under the LNA is found by solving the ODEs (2.27) and (2.41) over $(t, t + \Delta t]$ with $\eta_t = x_t$ and $V_t = 0$, where the approximating distribution of X_t is

$$X_t \sim N\left(\eta_t, V_t\right). \tag{2.43}$$

Algorithm 6 gives the steps to gain a realisation of X_t (at discrete times) for this alternative representation of the LNA incorporating a restart.

Algorithm 6 LNA (with restart) II

1. Set $t = 0$. Initialise θ and X_0. Set $\eta_0 = x_0$ and $V_0 = 0$ (a $d \times d$ matrix with all elements equal to zero).
2. Solve the system of ODEs (2.27) and (2.41) over $(t, t + \Delta t]$ to gain values of $\eta_{t+\Delta t}$ and $V_{t+\Delta t}$.
3. Draw $X_{t+\Delta t}$ from a $N(\eta_{t+\Delta t}, V_{t+\Delta t})$ distribution.
4. Set $t = t + \Delta t$, $\eta_t = x_t$ and $V_t = 0$.
5. Output t and x_t. If $t < T_{\max}$ return to step 2.

2.7.4 Example: Lotka-Volterra model

We now highlight the importance of incorporating the restart by means of an example. Let us consider a Lotka-Volterra model of predator-prey dynamics for $X_t = (X_{1,t}, X_{2,t})'$, ordered (prey, predator) at time t. The mass-action SDE representation of the system dynamics is given by

$$dX_t = \begin{pmatrix} \theta_1 X_{1,t} - \theta_2 X_{1,t} X_{2,t} \\ \theta_2 X_{1,t} X_{2,t} - \theta_3 X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \theta_1 X_{1,t} + \theta_2 X_{1,t} X_{2,t} & -\theta_2 X_{1,t} X_{2,t} \\ -\theta_2 X_{1,t} X_{2,t} & \theta_3 X_{2,t} + \theta_2 X_{1,t} X_{2,t} \end{pmatrix}^{\frac{1}{2}} dW_t. \tag{2.44}$$

Therefore, via ((2.27), (2.35) and (2.36)) the linear noise approximation of (2.44) has

$$\begin{aligned} \frac{d\eta_t}{dt} &= \begin{pmatrix} \theta_1 \eta_{1,t} - \theta_2 \eta_{1,t} \eta_{2,t} \\ \theta_2 \eta_{1,t} \eta_{2,t} - \theta_3 \eta_{2,t} \end{pmatrix}, \\ \frac{dP_t}{dt} &= \begin{pmatrix} \theta_1 - \theta_2 \eta_{2,t} & -\theta_2 \eta_{1,t} \\ \theta_2 \eta_{2,t} & \theta_2 \eta_{1,t} - \theta_3 \end{pmatrix} P_t, \\ \frac{d\psi_t}{dt} &= P_t^{-1} \begin{pmatrix} \theta_1 \eta_{1,t} + \theta_2 \eta_{1,t} \eta_{2,t} & -\theta_2 \eta_{1,t} \eta_{2,t} \\ -\theta_2 \eta_{1,t} \eta_{2,t} & \theta_3 \eta_{2,t} + \theta_2 \eta_{1,t} \eta_{2,t} \end{pmatrix} \left(P_t^{-1}\right)'. \end{aligned}$$

We aim to generate a realisation of X_t through the use of the above algorithms. Note that the ODE system above is intractable, and we use the `R` package `lsoda` to numerically solve the system in Algorithms 4 and 5.

We follow Boys et al. (2008) and set $\theta = (\theta_1, \theta_2, \theta_3)' = (0.5, 0.0025, 0.3)'$ with $x_0 = (71, 79)'$. Figure 2.3 depicts a single realisation of the Lotka-Volterra model from both the LNA (Algorithm 4) and the LNA with the restart included (Algorithm 5) for a time-step of $\Delta t = 0.1$. Whilst not conclusive, Figure 2.3 hints at the differences which can appear

over time through the incorporation of the restart. Investigating further, we compare 95% credible regions for the number of prey and predators from 100K simulations. Figure 2.4 illustrates these credible regions for both Algorithms 4 and 5 ($\Delta t = 0.1$), as well as the credible region for the true underlying stochastic kinetic model (obtained by direct sampling from the true underlying Markov jump process using the Gillespie algorithm, see Appendix C). Note that, whilst reactions within the Gillespie algorithm occur in continuous-time, we collect output every 0.1 to enable direct comparisons between the methods. It is evident from Figure 2.4 that without the inclusion of the restart, the LNA becomes 'out of sync' with the underlying stochastic process over time. Although a theoretical justification of the restarted LNA is missing from the literature, empirical evidence suggests that this approach can work well in practice. Henceforth, within this thesis, any implementation of the LNA will assume that the restart is included (unless stated otherwise).

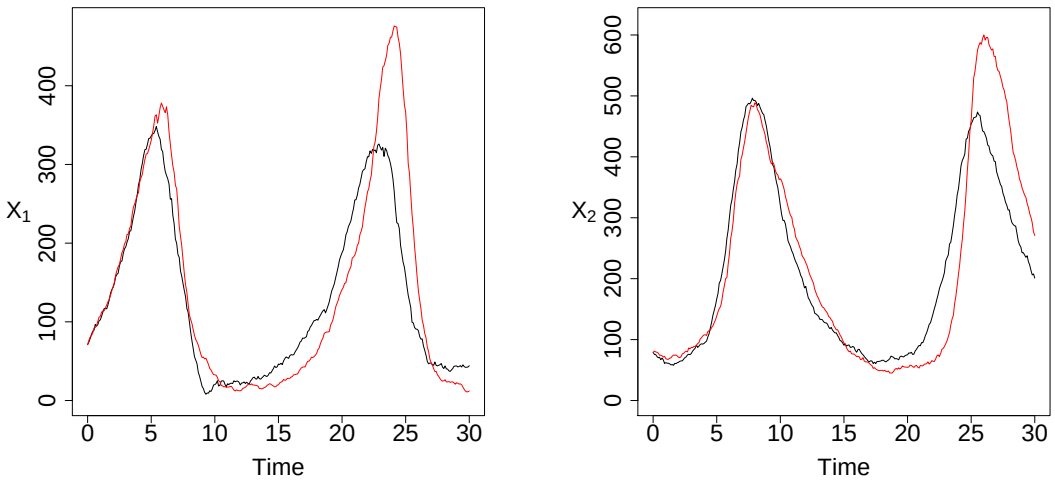

Figure 2.3: A single realisation of prey (X_1) and predator (X_2) in the Lotka-Volterra model, $x_0 = (71, 79)'$ and $\theta = (0.5, 0.0025, 0.3)'$ with time-step $\Delta t = 0.1$. Black: LNA (Algorithm 4). Red: LNA with restart (Algorithm 5).

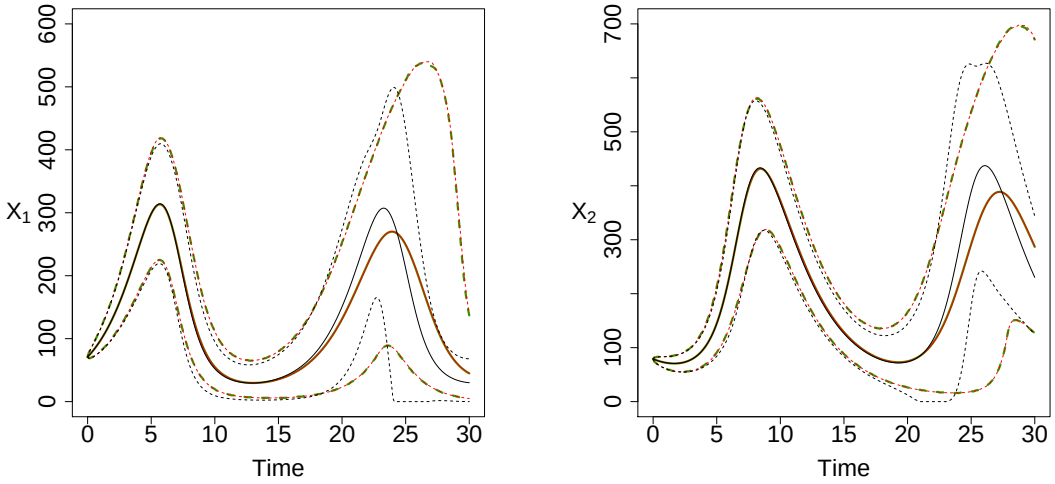

Figure 2.4: Lotka-Volterra model. 95% credible region (dashed line) and mean (solid line) for numbers of prey (X_1) and predator (X_2) on a uniform grid of step-size 0.1, $x_0 = (71, 79)'$ and $\theta = (0.5, 0.0025, 0.3)'$. Black: LNA (Algorithm 4). Red: LNA with restart (Algorithm 5). Green: True stochastic kinetic model.

Chapter 3

Bridge constructs for stochastic differential equations

As mentioned previously within this thesis, the transition densities of a diffusion process (satisfying an SDE) are likely to be intractable. However it is possible to numerically approximate these unavailable transition densities (Pedersen, 1995; Elerian et al., 2001; Eraker, 2001; Roberts and Stramer, 2001). This numerical approach can be seen as a data augmentation problem, where the simplest implementation augments low-frequency data by introducing latent values at intermediate time points between observation times. An Euler-Maruyama scheme is then applied by approximating the transition densities over the induced discretisation as Gaussian. Integrating over the uncertainty associated with the latent values typically requires the use of Monte Carlo, coupled with an appropriate proposal density for generating realisations of the latent values in between the observations. Such realisations are known as *diffusion bridges*. The designing of bridge constructs for irreducible nonlinear multivariate diffusions is a challenging problem and has received much attention in the recent literature.

Within this chapter we first discuss existing approaches to constructing *diffusion bridges*. The modified diffusion bridge of Durham and Gallant (2002) (see also extensions to the partial and noisy observation case in Golightly and Wilkinson (2008)) pushes the bridge process towards the observation in a linear way and provides the optimal sampling method when the drift and diffusion coefficients of the SDE are constant (Stramer and Yan, 2006). However, this construct does not produce efficient proposals when the process exhibits nonlinear dynamics. We therefore propose a novel class of bridge constructs that can capture nonlinear behaviour. Moreover, our approach is computationally and statistically efficient, simple to implement, and can be applied in scenarios where only partial and noisy measurements of the system are available. Essentially, the process is partitioned into

two parts, one that accounts for nonlinear dynamics in a deterministic way, and another as a residual stochastic process. A bridge construct is obtained for the target process by applying the modified diffusion bridge sampler to the end-point conditioned residual process. We consider two implementations of this approach. Firstly, we use a bridge introduced by Whitaker et al. (2016*a*) that constructs the residual process by subtracting the solution of an ordinary differential equation system based on the drift, from the target process. Secondly, we recognise that the intractable SDE governing the residual process can be approximated by a tractable process. We therefore extend the first approach by additionally subtracting the expectation of the approximate residual process and bridging the remainder with the modified diffusion bridge sampler. In addition, we adapt the guided proposal proposed by Schauer et al. (2016) to a partial and noisy observation regime. We conclude this chapter with three examples through which we showcase the differing properties of each individual bridge construct, whilst also assessing statistical efficiency by means of empirical acceptance probabilities.

3.1 Sampling a conditioned SDE

As in Chapter 2, let us consider a continuous-time d-dimensional Itô process $\{X_t, t \geq 0\}$ governed by an SDE paramaterised by $\theta = (\theta_1, \ldots, \theta_p)'$, of the form

$$dX_t = \alpha(X_t, \theta)\, dt + \sqrt{\beta(X_t, \theta)}\, dW_t, \quad X_0 = x_0. \tag{3.1}$$

For tractability, we make the same assumption as Golightly and Wilkinson (2008, 2011), Picchini (2014) and Lu et al. (2015) among others, that the process is observed at $t = T$ according to

$$Y_T = F' X_T + \epsilon_T, \qquad \epsilon_T | \Sigma \overset{indep}{\sim} N(0, \Sigma). \tag{3.2}$$

Here Y_T is a d_o-vector, F is a constant $d \times d_o$ matrix and ϵ_T is a random d_o-vector. Note that this setup allows for only observing a subset of components ($d_o < d$). For simplicity we also assume that the process is known exactly at $t = 0$. This is the case when a diffusion process is observed completely and without error. In the case of partial and/or noisy observations, typically the initial position is an unknown parameter in an MCMC scheme and a new bridge is created at each iteration conditional on the current parameter values, so in terms of the bridge, the initial position is effectively known. Without loss of generality we consider an interval $[0, T]$, and note that the arguments made within this chapter are easily scalable to multiple observations (and therefore multiple intervals). The case of multiple partial and/or noisy observations is discussed in Chapter 4.

Our aim is to generate discrete-time realisations of X_t conditional on x_0 and y_T. To this

end, we partition $[0, T]$ as

$$0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{m-1} < \tau_m = T, \tag{3.3}$$

giving m intervals of equal length $\Delta\tau = T/m$. Since, in general, the form of the SDE in (3.1) will not permit an analytic solution, we work with the Euler-Maruyama approximation which gives the change in the process over a small interval of length $\Delta\tau$ as a Gaussian random vector, see, for example, Kloeden and Platen (1992). Specifically, we have that

$$\Delta X_{\tau_k} = X_{\tau_{k+1}} - X_{\tau_k} = \alpha(X_{\tau_k}, \theta)\,\Delta\tau + \sqrt{\beta(X_{\tau_k}, \theta)}\,\Delta W_{\tau_k}, \tag{3.4}$$

where $\Delta W_{\tau_k} \sim N(0, \Delta\tau I_d)$. The continuous-time conditioned process is then approximated by the discrete-time skeleton bridge, with the latent values

$$x_{(0,T]} = (x_{\tau_1}, \ldots, x_{\tau_m} = x_T)'$$

having the (posterior) density

$$\pi\left(x_{(0,T]}|x_0, y_T, \theta, \Sigma\right) \propto \pi\left(y_T|x_T, \Sigma\right) \prod_{k=0}^{m-1} \pi\left(x_{\tau_{k+1}}|x_{\tau_k}, \theta\right), \tag{3.5}$$

where $\pi(x_{\tau_{k+1}}|x_{\tau_k}, \theta) = N\left(x_{\tau_{k+1}}\,;\, x_{\tau_k} + \alpha(x_{\tau_k}, \theta)\Delta\tau,\ \beta(x_{\tau_k}, \theta)\Delta\tau\right)$ is the transition density under the Euler-Maruyama approximation, $\pi(y_T|x_T, \Sigma) = N(y_T\,;\, F'x_T, \Sigma)$ and $N(\cdot; m, V)$ denotes the multivariate Gaussian density with mean vector m and variance matrix V. In the special case where x_T is known (so that $y_T = x_T$ and $F = I_d$), the latent values $x_{(0,T)} = (x_{\tau_1}, \ldots, x_{\tau_{m-1}})'$ have the density

$$\pi\left(x_{(0,T)}|x_0, x_T, \theta\right) \propto \prod_{k=0}^{m-1} \pi\left(x_{\tau_{k+1}}|x_{\tau_k}, \theta\right). \tag{3.6}$$

For nonlinear forms of the drift and diffusion coefficients, the products in (3.5) and (3.6) will be intractable and samples can be generated via computationally intensive algorithms such as Markov chain Monte Carlo (see Chapter 2) or importance sampling. We focus on the former but note that in either case, the efficiency of the algorithm will depend on the proposal mechanism used to generate the bridge. A common approach to constructing an efficient proposal is to factorise the target in (3.5) as

$$\pi\left(x_{(0,T]}|x_0, y_T, \theta, \Sigma\right) \propto \prod_{k=0}^{m-1} \pi\left(x_{\tau_{k+1}}|x_{\tau_k}, y_T, \theta, \Sigma\right). \tag{3.7}$$

The density in (3.6) can be factorised in a similar manner. This suggests seeking pro-

posal densities of the form $q(x_{\tau_{k+1}}|x_{\tau_k}, y_T, \theta, \Sigma)$ which aim to approximate the intractable constituent densities in (3.7). In what follows, we consider some existing approaches for generating bridges via approximation of $\pi(x_{\tau_{k+1}}|x_{\tau_k}, y_T, \theta, \Sigma)$ before outlining our contribution. For each bridge, the proposal densities take the form

$$q\left(x_{\tau_{k+1}}|x_{\tau_k}, y_T, \theta, \Sigma\right) = N\left\{x_{\tau_{k+1}}\,;\, x_{\tau_k} + \mu(x_{\tau_k})\Delta\tau\,,\, \Psi(x_{\tau_k})\Delta\tau\right\} \tag{3.8}$$

and our focus is on the choice of suitable $\mu(\cdot)$ and $\Psi(\cdot)$. For simplicity and where possible, we drop the parameters θ and Σ from the notation as they remain fixed throughout this chapter.

3.1.1 Myopic simulation

Ignoring the information in the observation y_T and simply applying the Euler-Maruyama approximation (3.4) over each interval of length $\Delta\tau$ leads to a proposal density of the form (3.8) with $\mu_{\mathrm{EM}}(x_{\tau_k}) = \alpha(x_{\tau_k})$ and $\Psi_{\mathrm{EM}}(x_{\tau_k}) = \beta(x_{\tau_k})$. Sampling recursively according to (3.8) for $k = 0, 1, \ldots, m-1$ gives a proposed bridge which we denote by $x^*_{(0,T]}$. The Metropolis-Hastings acceptance probability for a move from $x_{(0,T]}$ to $x^*_{(0,T]}$ is

$$\min\left\{1\,,\, \frac{\pi\left(y_T|x^*_T\right)}{\pi\left(y_T|x_T\right)}\right\}.$$

The simplified structure of the above is achieved as both the target and proposal densities are of the same form, leading to cancellations in the Metropolis-Hastings acceptance probability. This strategy is likely to work well provided that the observation y_T is not particularly informative, that is, when the measurement error dominates the intrinsic stochasticity of the process. However, as $\mathrm{tr}(\Sigma)$ is reduced, the Metropolis-Hastings acceptance rate decreases. A related approach can be found in Pedersen (1995), where it is assumed that x_T is known. In this case, a move from $x_{(0,T)}$ to $x^*_{(0,T)}$ is accepted with probability

$$\min\left\{1\,,\, \frac{\pi\big(x_T|x^*_{\tau_{m-1}}\big)}{\pi\big(x_T|x_{\tau_{m-1}}\big)}\right\},$$

which tends to 0 as $m \to \infty$ (or equivalently, $\Delta\tau \to 0$).

3.1.2 Modified diffusion bridge

For known x_T, Durham and Gallant (2002) (see also Golightly and Wilkinson (2006)) derive a linear Gaussian approximation of $\pi(x_{\tau_{k+1}}|x_{\tau_k}, x_T)$, leading to a sampler known as the modified diffusion bridge (MDB). Extensions to the partial and noisy observation

regime are considered in Golightly and Wilkinson (2008). Initially the joint distribution of $X_{\tau_{k+1}}$ and Y_T (conditional on x_{τ_k}) is approximated before multivariate normal theory is used to condition on Y_T.

We approximate the distribution of Y_T conditional on $X_{\tau_{k+1}}$ via a very crude Euler approximation, giving

$$Y_T|X_{\tau_{k+1}} \sim N\left(F'\left\{X_{\tau_{k+1}} + \alpha\left(X_{\tau_{k+1}}\right)\Delta_{k+1}\right\}, F'\beta\left(X_{\tau_{k+1}}\right)F\Delta_{k+1} + \Sigma\right) \tag{3.9}$$

where $\Delta_{k+1} = T - \tau_{k+1}$. To obtain a linear Gaussian structure we approximate (3.9) using the assumption that α and β (the drift and diffusion coefficient) are locally constant. We therefore estimate $\alpha(X_{\tau_{k+1}})$ and $\beta(X_{\tau_{k+1}})$ by $\alpha_k = \alpha(x_{\tau_k})$ and $\beta_k = \beta(x_{\tau_k})$ respectively. Thus, we obtain

$$Y_T|X_{\tau_{k+1}} \sim N\left\{F'\left(X_{\tau_{k+1}} + \alpha_k\Delta_{k+1}\right), F'\beta_k F\Delta_{k+1} + \Sigma\right\}. \tag{3.10}$$

Through the Euler-Maruyama approximation we have that the distribution of $X_{\tau_{k+1}}$ conditional on x_{τ_k} is

$$X_{\tau_{k+1}}|x_{\tau_k} \sim N(x_{\tau_k} + \alpha_k\Delta\tau, \beta_k\Delta\tau). \tag{3.11}$$

Hence, the joint distribution of $X_{\tau_{k+1}}$ and Y_T (conditional on x_{τ_k}) is approximated by

$$\begin{pmatrix} X_{\tau_{k+1}} \\ Y_T \end{pmatrix} \Bigg| x_{\tau_k} \sim N\left\{\begin{pmatrix} x_{\tau_k} + \alpha_k\Delta\tau \\ F'\left(x_{\tau_k} + \alpha_k\Delta_k\right) \end{pmatrix}, \begin{pmatrix} \beta_k\Delta\tau & \beta_k F\Delta\tau \\ F'\beta_k\Delta\tau & F'\beta_k F\Delta_k + \Sigma \end{pmatrix}\right\}$$

where $\Delta_k = T - \tau_k$. Conditioning further on $Y_T = y_T$ gives

$$X_{\tau_{k+1}}|x_{\tau_k}, y_T \sim N\left\{x_{\tau_k} + \mu_{\text{MDB}}\left(x_{\tau_k}\right)\Delta\tau\,,\ \Psi_{\text{MDB}}\left(x_{\tau_k}\right)\Delta\tau\right\}$$

where

$$\mu_{\text{MDB}}\left(x_{\tau_k}\right) = \alpha_k + \beta_k F\left(F'\beta_k F\Delta_k + \Sigma\right)^{-1}\left\{y_T - F'\left(x_{\tau_k} + \alpha_k\Delta_k\right)\right\} \tag{3.12}$$

and

$$\Psi_{\text{MDB}}\left(x_{\tau_k}\right) = \beta_k - \beta_k F\left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} F'\beta_k\Delta\tau. \tag{3.13}$$

In the case of no measurement error and observation of all components (so that x_T is known), (3.12) and (3.13) become

$$\mu^*_{\text{MDB}}\left(x_{\tau_k}\right) = \frac{x_T - x_{\tau_k}}{T - \tau_k} \quad \text{and} \quad \Psi^*_{\text{MDB}}\left(x_{\tau_k}\right) = \frac{T - \tau_{k+1}}{T - \tau_k}\beta_k.$$

Connection with continuous-time conditioned processes

Consider the case of no measurement error and full observation of all components. The SDE satisfied by the conditioned process $\{X_t, t \in [0,T]\}$, takes the form

$$dX_t = \tilde{\alpha}\left(X_t\right)\, dt + \sqrt{\beta\left(X_t\right)}\, dW_t, \quad X_0 = x_0, \tag{3.14}$$

where the drift is

$$\tilde{\alpha}\left(X_t\right) = \alpha\left(X_t\right) + \beta\left(X_t\right) \nabla_{x_t} \log p\left(x_T | x_t\right). \tag{3.15}$$

See, for example, Chapter IV.39 of Rogers and Williams (2000) for a derivation. Note that $p(x_T|x_t)$ denotes the (intractable) transition density of the unconditioned process defined in (3.1). Approximating $\alpha(X_t)$ and $\beta(X_t)$ in (3.1) by the constants $\alpha(x_T)$ and $\beta(x_T)$ yields a process for which $p(x_T|x_t)$ is tractable. The corresponding conditioned process satisfies

$$dX_t = \frac{X_T - X_t}{T - t}\, dt + \sqrt{\beta(X_t)}\, dW_t. \tag{3.16}$$

Use of (3.16) as a proposal process has been justified by Delyon and Hu (2006) (see also Stramer and Yan (2006), Marchand (2011) and Papaspiliopoulos et al. (2013)), who show that the distribution of the target process (conditional on x_T) is absolutely continuous with respect to the distribution of the solution to (3.16). As discussed by Papaspiliopoulos et al. (2013), it is impossible to simulate exact (discrete-time) realisations of (3.16) unless $\beta(\cdot)$ is constant. They also note that performing a local linearisation of (3.16) according to Shoji and Ozaki (1998) (see also Shoji (2011)) gives a tractable process with transition density

$$q(x_{\tau_{k+1}} | x_{\tau_k}, x_T) = N\left\{x_{\tau_{k+1}}\,;\, x_{\tau_k} + \frac{x_T - x_{\tau_k}}{T - \tau_k}\Delta\tau\,,\, \frac{T - \tau_{k+1}}{T - \tau_k}\beta\left(x_{\tau_k}\right)\Delta\tau\right\},$$

that is, the transition density of the modified diffusion bridge discussed in the previous section. Plainly, taking the Euler-Maruyama approximation of (3.16) yields the MDB construct, albeit without the time dependent multiplier of β_k in the variance. As observed by Durham and Gallant (2002) and discussed in Papaspiliopoulos and Roberts (2012) and Papaspiliopoulos et al. (2013), the inclusion of the time dependent multiplier can lead to improved empirical performance.

Whilst this construct can, in principle, be applied to arbitrary nonlinear multivariate diffusion processes, the effect of the Gaussian approximation is to guide the bridge towards the observation in a linear way, unless there is large uncertainty in the observation process. This effect is exacerbated in the case of no measurement error, in which case the resulting construct is independent of the drift of the target process (see (3.16)). Consequently, use

of the MDB as a proposal mechanism (in a Metropolis-Hastings independence sampler) is likely to result in low acceptance rates unless the drift is of little importance in dictating the dynamics of the target process between observation times. In other words, the MDB is likely to be unsatisfactory in situations where realisations of the target SDE (with the same initial condition) exhibit strong and similar nonlinearity over the inter-observation time.

3.1.3 Lindström bridge

A bridge construct that combines the myopic sampler with the MDB has been proposed in Lindström (2012), for the special case of known x_T. Lindström's approach is to use the insight that it is worse to have a proposal distribution which is too light-tailed than too heavy-tailed, in the designing of the construct (see Geweke (1989) and Koopman et al. (2009)). Extending the sampler to the observation scenario in (3.2) is straightforward. Whereas the MDB approximates the variance of $Y_T|x_{\tau_k}$ by $F'\beta_k F\Delta_k + \Sigma$, the simplest version of the Lindström bridge (LB) has that

$$\mathrm{Var}\left(Y_T|x_{\tau_k}\right) \simeq F'\left\{\beta_k\Delta_k + C\left(\Delta_{k+1}\right)^2\right\}F + \Sigma,$$

where $C(\Delta_{k+1})^2$ is the squared bias of $X_T|x_{\tau_{k+1}}$ using a single Euler-Maruyama time-step and C is an unknown matrix. By assuming that the squared bias is a fraction γ of the variance over an interval of length $\Delta\tau$, a heuristic choice of C is given by

$$C_{\mathrm{Heur}} = \frac{\gamma\beta_k}{\Delta\tau},$$

with $\gamma > 0$. This particular choice of C_{Heur} ensures that $\mathrm{Var}(Y_T|x_{\tau_k})$ is a positive definite matrix. The joint distribution of $X_{\tau_{k+1}}$ and Y_T (conditional on x_{τ_k}) is then approximated by

$$\begin{pmatrix} X_{\tau_{k+1}} \\ Y_T \end{pmatrix}\Bigg| x_{\tau_k} \sim N\left\{\begin{pmatrix} x_{\tau_k} + \alpha_k\Delta\tau \\ F'\left(x_{\tau_k} + \alpha_k\Delta_k\right)\end{pmatrix}, \begin{pmatrix} \beta_k\Delta\tau & \beta_k F\Delta\tau \\ F'\beta_k\Delta\tau & F'\beta_k F\Delta_k^\gamma + \Sigma\end{pmatrix}\right\},$$

where $\Delta_k^\gamma = \Delta_k + \gamma(\Delta_{k+1})^2/\Delta\tau$. Conditioning further on $Y_T = y_T$ gives

$$X_{\tau_{k+1}}|x_{\tau_k}, y_T \sim N\left\{x_{\tau_k} + \mu_{\mathrm{LB}}\left(x_{\tau_k}\right)\Delta\tau\,,\, \Psi_{\mathrm{LB}}\left(x_{\tau_k}\right)\Delta\tau\right\}$$

where

$$\mu_{\mathrm{LB}}\left(x_{\tau_k}\right) = \alpha_k + \beta_k F\left(F'\beta_k F\Delta_k^\gamma + \Sigma\right)^{-1}\left\{y_T - F'\left(x_{\tau_k} + \alpha_k\Delta_k\right)\right\} \tag{3.17}$$

and

$$\Psi_{\mathrm{LB}}\left(x_{\tau_k}\right) = \beta_k - \beta_k F\left(F'\beta_k F\Delta_k^{\gamma} + \Sigma\right)^{-1} F'\beta_k \Delta\tau. \tag{3.18}$$

In the case of no measurement error and observation of all components, (3.17) and (3.18) become

$$\mu^*_{\mathrm{LB}}\left(x_{\tau_k}\right) = w_k^{\gamma}\mu^*_{\mathrm{MDB}}\left(x_{\tau_k}\right) + \left(1 - w_k^{\gamma}\right)\alpha_k$$

and

$$\Psi^*_{\mathrm{LB}}\left(x_{\tau_k}\right) = w_k^{\gamma}\Psi^*_{\mathrm{MDB}}\left(x_{\tau_k}\right) + \left(1 - w_k^{\gamma}\right)\beta_k,$$

where

$$w_k^{\gamma} = \frac{\left(\tau_{k+1} - \tau_k\right)\left(T - \tau_k\right)}{\left(\tau_{k+1} - \tau_k\right)\left(T - \tau_k\right) + \gamma\left(T - \tau_{k+1}\right)^2}.$$

The LB can therefore be seen as a convex combination of the MDB and myopic samplers, with $\gamma = 0$ giving the MDB and $\gamma = \infty$ giving the myopic approach. In practice, Lindström (2012) suggests that $\gamma \in [0.01, 1]$, given that these values have proved successful in simulation experiments. Note also that for a fixed γ, if $T - \tau_{k+1} \gg \Delta\tau$ then $w_k^{\gamma} \simeq 0$ and the myopic sampler dominates. However, as τ_{k+1} approaches T, w_k^{γ} approaches 1 and the LB is dominated by the MDB.

Whilst the LB attempts to account for nonlinear dynamics by combining the MDB with the myopic approach, having to specify a model-dependent tuning parameter is unsatisfactory since different choices of γ will lead to different properties of the proposed bridges. Moreover, the link between the regularised sampler and the continuous-time conditioned process is unclear.

3.2 Improved bridge constructs

In this section we describe a novel class of bridge constructs that require no tuning parameters, are simple to implement (even when only a subset of components are observed with Gaussian noise) and can account for nonlinear dynamics driven by the drift. In addition, we discuss the recently proposed bridging strategy of Schauer et al. (2016) and describe an implementation method in the case of partial observation with additive Gaussian measurement error.

3.2.1 Bridges based on residual processes

Suppose that X_t is partitioned as $X_t = \zeta_t + R_t$ where $\{\zeta_t, t \geq 0\}$ is a deterministic process and $\{R_t, t \geq 0\}$ is a residual stochastic process, satisfying

$$\begin{aligned} d\zeta_t &= f(\zeta_t)\, dt, \quad \zeta_0 = x_0, \\ dR_t &= \{\alpha(X_t) - f(\zeta_t)\}\, dt + \sqrt{\beta(X_t)}\, dW_t, \quad R_0 = 0. \end{aligned} \tag{3.19}$$

We then aim to choose ζ_t (and therefore $f(\cdot)$) to adequately account for nonlinear dynamics (so that the drift in (3.19) is approximately constant), and construct the MDB of Section 3.1.2 for the *residual stochastic process* rather than the target process itself. It should be clear from the discussion in Section 3.1.2 that for known x_T, the MDB approximates the density of $R_{\tau_{k+1}} | r_{\tau_k}, r_T$ by

$$q(r_{\tau_{k+1}} | r_{\tau_k}, r_T) = N\left\{ r_{\tau_{k+1}} ;\, r_{\tau_k} + \frac{r_T - r_{\tau_k}}{T - \tau_k}\Delta\tau\,,\, \frac{T - \tau_{k+1}}{T - \tau_k}\beta(x_{\tau_k})\,\Delta\tau \right\}. \tag{3.20}$$

In this case, the connection between (3.20) and the intractable continuous-time conditioned residual process can be established by following the arguments of Section 3.1.2. By approximating the drift and diffusion matrix in (3.19) by the constants $\alpha(x_T) - f(\zeta_T)$ and $\beta(x_T)$ gives a process with a tractable transition density. The corresponding conditioned process then satisfies

$$dR_t = \frac{R_T - R_t}{T - t}\, dt + \sqrt{\beta(X_t)}\, dW_t. \tag{3.21}$$

The density in (3.20) is then obtained by a local linearisation of (3.21).

It remains for us to choose ζ_t to balance the accuracy and computational efficiency of the resulting construct. We explore two possible choices in the remainder of this section.

Subtracting the drift

In the simplest approach to account for dynamics based on the drift, we take $\zeta_t = \eta_t$ and $f(\cdot) = \alpha(\cdot)$ where

$$\frac{d\eta_t}{dt} = \alpha(\eta_t), \quad \eta_0 = x_0. \tag{3.22}$$

In other words, we take a deterministic process satisfying the ODE based on the drift, so that

$$dR_t = \{\alpha(X_t) - \alpha(\eta_t)\}\, dt + \sqrt{\beta(X_t)}\, dW_t, \quad R_0 = 0. \tag{3.23}$$

Note that this approach explicitly partitions X_t as $X_t = \eta_t + R_t$. This is the same partition used by Fearnhead et al. (2014) (see also Section 2.7) to derive a tractable approximation to the intractable transition densities governing X_t, whereas our primary motivation for this partition is the application of the MDB to the residual process, thus giving a proposal that is likely to perform well for arbitrarily fine discretisations and explicitly incorporates the drift of the target SDE. The MDB can be constructed for the residual process by approximating the joint distribution of $R_{\tau_{k+1}}$ and $Y_T - F'\eta_T$ (conditional on r_{τ_k}), where $Y_T - F'\eta_T$ can be seen as a partial and noisy observation of R_T since

$$Y_T - F'\eta_T = F'R_T + \epsilon_T, \qquad \epsilon_T|\Sigma \sim N(0, \Sigma).$$

As in Section 3.1.2, we obtain the (approximate) joint distribution

$$\begin{pmatrix} R_{\tau_{k+1}} \\ Y_T - F'\eta_T \end{pmatrix} \Big| r_{\tau_k} \sim N\left\{ \begin{pmatrix} r_{\tau_k} + \left(\alpha_k - \alpha_k^\eta\right)\Delta\tau \\ F'\left\{r_{\tau_k} + \left(\alpha_k - \alpha_k^\eta\right)\Delta_k\right\} \end{pmatrix}, \begin{pmatrix} \beta_k\Delta\tau & \beta_k F\Delta\tau \\ F'\beta_k\Delta\tau & F'\beta_k F\Delta_k + \Sigma \end{pmatrix} \right\}, \tag{3.24}$$

where $\alpha_k^\eta = \alpha(\eta_{\tau_k})$ and α_k, β_k and Δ_k are as defined in Section 3.1.2. Note that the mean in (3.24) uses the tangent α_k^η at (τ_k, η_{τ_k}) to approximate $d\eta_t/dt$ over time intervals of length $\Delta\tau$ and Δ_k. Since $\eta_{\tau_{k+1}}$ will be available either exactly from the solution of (3.22) or from the output of a (stiff) ODE solver, we propose to approximate $d\eta_t/dt$ via the chord between (τ_k, η_{τ_k}) and $(\tau_{k+1}, \eta_{\tau_{k+1}})$, that is, by

$$\delta_k^\eta = \frac{\eta_{\tau_{k+1}} - \eta_{\tau_k}}{\Delta\tau}.$$

Replacing α_k^η in (3.24) with δ_k^η gives

$$\begin{pmatrix} R_{\tau_{k+1}} \\ Y_T - F'\eta_T \end{pmatrix} \Big| r_{\tau_k} \sim N\left\{ \begin{pmatrix} r_{\tau_k} + \left(\alpha_k - \delta_k^\eta\right)\Delta\tau \\ F'\left\{r_{\tau_k} + \left(\alpha_k - \delta_k^\eta\right)\Delta_k\right\} \end{pmatrix}, \begin{pmatrix} \beta_k\Delta\tau & \beta_k F\Delta\tau \\ F'\beta_k\Delta\tau & F'\beta_k F\Delta_k + \Sigma \end{pmatrix} \right\}. \tag{3.25}$$

Conditioning further on $y_T - F'\eta_T$ (and using the partition $X_t = \eta_t + R_t$) we obtain

$$X_{\tau_{k+1}}|x_{\tau_k}, y_T \sim N\left\{x_{\tau_k} + \mu_{\mathrm{RB}}\left(x_{\tau_k}\right)\Delta\tau\,,\ \Psi_{\mathrm{RB}}\left(x_{\tau_k}\right)\Delta\tau\right\}$$

where $\Psi_{\mathrm{RB}}(x_{\tau_k}) = \Psi_{\mathrm{MDB}}(x_{\tau_k})$ and

$$\mu_{\mathrm{RB}}\left(x_{\tau_k}\right) = \alpha_k + \beta_k F\left(F'\beta_k F\Delta_k + \Sigma\right)^{-1}\left[y_T - F'\left\{\eta_T + r_{\tau_k} + \left(\alpha_k - \delta_k^\eta\right)\Delta_k\right\}\right]. \tag{3.26}$$

Note that in the case of known x_T, $\Psi^*_{\mathrm{RB}}(x_{\tau_k}) = \Psi^*_{\mathrm{MDB}}(x_{\tau_k})$ and (3.26) becomes

$$\mu^*_{\mathrm{RB}}\left(x_{\tau_k}\right) = \delta_k^\eta + \frac{(x_T - x_{\tau_k}) - (\eta_T - \eta_{\tau_k})}{T - \tau_k}.$$

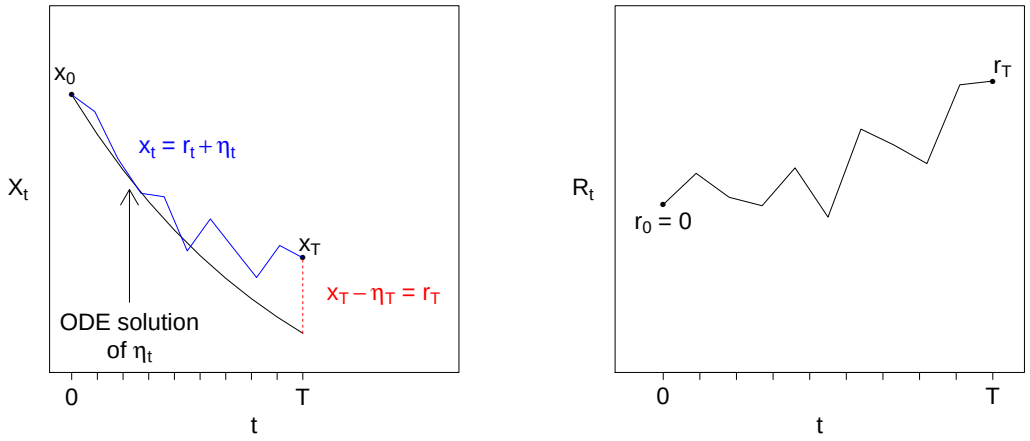

Figure 3.1: An illustration of the RB construct. Left: The full bridge. Right: A sample path of R_t.

The scheme is illustrated in Figure 3.1.

Further subtraction using the linear noise approximation

Whilst the solution of the SDE governing the residual stochastic process in (3.23) is unavailable in closed form, a tractable approximation can be obtained. Therefore, in situations where η_t fails to adequately capture the target process dynamics, we propose to further subtract an approximation of the conditional expectation $\rho_t = \mathrm{E}(R_t|r_0, y_T)$, which we denote by $\hat{\rho}_t = \mathrm{E}(\hat{R}_t|r_0, y_T)$. Here, $\{\hat{R}_t, t \in [0,T]\}$ is obtained through the linear noise approximation (LNA) of (3.23) (see Chapter 2).

Recall

$$d\hat{R}_t = H_t \hat{R}_t \, dt + \sqrt{\beta\left(\eta_t\right)} \, dW_t,$$

with η_t satisfying (3.22) and H_t being the Jacobian matrix with (i,j)th element

$$\left(H_t\right)_{i,j} = \frac{\partial \alpha_i\left(\eta_t\right)}{\partial \eta_{j,t}}.$$

For a fixed initial condition $\hat{R}_0 = \hat{r}_0$, we have from (2.34) that

$$\hat{R}_t|\hat{R}_0 = \hat{r}_0 \sim N\left(P_t \hat{r}_0 \, , \, P_t \psi_t P_t'\right), \tag{3.27}$$

where P_t and ψ_t satisfy the ODE system

$$\frac{dP_t}{dt} = H_t P_t, \quad P_0 = I_d, \tag{3.28}$$

$$\frac{d\psi_t}{dt} = P_t^{-1}\beta\left(\eta_t\right)\left(P_t^{-1}\right)', \quad \psi_0 = 0, \tag{3.29}$$

where $\psi_0 = 0$ is a $d \times d$ matrix with all elements equal to zero.

Along with (3.27) we have that

$$\hat{R}_t|\hat{R}_s \sim N\left(P_{t|s}\hat{R}_s\,,\, P_{t|s}\psi_{t|s}P_{t|s}'\right), \tag{3.30}$$

where $P_{t|s}$ and $\psi_{t|s}$ are found by integrating (3.28) and (3.29) from s to t with $P_s = I_d$, $\psi_s = 0$, but the ODE for η_t (3.22) is not restarted. Now we can write

$$\hat{R}_t = P_{t|s}\hat{R}_s + \epsilon_t, \qquad \epsilon_t \sim N\left(0, P_{t|s}\psi_{t|s}P_{t|s}'\right). \tag{3.31}$$

Hence

$$\mathrm{E}\left(\hat{R}_t\right) = P_{t|s}\mathrm{E}\left(\hat{R}_s\right) = P_{t|s}P_s\hat{R}_0 = P_t\hat{R}_0, \tag{3.32}$$

which implies that

$$P_{t|s} = P_t P_s^{-1}. \tag{3.33}$$

Furthermore,

$$\begin{aligned}\mathrm{Cov}\left(\hat{R}_t, \hat{R}_s\right) &= \mathrm{Cov}\left(P_{t|s}\hat{R}_s, \hat{R}_s\right)\\ &= P_{t|s}\mathrm{Var}\left(\hat{R}_s\right)\\ &= P_{t|s}P_s\psi_s P_s'.\end{aligned}$$

Using (3.33) gives

$$\mathrm{Cov}\left(\hat{R}_t, \hat{R}_s\right) = P_t\psi_s P_s'. \tag{3.34}$$

A useful identity (which is used in Section 3.2.2) should be noted at this point, and states that $\psi_{t|s} = P_s\left(\psi_t - \psi_s\right)P_s'$. From (3.31) we have that

$$\begin{aligned}\mathrm{Var}\left(\hat{R}_t\right) &= P_{t|s}\mathrm{Var}\left(\hat{R}_s\right)P_{t|s}' + \mathrm{Var}\left(\epsilon_t\right)\\ &= P_{t|s}P_s\psi_s P_s'P_{t|s}' + P_{t|s}\psi_{t|s}P_{t|s}'.\end{aligned}$$

Through the use of (3.33) we obtain

$$\mathrm{Var}\left(\hat{R}_t\right) = P_t\psi_s P_t' + P_{t|s}\psi_{t|s}P_{t|s}'.$$

Hence

$$\begin{aligned} P_t\psi_t P_t' &= P_t\psi_s P_t' + P_{t|s}\psi_{t|s}P_{t|s}' \\ \implies \quad P_{t|s}\psi_{t|s}P_{t|s}' &= P_t\psi_t P_t' - P_t\psi_s P_t' \\ &= P_t\left(\psi_t - \psi_s\right)P_t' \\ \implies \quad \psi_{t|s} &= P_{t|s}^{-1}P_t\left(\psi_t - \psi_s\right)P_t'\left(P_{t|s}^{-1}\right)', \end{aligned}$$

which via (3.33) gives

$$\psi_{t|s} = P_s\left(\psi_t - \psi_s\right)P_s'. \tag{3.35}$$

Through (3.32) and (3.34) we can construct the joint distribution of $\hat{R}_t$ and $Y_T - F'\eta_T$ (conditional on $\hat{r}_0$) as

$$\begin{pmatrix} \hat{R}_t \\ Y_T - F'\eta_T \end{pmatrix} \Big| \hat{r}_0 \sim N\left\{ \begin{pmatrix} P_t\hat{r}_0 \\ F'P_T\hat{r}_0 \end{pmatrix}, \begin{pmatrix} P_t\psi_t P_t' & P_t\psi_t P_T'F \\ F'P_T\psi_t P_t' & F'P_T\psi_T P_T'F + \Sigma \end{pmatrix} \right\}. \tag{3.36}$$

Conditioning further on $y_T - F'\eta_T$ and noting that $\hat{r}_0 = r_0 = 0$ gives

$$\hat{\rho}_t = \mathrm{E}\left(\hat{R}_t | r_0, y_T\right) = P_t\psi_t P_T'F\left(F'P_T\psi_T P_T'F + \Sigma\right)^{-1}\left(y_T - F'\eta_T\right).$$

Having obtained an explicit closed-form (subject to the solution of (3.22), (3.28) and (3.29)) approximation of the expected conditioned residual process, we adopt the partition $X_t = \eta_t + \hat{\rho}_t + R_t^-$ where $\{R_t^-, t \in [0,T]\}$ is the residual stochastic process resulting from the additional decomposition of X_t. Although the SDE satisfied by R_t^- will be intractable, the joint distribution of $R_{\tau_{k+1}}^-$ and $Y_T - F'(\eta_T + \hat{\rho}_T)$ can be approximated (conditional on $r_{\tau_k}^-$) by

$$\begin{pmatrix} R_{\tau_{k+1}}^- \\ Y_T - F'(\eta_T + \hat{\rho}_T) \end{pmatrix} \Big| r_{\tau_k}^- \sim N\left\{ \begin{pmatrix} r_{\tau_k}^- + \left(\alpha_k - \delta_k^\eta - \delta_k^\rho\right)\Delta\tau \\ F'\left\{r_{\tau_k}^- + \left(\alpha_k - \delta_k^\eta - \delta_k^\rho\right)\Delta_k\right\} \end{pmatrix}, \begin{pmatrix} \beta_k\Delta\tau & \beta_k F\Delta\tau \\ F'\beta_k\Delta\tau & F'\beta_k F\Delta_k + \Sigma \end{pmatrix} \right\},$$

where again we use the chord

$$\delta_k^\rho = \frac{\hat{\rho}_{\tau_{k+1}} - \hat{\rho}_{\tau_k}}{\Delta\tau}$$

in preference to the tangent. Hence, conditioning further on $Y_T - F'(\eta_T + \hat{\rho}_T)$ we obtain

$$X_{\tau_{k+1}}|x_{\tau_k}, y_T \sim N\left\{x_{\tau_k} + \mu_{\text{RB}^-}\left(x_{\tau_k}\right)\Delta\tau\,,\ \Psi_{\text{RB}^-}\left(x_{\tau_k}\right)\Delta\tau\right\}$$

where $\Psi_{\text{RB}^-}(x_{\tau_k}) = \Psi_{\text{MDB}}(x_{\tau_k})$ and

$$\mu_{\text{RB}^-}\left(x_{\tau_k}\right) = \alpha_k + \beta_k F\left(F'\beta_k F\Delta_k + \Sigma\right)^{-1}\left[y_T - F'\left\{\eta_T + \hat{\rho}_T + r^-_{\tau_k} + \left(\alpha_k - \delta^\eta_k - \delta^\rho_k\right)\Delta_k\right\}\right]. \tag{3.37}$$

Note that in the case of known x_T, $\Psi^*_{\text{RB}^-}(x_{\tau_k}) = \Psi^*_{\text{MDB}}(x_{\tau_k})$ and (3.37) becomes

$$\mu^*_{\text{RB}^-}\left(x_{\tau_k}\right) = \delta^\eta_k + \delta^\rho_k + \frac{(x_T - x_{\tau_k}) - (\eta_T - \eta_{\tau_k}) - (\hat{\rho}_T - \hat{\rho}_{\tau_k})}{T - \tau_k}.$$

3.2.2 Guided proposals

Methods based on guided diffusion processes have been examined by, for example, Clark (1990), Delyon and Hu (2006), Papaspiliopoulos and Roberts (2012) and Schauer et al. (2016). For known x_T, van der Meulen and Schauer (2015) (see also Schauer et al. (2016)) derive a bridge construct which they term a *guided proposal* (GP). They take the SDE satisfied by the conditioned process $\{X_t, t \in [0,T]\}$ in (3.14) and (3.15) but replace the intractable $p(x_T|x_t)$ with the transition density associated with a class of linear processes $\{\hat{X}_t, t \in [0,T]\}$ satisfying

$$d\hat{X}_t = B(t)\hat{X}_t\,dt + b(t)\,dt + \sqrt{\sigma(t)}\,dW_t, \quad \hat{X}_0 = x_. \tag{3.38}$$

Here, $B(t)$ and $\sigma(t)$ are $d \times d$ matrices and $b(t)$ is a d-vector. Note that the LNA (see Sections 2.7 or 3.2.1) satisfies (3.38) with $B(t) = H_t$, $b(t) = \alpha(\eta_t) - H_t\eta_t$ and $\sigma(t) = \beta(\eta_t)$.

The guided proposal can be extended to the Gaussian additive noise regime in (3.2) by noting that in this case, the drift in (3.15) becomes

$$\tilde{\alpha}(X_t) = \alpha\left(X_t\right) + \beta\left(X_t\right)\nabla_{x_t}\log p\left(y_T|x_t\right). \tag{3.39}$$

Given a tractable approximation of $p(y_T|x_t)$, the Euler-Maruyama approximation of (3.14) can be applied over the discretisation of $[0,T]$ to give a proposal density of the form (3.8) with $\mu_{\text{GP}}(x_{\tau_k}) = \tilde{\alpha}(x_{\tau_k})$ and $\Psi_{\text{GP}}(x_{\tau_k}) = \beta_k$.

We will approximate $p(y_T|x_t)$ using the LNA. Using the partition $\hat{X}_t = \eta_t + \hat{R}_t$ and combining the transition density of $\hat{R}_t$ in (3.27) with the observation regime defined in (3.2) gives

$$\hat{p}\left(y_T|x_t\right) = N\left(y_T\,;\ F'\left\{\eta_T + P_{T|t}\left(x_t - \eta_t\right)\right\}, F'P_{T|t}\psi_{T|t}P'_{T|t}F + \Sigma\right),$$

where $P_{T|t}$ and $\psi_{T|t}$ are found by integrating the ODE system in (3.28) and (3.29) from t to T with $P_{t|t} = I_d$ and $\psi_{t|t} = 0$. Clearly

$$\log \hat{p}\left(y_T|x_t\right) = constant - \frac{1}{2}\left[y_T - F'\left\{\eta_T + P_{T|t}\left(x_t - \eta_t\right)\right\}\right]' \left(F'P_{T|t}\psi_{T|t}P'_{T|t}F + \Sigma\right)^{-1} \times \left[y_T - F'\left\{\eta_T + P_{T|t}\left(x_t - \eta_t\right)\right\}\right],$$

and therefore

$$\nabla_{x_t} \log \hat{p}\left(y_T|x_t\right) = P'_{T|t}F\left(F'P_{T|t}\psi_{T|t}P'_{T|t}F + \Sigma\right)^{-1}\left[y_T - F'\left\{\eta_T + P_{T|t}\left(x_t - \eta_t\right)\right\}\right].$$

Hence the drift (3.39) becomes

$$\tilde{\alpha}\left(X_t\right) = \alpha\left(X_t\right) + \beta\left(X_t\right)P'_{T|t}F\left(F'P_{T|t}\psi_{T|t}P'_{T|t}F + \Sigma\right)^{-1}\left[y_T - F'\left\{\eta_T + P_{T|t}\left(x_t - \eta_t\right)\right\}\right]. \tag{3.40}$$

Note that a computationally efficient implementation of this approach is obtained by using the identities $P_{T|t} = P_T P_t^{-1}$ (see (3.33)) and $\psi_{T|t} = P_t(\psi_T - \psi_t)P'_t$ (see (3.35)). Hence, the LNA ODEs in (3.22), (3.28) and (3.29) need only be integrated *once* over the interval $[0, T]$. Unfortunately, we find that this approach does not work well in practice, unless the total measurement error $\mathrm{tr}(\Sigma)$ is large relative to the infinitesimal variance $\beta(\cdot)$. Note that the variance of $Y_T|x_t$ under the LNA is a function of the deterministic process η_t. If η_t and x_t diverge as t is increased, the guiding term in (3.40) will result in an over or under dispersed proposal mechanism (relative to the target conditioned process) at times close to T. The problem is exacerbated in the case of no measurement error, where the discrepancy between x_t and η_t can result in a singularity in the guiding term in (3.40) at time T. This naive approach (henceforth referred to as GP-N) can be alleviated by integrating the ODE system given by (3.22), (3.28) and (3.29) *for each interval* $[\tau_k, T]$, $k = 0, 1, \ldots, m-1$, with $\eta_{\tau_k} = x_{\tau_k}$. In this case, the drift (3.39) is given by

$$\tilde{\alpha}\left(X_t\right) = \alpha\left(X_t\right) + \beta\left(X_t\right)P'_{T|t}F\left(F'P_{T|t}\psi_{T|t}P'_{T|t}F + \Sigma\right)^{-1}\left(y_T - F'\eta_T\right).$$

Explicitly, we have that $\Psi_{\text{GP-N}}(x_{\tau_k}) = \Psi_{\text{GP}}(x_{\tau_k}) = \beta_k$,

$$\mu_{\text{GP-N}}\left(x_{\tau_k}\right) = \alpha_k + \beta_k P'_{T|\tau_k}F\left(F'P_{T|\tau_k}\psi_{T|\tau_k}P'_{T|\tau_k}F + \Sigma\right)^{-1}\left[y_T - F'\left\{\eta_T + P_{T|\tau_k}\left(x_{\tau_k} - \eta_{\tau_k}\right)\right\}\right]$$

and

$$\mu_{\text{GP}}\left(x_{\tau_k}\right) = \alpha_k + \beta_k P'_{T|\tau_k}F\left(F'P_{T|\tau_k}\psi_{T|\tau_k}P'_{T|\tau_k}F + \Sigma\right)^{-1}\left(y_T - F'\eta_T\right).$$

In the special case that x_T is known, we have that $\Psi^*_{\text{GP-N}}(x_{\tau_k}) = \Psi^*_{\text{GP}}(x_{\tau_k}) = \beta_k$,

$$\mu^*_{\text{GP-N}}(x_{\tau_k}) = \alpha_k + \beta_k P'_{T|\tau_k}\left(P_{T|\tau_k}\psi_{T|\tau_k}P'_{T|\tau_k}\right)^{-1}\left[x_T - \left\{\eta_T + P_{T|\tau_k}(x_{\tau_k} - \eta_{\tau_k})\right\}\right]$$

and

$$\mu^*_{\text{GP}}(x_{\tau_k}) = \alpha_k + \beta_k P'_{T|\tau_k}\left(P_{T|\tau_k}\psi_{T|\tau_k}P'_{T|\tau_k}\right)^{-1}(x_T - \eta_T).$$

The limiting (as $\Delta\tau \to 0$) form of the acceptance rate in this case can be found in Schauer et al. (2016), who also remark that a key requirement for absolute continuity of the target and proposal process is that $\sigma(T) = \beta(x_T)$. For the LNA, we have $\sigma(t) = \beta(\eta_t)$. Again, we note that the naive implementation of the guided proposal (GP-N) will not meet this condition in general (when x_T is known). Ensuring that $\sigma(t) \to \beta(x_T)$ as $t \to T$ by integrating (3.22), (3.28) and (3.29) for each τ_k is likely to be time consuming, unless the LNA ODE system is tractable. In the case of exact observations only, a computationally less demanding approach is obtained in van der Meulen and Schauer (2015) by taking the transition density of (3.38) with $B(t) = 0$ and $\sigma(t) = \beta(x_T)$ to construct the guided proposal. Setting $b(t) = \alpha(\eta_t)$ leads to a proposal density for the simplified guided proposal (GP-S) of the form (3.8) with $\Psi^*_{\text{GP-S}}(x_{\tau_k}) = \beta_k$ and

$$\mu^*_{\text{GP-S}}(x_{\tau_k}) = \alpha_k + \beta_k\beta(x_T)^{-1}\left\{\frac{x_T - x_{\tau_k} - (\eta_T - \eta_{\tau_k})}{T - \tau_k}\right\}.$$

Further (example-dependent) methods for constructing guided proposals in the case of known x_T can be found in van der Meulen and Schauer (2015).

Use of the MDB variance

Using the Euler-Maruyama approximation of (3.14) gives the variance of $X_{\tau_{k+1}}|x_{\tau_k}, y_T$ in the guided proposal process as $\Psi_{\text{GP}}(x_{\tau_k})\Delta\tau = \beta_k\Delta\tau$. In Section 3.5 we investigate the effect of using the variance (3.13) of the modified diffusion bridge construct by taking $\Psi_{\text{GP}}(x_{\tau_k}) = \Psi_{\text{MDB}}(x_{\tau_k})$. Although in this case, deriving the limiting form of the acceptance rate under the resulting proposal is problematic, we observe a worthwhile increase in empirical performance (see Section 3.5). In the case of known x_T, use of the MDB variance in place of $\beta_k\Delta\tau$ comes at almost no additional computational cost. We denote this *diffusion bridge* by GP-MDB.

3.3 Computational considerations

For the observation regime in (3.2), where we observe d_o components, all bridge constructs (with the exception of the myopic approach) require the inversion of a $d_o \times d_o$ matrix at

each intermediate time τ_k, $k = 1, 2 \ldots, m-1$ and for each skeleton bridge required. For known x_T, the proposal densities associated with each construct simplify. In this case, only the LNA-based residual bridge (RB^-) and guided proposal require the inversion of a $d \times d$ matrix at each intermediate time.

The Lindström bridge (LB) and modified diffusion bridge (MDB) have roughly the same computational cost. The bridges based on residual processes incur an additional computational cost of having to solve a system of either d (when subtracting η_t) or order d^2 (when further subtracting ρ_t) coupled ODEs. However, we note that for known x_0, the ODE system need only be solved once, irrespective of the number of skeleton bridges required. This is also true of the naive and simplified guided proposals. However, we note that in the case of known x_T, the guided proposal requires solving order d^2 ODEs over each interval $[\tau_k, T]$, $k = 0, 1, \ldots, m-1$ for each simulated skeleton bridge, in order to maintain reasonable statistical efficiency (as measured by, for example, the estimated acceptance rate of a Metropolis-Hastings independence sampler).

3.4 Summary of bridge constructs

All the *diffusion bridges* discussed above are formed using an Euler-Maruyama approximation over the partition (3.3), taking the form $N\left(x_{\tau_{k+1}}\,;\, x_{\tau_k} + \mu(x_{\tau_k})\Delta\tau\,,\, \Psi(x_{\tau_k})\Delta\tau\right)$, or indeed $\mu^*(x_{\tau_k})$ and $\Psi^*(x_{\tau_k})$ in the special case of known x_T. Summaries of $(\mu(x_{\tau_k}), \Psi(x_{\tau_k}))$ and $(\mu^*(x_{\tau_k}), \Psi^*(x_{\tau_k}))$ for each construct are presented in Table 3.1 and Table 3.2 respectively. GP-S appears only in Table 3.2 as it applies only in the special case when x_T is known. We note the following definitions, $\alpha_k = \alpha(x_{\tau_k})$, $\beta_k = \beta(x_{\tau_k})$, $\Delta_k = T - \tau_k$, $\Delta_k^\gamma = \Delta_k + \gamma(\Delta_{k+1})^2/\Delta\tau$,

$$\delta_k^\eta = \frac{\eta_{\tau_{k+1}} - \eta_{\tau_k}}{\Delta\tau}, \quad \delta_k^\rho = \frac{\hat{\rho}_{\tau_{k+1}} - \hat{\rho}_{\tau_k}}{\Delta\tau} \quad \text{and} \quad w_k^\gamma = \frac{(\tau_{k+1} - \tau_k)\,(T - \tau_k)}{(\tau_{k+1} - \tau_k)\,(T - \tau_k) + \gamma\,(T - \tau_{k+1})^2}.$$

3.5 Bridge construct performance

We now compare the accuracy and efficiency of the bridge constructs discussed in the previous sections, by using them to make proposals inside a Metropolis-Hastings independence sampler (see Algorithm 1). We consider three examples: a simple birth-death model in which the ODEs governing the LNA are tractable, a Lotka-Volterra system in which the use of numerical solvers are required, and a model of aphid growth inspired by real data taken from Matis et al. (2008). Generating discrete-time realisations from the SDE model of aphid growth is particularly challenging due to nonlinear dynamics, and

Bridge construct	$\mu(x_{\tau_k})$
EM	α_k
MDB	$\alpha_k + \beta_k F \left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} \left\{y_T - F'\left(x_{\tau_k} + \alpha_k\Delta_k\right)\right\}$
LB	$\alpha_k + \beta_k F \left(F'\beta_k F\Delta_k^{\gamma} + \Sigma\right)^{-1} \left\{y_T - F'\left(x_{\tau_k} + \alpha_k\Delta_k\right)\right\}$
RB	$\alpha_k + \beta_k F \left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} \left[y_T - F'\left\{\eta_T + r_{\tau_k} + \left(\alpha_k - \delta_k^{\eta}\right)\Delta_k\right\}\right]$
RB$^-$	$\alpha_k + \beta_k F \left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} \left[y_T - F'\left\{\eta_T + \hat{\rho}_T + r_{\tau_k}^{-} + \left(\alpha_k - \delta_k^{\eta} - \delta_k^{\rho}\right)\Delta_k\right\}\right]$
GP	$\alpha_k + \beta_k P'_{T\|\tau_k} F \left(F'P_{T\|\tau_k}\psi_{T\|\tau_k}P'_{T\|\tau_k}F + \Sigma\right)^{-1} \left(y_T - F'\eta_T\right)$
GP-N	$\alpha_k + \beta_k P'_{T\|\tau_k} F \left(F'P_{T\|\tau_k}\psi_{T\|\tau_k}P'_{T\|\tau_k}F + \Sigma\right)^{-1} \left[y_T - F'\left\{\eta_T + P_{T\|\tau_k}\left(x_{\tau_k} - \eta_{\tau_k}\right)\right\}\right]$
GP-MDB	$\alpha_k + \beta_k P'_{T\|\tau_k} F \left(F'P_{T\|\tau_k}\psi_{T\|\tau_k}P'_{T\|\tau_k}F + \Sigma\right)^{-1} \left(y_T - F'\eta_T\right)$
	$\Psi(x_{\tau_k})$
EM	β_k
MDB	$\beta_k - \beta_k F \left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} F'\beta_k\Delta\tau$
LB	$\beta_k - \beta_k F \left(F'\beta_k F\Delta_k^{\gamma} + \Sigma\right)^{-1} F'\beta_k\Delta\tau$
RB	$\beta_k - \beta_k F \left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} F'\beta_k\Delta\tau$
RB$^-$	$\beta_k - \beta_k F \left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} F'\beta_k\Delta\tau$
GP	β_k
GP-N	β_k
GP-MDB	$\beta_k - \beta_k F \left(F'\beta_k F\Delta_k + \Sigma\right)^{-1} F'\beta_k\Delta\tau$

Table 3.1: Summaries of $\mu(x_{\tau_k})$ and $\Psi(x_{\tau_k})$.

an observation regime in which only one component is observed and is subject to additive Gaussian noise.

In what follows, all results are based on 100K iterations of a Metropolis-Hastings independence sampler targeting either (3.5) or (3.6), depending on the observation regime. We measure the statistical efficiency of each bridge via their empirical acceptance probability. Note that these empirical acceptance probabilities are reasonably accurate, with repeated implementations of the independence sampler typically leading to only small differences in the third decimal place. `R` code for the implementation of the Metropolis-Hastings scheme can be found at *https://github.com/gawhitaker/bridges-apps*

Bridge construct	$\mu^*(x_{\tau_k})$	$\Psi^*(x_{\tau_k})$
EM	α_k	β_k
MDB	$\frac{x_T - x_{\tau_k}}{T-\tau_k}$	$\frac{T-\tau_{k+1}}{T-\tau_k}\beta_k$
LB	$w_k^\gamma \frac{x_T - x_{\tau_k}}{T-\tau_k} + (1-w_k^\gamma)\,\alpha_k$	$w_k^\gamma \frac{T-\tau_{k+1}}{T-\tau_k}\beta_k + (1-w_k^\gamma)\,\beta_k$
RB	$\delta_k^\eta + \frac{(x_T - x_{\tau_k}) - (\eta_T - \eta_{\tau_k})}{T-\tau_k}$	$\frac{T-\tau_{k+1}}{T-\tau_k}\beta_k$
RB$^-$	$\delta_k^\eta + \delta_k^\rho + \frac{(x_T - x_{\tau_k}) - (\eta_T - \eta_{\tau_k}) - (\hat{\rho}_T - \hat{\rho}_{\tau_k})}{T-\tau_k}$	$\frac{T-\tau_{k+1}}{T-\tau_k}\beta_k$
GP	$\alpha_k + \beta_k P'_{T\|\tau_k}\left(P_{T\|\tau_k}\psi_{T\|\tau_k}P'_{T\|\tau_k}\right)^{-1}(x_T - \eta_T)$	β_k
GP-N	$\alpha_k + \beta_k P'_{T\|\tau_k}\left(P_{T\|\tau_k}\psi_{T\|\tau_k}P'_{T\|\tau_k}\right)^{-1} \times \left[x_T - \left\{\eta_T + P_{T\|\tau_k}\left(x_{\tau_k} - \eta_{\tau_k}\right)\right\}\right]$	β_k
GP-S	$\alpha_k + \beta_k \beta\left(x_T\right)^{-1}\left\{\frac{x_T - x_{\tau_k} - (\eta_T - \eta_{\tau_k})}{T-\tau_k}\right\}$	β_k
GP-MDB	$\alpha_k + \beta_k P'_{T\|\tau_k}\left(P_{T\|\tau_k}\psi_{T\|\tau_k}P'_{T\|\tau_k}\right)^{-1}(x_T - \eta_T)$	$\frac{T-\tau_{k+1}}{T-\tau_k}\beta_k$

Table 3.2: Summaries of $\mu^*(x_{\tau_k})$ and $\Psi^*(x_{\tau_k})$.

The bridge constructs used in each example, together with their relative computational cost can be found in Table 3.3. Note that in contrast to Lindström (2012), we found that $\gamma \in [0.001, 0.3]$ was required in order to find a near-optimal γ. Where LB is used, we only present results for the value of γ that maximised empirical performance.

3.5.1 Birth-death model

We consider a simple birth-death process with birth rate θ_1 and death rate θ_2, characterised by the SDE

$$dX_t = (\theta_1 - \theta_2)\, X_t\, dt + \sqrt{(\theta_1 + \theta_2)\, X_t}\, dW_t, \quad X_0 = x_0, \tag{3.41}$$

which can be seen as a degenerate case of a Feller square-root diffusion (Feller, 1952). Here, $H_t = (\theta_1 - \theta_2)$ and the ODE system ((3.22), (3.28) and (3.29)) governing the linear noise approximation of (3.41) is given by

$$\frac{d\eta_t}{dt} = (\theta_1 - \theta_2)\,\eta_t,$$
$$\frac{dP_t}{dt} = (\theta_1 - \theta_2)\,P_t,$$

		Birth-death	Lotka-Volterra	Aphid
Myopic Euler-Maruyama	(EM)	–	–	1.0
Modified diffusion bridge	(MDB)	1.0	1.0	–
Lindström bridge	(LB)	1.1	1.1	–
Residual bridge, subtract η_t	(RB)	1.0	1.0	7.3
RB, further subtract ρ_t	(RB^-)	1.0	1.0	7.9
Guided proposal	(GP)	1.2	30.7	7.1
GP with MDB variance	(GP-MDB)	1.3	31.0	7.9
Naive GP	(GP-N)	1.2	–	–
Simplified GP	(GP-S)	1.1	–	–

Table 3.3: Example and bridge specific relative CPU cost for 100K iterations of a Metropolis-Hastings independence sampler. Due to well known poor performance in the case of known x_T, EM is not implemented for the first two examples. Likewise, due to poor performance, we omit results based on GP-N and GP-S in the second example, and results based on MDB and LB in the final example.

$$\frac{d\psi_t}{dt} = P_t^{-1}\left(\theta_1+\theta_2\right)\eta_t\left(P_t^{-1}\right)'.$$

For this model the system of ODEs is tractable, and we obtain $\eta_t = x_0 e^{(\theta_1-\theta_2)t}$, $P_t = e^{(\theta_1-\theta_2)t}$ and

$$\psi_t = \frac{\left(\theta_1+\theta_2\right)x_0}{\theta_1-\theta_2}\left(1-e^{-(\theta_1-\theta_2)t}\right).$$

Derivations of these solutions can be found in Appendix B.

In this example we assume that x_T is known and to adequately assess the performance of each bridge construct, we take x_T to be either the 5%, 50% or 95% quantile (denoted by $x_{T,(5)}$, $x_{T,(50)}$ and $x_{T,(95)}$ respectively) of $X_T|X_0 = x_0$, found by repeatedly applying the Euler-Maruyama approximation to (3.41) with a small time-step ($\Delta t = 0.01$). To allow for different inter-observation intervals, we take $T \in \{1, 2\}$. An initial condition of $x_0 = 50$ and parameter values $\theta = (0.1, 0.8)'$ gives $(x_{1,(5)}, x_{1,(50)}, x_{1,(95)}) = (18.49, 24.62, 31.68)$ and $(x_{2,(5)}, x_{2,(50)}, x_{2,(95)}) = (6.97, 12.00, 18.35)$. Note that here, the parameter choice leads to a moribund system.

Since the ODE system governing the LNA is tractable for this example, there is little difference in CPU cost between the bridges (see Table 3.3). Therefore, we use statistical efficiency (as measured by empirical Metropolis-Hastings acceptance probabilities) as a proxy for overall efficiency of each bridge, with higher probabilities preferred. When there is a difference in CPU time between the bridges we propose to further assess overall efficiency via the minimum effective sample size (see Section 3.5.2).

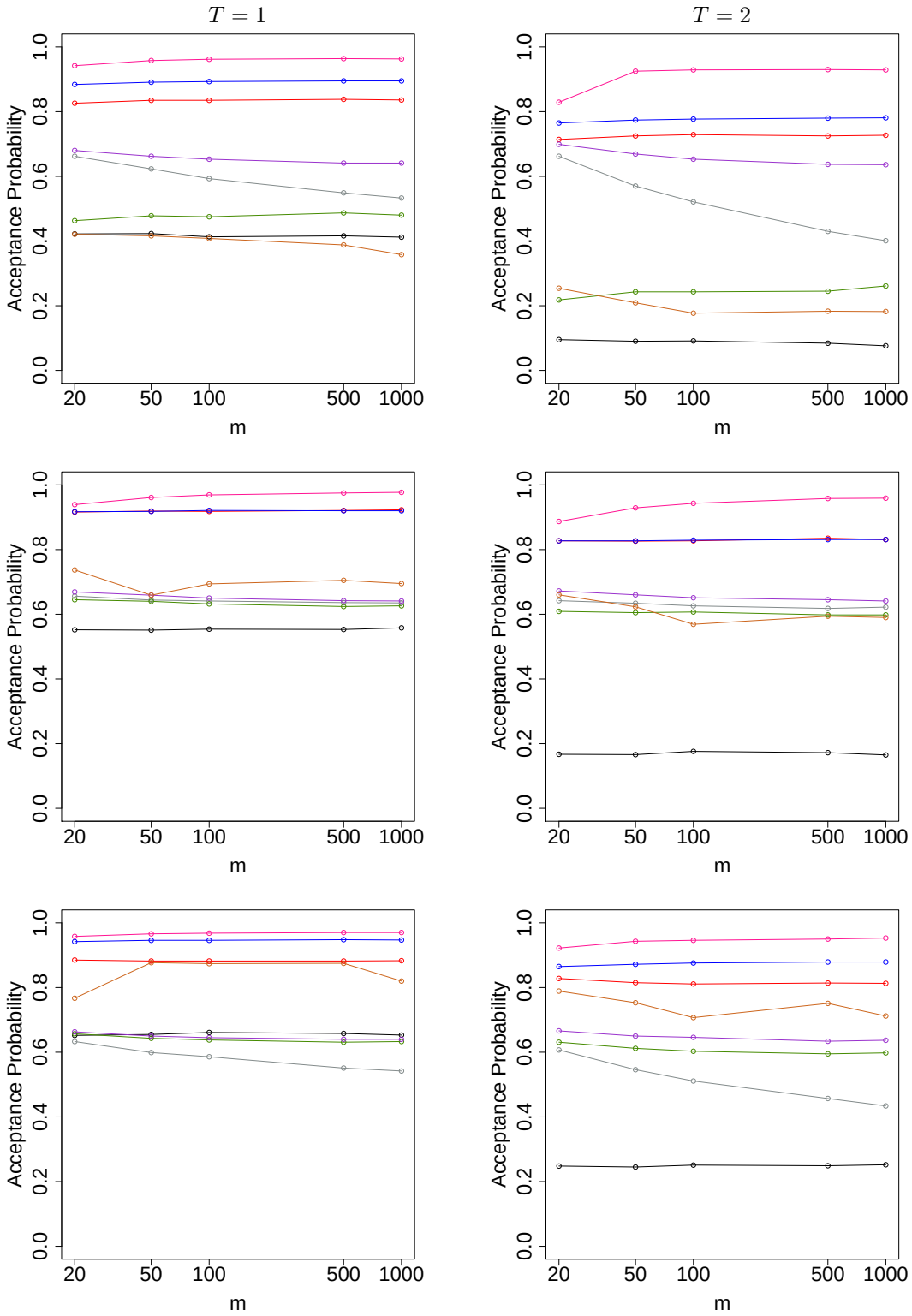

Figure 3.2: Birth-death model. Empirical acceptance probability against m with $x_T = x_{T,(5)}$ (1^{st} row), $x_T = x_{T,(50)}$ (2^{nd} row) and $x_T = x_{T,(95)}$ (3^{rd} row). The results are based on 100K iterations of a Metropolis-Hastings independence sampler. Black: MDB. Brown: LB. Red: RB. Blue: RB$^-$. Grey: GP-N. Green: GP-S. Purple: GP. Pink: GP-MDB.

Figure 3.2 shows empirical acceptance probabilities against the number of sub-intervals m for each bridge and each x_T. The empirical acceptance probabilities are also presented in Tables 3.4 and 3.5 for $T = 1$ and $T = 2$ respectively. Figures 3.3 and 3.4 compare 95% credible regions of the proposal under various bridging strategies with the true conditioned process (obtained from the output of the Metropolis-Hastings independence sampler). It is clear from the figures that as T is increased, the MDB fails to adequately account for the nonlinear behaviour of the conditioned process. Indeed, in terms of empirical acceptance rate, MDB is outperformed by all other bridges for $T = 2$. As m is increased so that the discretisation gets finer, the acceptance rates under all bridges (with the exception of GP-N) stay roughly constant. For GP-N, the acceptance rates decrease with m when x_T is either the 5% or 95% quantile of $X_T|X_0 = 50$. In this case, the variance associated with the approximate transition density either overestimates (when x_T is the 5% quantile) or underestimates (when x_T is the 95% quantile) the true variance at the end-point. For example, when x_T is the 95% quantile, this results (see Figure 3.4) in a 'tapering in' of the proposal relative to the true conditioned process. GP-S, GP and LB give similar performance, although we note that GP-S and LB perform particularly poorly when x_T is the 5% quantile. Moreover, LB requires the specification of a tuning parameter γ and we found that the acceptance rate was fairly sensitive to the choice of γ. In all scenarios, RB, RB$^-$ and GP-MDB comprehensively outperform all other bridge constructs. When x_T is the median of $X_T|X_0 = 50$, we see that RB and RB$^-$ (red and blue lines in Figure 3.2) give near identical performance, with η_t adequately accounting for the observed nonlinear dynamics. In terms of statistical efficiency, GP-MDB outperforms both RB and RB$^-$ in all scenarios, although the relative difference is small.

3.5.2 Lotka-Volterra model

In this example we consider a Lotka-Volterra model of predator-prey dynamics. We denote the system state at time t by $X_t = (X_{1,t}, X_{2,t})'$, ordered as prey, predators. The mass-action SDE representation of system dynamics with initial condition $X_0 = x_0$ takes the form

$$dX_t = \begin{pmatrix} \theta_1 X_{1,t} - \theta_2 X_{1,t} X_{2,t} \\ \theta_2 X_{1,t} X_{2,t} - \theta_3 X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \theta_1 X_{1,t} + \theta_2 X_{1,t} X_{2,t} & -\theta_2 X_{1,t} X_{2,t} \\ -\theta_2 X_{1,t} X_{2,t} & \theta_3 X_{2,t} + \theta_2 X_{1,t} X_{2,t} \end{pmatrix}^{\frac{1}{2}} dW_t. \tag{3.42}$$

The components of $\theta = (\theta_1, \theta_2, \theta_3)'$ can be interpreted as prey reproduction rate, prey death and predator reproduction rate, and predator death. For this model

$$H_t = \begin{pmatrix} \theta_1 - \theta_2 \eta_{2,t} & -\theta_2 \eta_{1,t} \\ \theta_2 \eta_{2,t} & \theta_2 \eta_{1,t} - \theta_3 \end{pmatrix},$$

Bridge	$m = 20$	$m = 50$	$m = 100$	$m = 500$	$m = 1000$
	(0.422)	(0.423)	(0.413)	(0.416)	(0.412)
MDB	0.552	0.551	0.554	0.553	0.558
	[0.652]	[0.655]	[0.661]	[0.658]	[0.653]
	(0.826)	(0.835)	(0.835)	(0.838)	(0.836)
RB	0.916	0.919	0.918	0.921	0.923
	[0.885]	[0.882]	[0.882]	[0.882]	[0.883]
	(0.884)	(0.891)	(0.893)	(0.895)	(0.895)
RB$^-$	0.917	0.918	0.921	0.920	0.920
	[0.942]	[0.946]	[0.946]	[0.948]	[0.947]
	(0.662)	(0.623)	(0.593)	(0.549)	(0.533)
GP-N	0.656	0.644	0.641	0.636	0.635
	[0.633]	[0.599]	[0.586]	[0.551]	[0.542]
	(0.463)	(0.478)	(0.475)	(0.487)	(0.480)
GP-S	0.645	0.640	0.632	0.624	0.626
	[0.657]	[0.643]	[0.638]	[0.631]	[0.633]
	(0.680)	(0.662)	(0.653)	(0.641)	(0.641)
GP	0.669	0.659	0.650	0.642	0.641
	[0.663]	[0.650]	[0.645]	[0.640]	[0.640]
	(0.942)	(0.958)	(0.962)	(0.964)	(0.963)
GP-MDB	0.939	0.961	0.969	0.975	0.977
	[0.958]	[0.966]	[0.968]	[0.970]	[0.970]
	(0.421/0.0025)	(0.416/0.001)	(0.408/0.001)	(0.388/0.001)	(0.358/0.001)
LB/γ	0.737/0.1	0.659/0.1	0.694/0.01	0.705/0.0025	0.695/0.0025
	[0.767/0.01]	[0.877/0.01]	[0.874/0.005]	[0.875/0.001]	[0.820/0.001]

Table 3.4: Birth-death model. Empirical acceptance probability against m with ($x_1 = x_{1,(5)}$), $x_1 = x_{1,(50)}$ and $[x_1 = x_{1,(95)}]$. The results are based on 100K iterations of a Metropolis-Hastings independence sampler.

and hence, the ODE system ((3.22), (3.28) and (3.29)) governing the linear noise approximation of (3.42) is

$$\frac{d\eta_t}{dt} = \begin{pmatrix} \theta_1\eta_{1,t} - \theta_2\eta_{1,t}\eta_{2,t} \\ \theta_2\eta_{1,t}\eta_{2,t} - \theta_3\eta_{2,t} \end{pmatrix},$$

$$\frac{dP_t}{dt} = \begin{pmatrix} \theta_1 - \theta_2\eta_{2,t} & -\theta_2\eta_{1,t} \\ \theta_2\eta_{2,t} & \theta_2\eta_{1,t} - \theta_3 \end{pmatrix} P_t,$$

$$\frac{d\psi_t}{dt} = P_t^{-1} \begin{pmatrix} \theta_1\eta_{1,t} + \theta_2\eta_{1,t}\eta_{2,t} & -\theta_2\eta_{1,t}\eta_{2,t} \\ -\theta_2\eta_{1,t}\eta_{2,t} & \theta_3\eta_{2,t} + \theta_2\eta_{1,t}\eta_{2,t} \end{pmatrix} \left(P_t^{-1}\right)'.$$

Bridge	$m = 20$	$m = 50$	$m = 100$	$m = 500$	$m = 1000$
	(0.095)	(0.090)	(0.091)	(0.084)	(0.076)
MDB	0.167	0.166	0.176	0.172	0.165
	[0.248]	[0.245]	[0.251]	[0.249]	[0.252]
	(0.714)	(0.725)	(0.729)	(0.725)	(0.727)
RB	0.827	0.826	0.827	0.835	0.831
	[0.828]	[0.815]	[0.811]	[0.814]	[0.813]
	(0.765)	(0.774)	(0.777)	(0.780)	(0.781)
RB^-	0.827	0.827	0.829	0.831	0.831
	[0.865]	[0.872]	[0.876]	[0.879]	[0.879]
	(0.662)	(0.570)	(0.521)	(0.430)	(0.401)
GP-N	0.642	0.634	0.626	0.618	0.622
	[0.607]	[0.546]	[0.511]	[0.457]	[0.434]
	(0.218)	(0.243)	(0.243)	(0.245)	(0.261)
GP-S	0.609	0.605	0.607	0.598	0.598
	[0.631]	[0.612]	[0.603]	[0.595]	[0.598]
	(0.699)	(0.669)	(0.653)	(0.637)	(0.636)
GP	0.672	0.660	0.651	0.645	0.641
	[0.666]	[0.650]	[0.646]	[0.634]	[0.637]
	(0.829)	(0.925)	(0.929)	(0.930)	(0.929)
GP-MDB	0.887	0.929	0.943	0.958	0.959
	[0.922]	[0.943]	[0.946]	[0.950]	[0.953]
	(0.254/0.3)	(0.209/0.1)	(0.177/0.1)	(0.183/0.0075)	(0.182/0.005)
LB/γ	0.660/0.3	0.623/0.1	0.569/0.1	0.594/0.01	0.590/0.005
	[0.789/0.1]	[0.753/0.025]	[0.707/0.01]	[0.751/0.0025]	[0.712/0.0025]

Table 3.5: Birth-death model. Empirical acceptance probability against m with $(x_2 = x_{2,(5)})$, $x_2 = x_{2,(50)}$ and $[x_2 = x_{2,(95)}]$. The results are based on 100K iterations of a Metropolis-Hastings independence sampler.

Note that this ODE system is intractable and we therefore use the `R` package `lsoda` to numerically solve the system when necessary.

Following Boys et al. (2008) we adopt the parameter values $\theta = (\theta_1, \theta_2, \theta_3)' = (0.5, 0.0025, 0.3)'$ and let $x_0 = (71, 79)'$. We assume that x_T is known and generate a number of challenging scenarios by taking x_T as either the 5%, 50% or 95% marginal quantiles of $X_T|X_0 = (71, 79)'$ for $T \in \{1, 2, 3, 4\}$. These quantiles are shown in Table 3.6 and illustrated in Figure 3.5. Note that for this parameter choice, the expectation of $X_t|X_0 = (71, 79)'$ is approximately periodic with a period of around 17.

We fixed the discretisation by taking $m = 50$, but note no appreciable difference in results for finer discretisations (for example, $m = 1000$). In the previous example, GP-N and

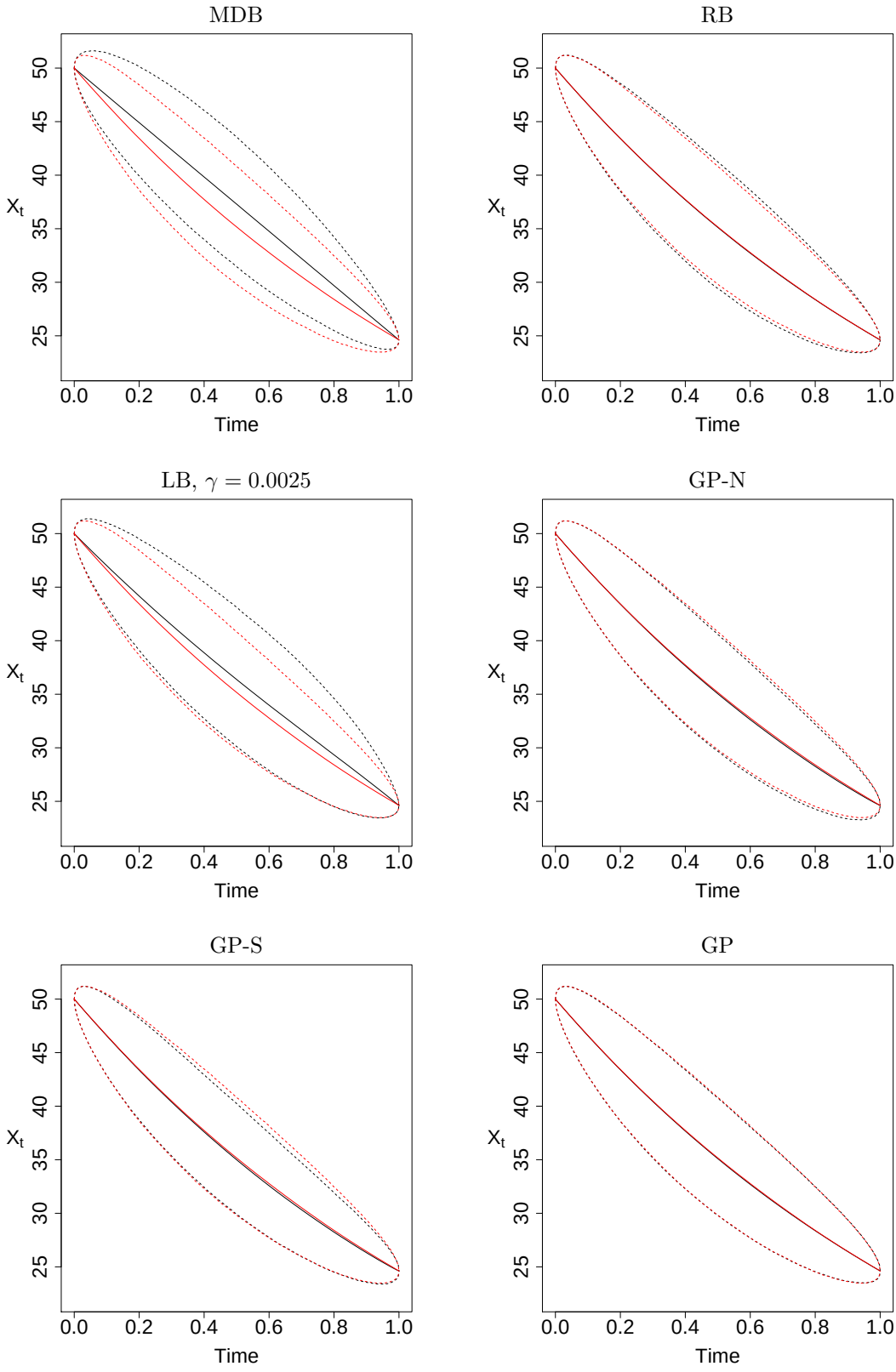

Figure 3.3: Birth-death model. 95% credible region (dashed line) and mean (solid line) of the true conditioned process (red) and various bridge constructs (black) using $x_T = x_{1,(50)}$.

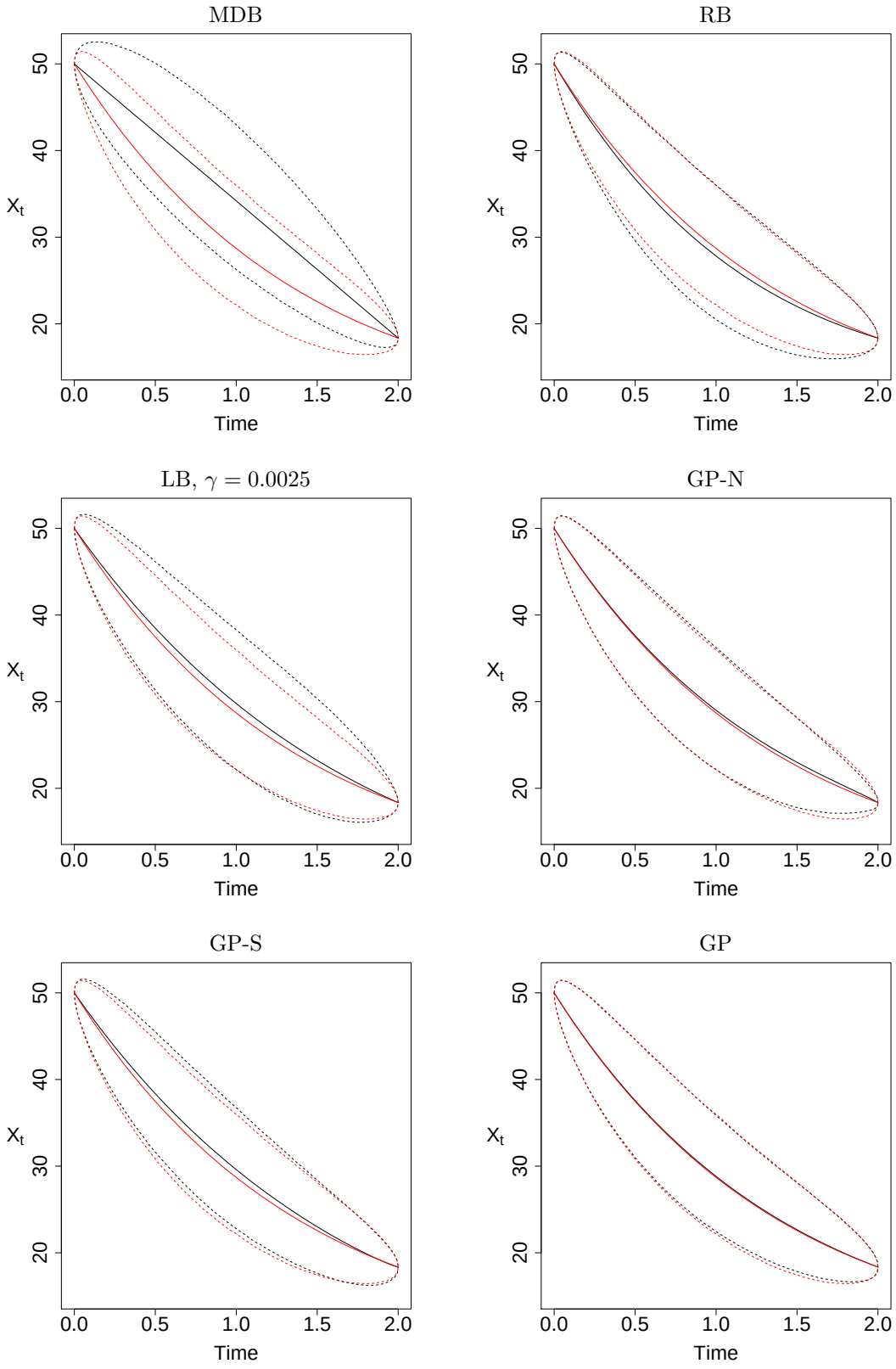

Figure 3.4: Birth-death model. 95% credible region (dashed line) and mean (solid line) of the true conditioned process (red) and various bridge constructs (black) using $x_T = x_{2,(95)}$.

	$T = 1$	$T = 2$	$T = 3$	$T = 4$
$x_{T,(5)}$	(82.47,62.78)	(107.35,57.95)	(142.00,60.02)	(185.04,71.23)
$x_{T,(50)}$	(96.82,71.93)	(133.35,70.75)	(182.64,77.36)	(242.08,97.23)
$x_{T,(95)}$	(112.13,81.58)	(162.28,84.63)	(228.82,97.12)	(308.58,128.76)

Table 3.6: Lotka-Volterra model. Quantiles of $X_T|X_0 = (71, 79)'$ found by repeatedly simulating from the Euler-Maruyama approximation of (3.42) with $\theta = (0.5, 0.0025, 0.3)'$.

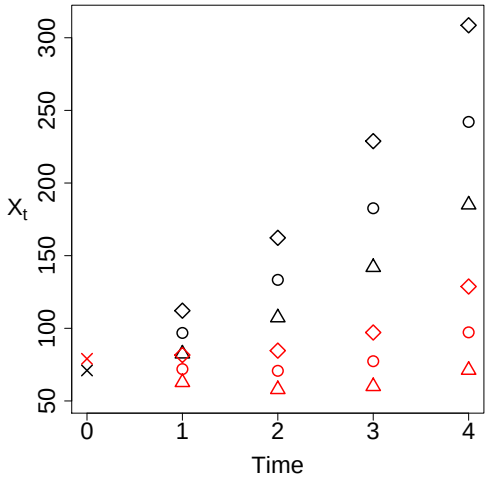

Figure 3.5: Lotka-Volterra model. Quantiles of $X_T|X_0 = (71, 79)'$ found by repeatedly simulating from the Euler-Maruyama approximation of (3.42) with $\theta = (0.5, 0.0025, 0.3)'$. The 5%, 50% and 95% quantiles are represented by triangles, circles and diamonds respectively for times $t = 1, 2, 3, 4$. Black: Prey ($X_{1,t}$). Red: Predator ($X_{2,t}$).

GP-S perform relatively poorly, and so in what follows we omit these bridges from the results. Note that we include MDB for reference. Figure 3.6 shows empirical acceptance probabilities against T for each bridge and each x_T, with the explicit values also given in Table 3.7. Figure 3.7 compares 95% credible regions of the proposal under various bridging strategies with the true conditioned process (obtained from the output of the Metropolis-Hastings independence sampler).

Unsurprisingly, as T is increased, MDB fails to adequately account for the nonlinear behaviour of the conditioned process. LB offers a modest improvement (except when $x_T = x_{T,(5)}$) but is generally outperformed by the other bridge constructs. We found that as T was increased, LB required larger values of γ, reflecting the need for more weight to be placed on the myopic component of the construct. As in the previous example, unless x_T is the median of $X_T|x_0$, RB is comprehensively outperformed by RB$^-$ (see Figure 3.7 for the effect of increasing T on RB and RB$^-$). However, we see that the acceptance probabilities are decreasing in T for both constructs. As noted by Fearnhead et al. (2014), the LNA

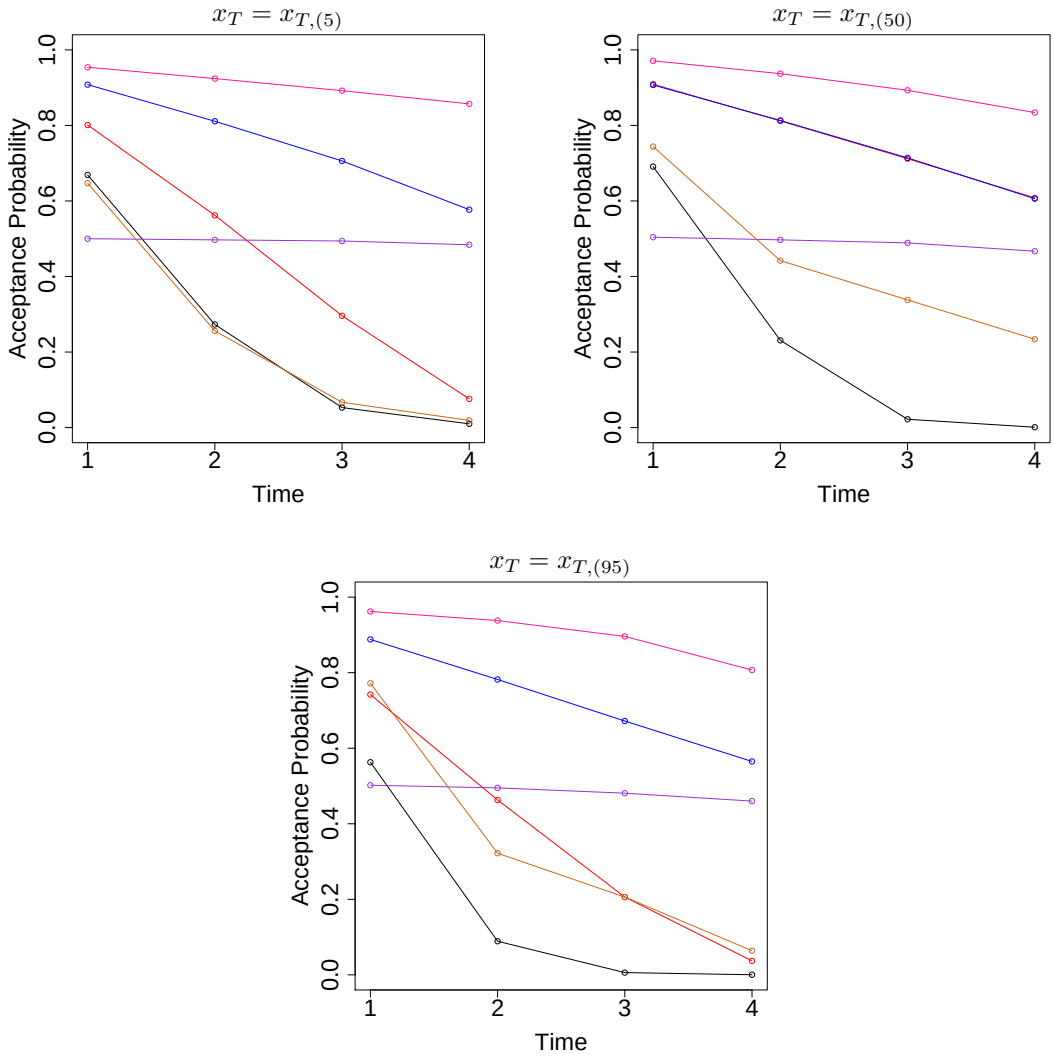

Figure 3.6: Lotka-Volterra model. Empirical acceptance probabilities against T. The results are based on 100K iterations of a Metropolis-Hastings independence sampler. Black: MDB. Brown: LB. Red: RB. Blue: RB$^-$. Purple: GP. Pink: GP-MDB.

can become poor as T increases, with the implication here being that the approximation of the expected residual (as used in RB$^-$) degrades with T.

We note that the estimated acceptance probabilities are roughly constant for GP and (to a lesser extent) GP-MDB, and in terms of statistical efficiency for a fixed number of iterations, GP-MDB should be preferred over all other algorithms considered in this chapter. However, the difference in estimated acceptance probabilities between GP-MDB and RB$^-$ is fairly small, even when $T = 4$; for example, 0.857 vs 0.577 when $x_T = x_{T,(5)}$ and 0.834 vs 0.606 when $x_T = x_{T,(50)}$. We also note that a Metropolis-Hastings scheme that uses RB or RB$^-$ is some 30 times faster than a scheme with GP or GP-MDB, since

the latter require solving the LNA ODE system for each sub-interval $[\tau_k, T]$ to maintain reasonable statistical efficiency for a given m. Therefore, we further compare RB, RB$^-$, GP and GP-MDB by computing the minimum effective sample size (minESS) at time $T/2$ (where the minimum is over each component of $X_{T/2}$) divided by CPU cost (in seconds). We denote this measure of overall efficiency by minESS/sec. The effective sample size (ESS) is equivalent to the number of independent samples, and is characterised as

$$\text{ESS} = \frac{\text{number of iterations}}{1 + \sum_{k=1}^{\infty} \rho_k},$$

where ρ_k is the lag-k autocorrelation. Table 3.8 shows the minESS/sec for all bridge constructs considered in this example for $x_T = x_{1,(5)}$, $x_T = x_{1,(50)}$, $x_T = x_{4,(5)}$ and $x_T = x_{4,(50)}$. As T increases the minESS/sec for MDB falls, due to the bridge construct failing to adequately account for the nonlinear dynamics of the conditioned process. For $x_T = x_{T,(50)}$, LB performs reasonably well, however the minESS/sec decreases when $x_T = x_{T,(5)}$. When $x_T = x_{T,(5)}$ and $T = 1$, minESS/sec scales roughly as $1 : 3 : 56 : 83$ for GP : GP-MDB : RB : RB$^-$. When $T = 4$, minESS/sec scales roughly as $1 : 3 : 1 : 17$. Hence, for this example, RB$^-$ is to be preferred in terms of overall efficiency, although the relative difference between RB$^-$ and GP-MDB appears to decrease as T is increased, consistent with the behaviour of the empirical acceptance rates observed in Figure 3.6.

3.5.3 Aphid growth model

Matis et al. (2008) describe a stochastic model for aphid dynamics in terms of population size (N_t) and cumulative population size (C_t). The diffusion approximation of their model is given by

$$\begin{pmatrix} dN_t \\ dC_t \end{pmatrix} = \begin{pmatrix} \theta_1 N_t - \theta_2 N_t C_t \\ \theta_1 N_t \end{pmatrix} dt + \begin{pmatrix} \theta_1 N_t + \theta_2 N_t C_t & \theta_1 N_t \\ \theta_1 N_t & \theta_1 N_t \end{pmatrix}^{1/2} dW_t \tag{3.43}$$

where the components of $\theta = (\theta_1, \theta_2)'$ characterise the birth and death rate respectively, and we have initial condition $X_0 = x_0$. Matis et al. (2008) also provide a dataset consisting of cotton aphid counts recorded at times $t = 0, 1.14, 2.29, 3.57$ and 4.57 weeks, and collected for 27 different treatment block combinations. The analysis of these data via a stochastic differential mixed-effects model driven by (3.43) is the focus of Whitaker et al. (2016*a*) and the subject of Section 5.2.

Driven by the real data of Matis et al. (2008) and to illustrate the proposed methodology in a challenging partial observation scenario, we assume that X_T cannot be measured

Bridge	$T = 1$	$T = 2$	$T = 3$	$T = 4$
	(0.669)	(0.273)	(0.053)	(0.010)
MDB	0.691	0.231	0.022	0.001
	[0.563]	[0.089]	[0.006]	[0.0003]
	(0.801)	(0.562)	(0.296)	(0.076)
RB	0.909	0.812	0.712	0.608
	[0.742]	[0.463]	[0.206]	[0.037]
	(0.908)	(0.811)	(0.706)	(0.577)
RB^-	0.907	0.813	0.714	0.606
	[0.888]	[0.782]	[0.672]	[0.565]
	(0.500)	(0.497)	(0.494)	(0.484)
GP	0.504	0.497	0.489	0.467
	[0.502]	[0.495]	[0.481]	[0.460]
	(0.954)	(0.924)	(0.892)	(0.857)
GP-MDB	0.971	0.937	0.893	0.834
	[0.962]	[0.938]	[0.896]	[0.807]
	(0.647/0.001)	(0.256/0.001)	(0.067/0.001)	(0.019/0.1 & 0.2)
LB/γ	0.744/0.01	0.442/0.1	0.338/0.1	0.234/0.2
	[0.772/0.01]	[0.322/0.01]	[0.206/0.1]	[0.064/0.1]

Table 3.7: Lotka-Volterra model. Empirical acceptance probabilities against T, with $(x_T = x_{T,(5)})$, $x_T = x_{T,(50)}$ and $[x_T = x_{T,(95)}]$. The results are based on 100K iterations of a Metropolis-Hastings independence sampler.

exactly. Rather, we observe

$$Y_T = F'X_T + \epsilon_T, \qquad \epsilon_T|\Sigma \sim N(0, \Sigma),$$

where $\Sigma = \sigma^2$ and $F = (1, 0)'$ so that only noisy observation of the population size N_T is possible, and the cumulative population size C_T is not observed at all. We consider a single treatment-block combination and consider the dynamics of the process over an observation time interval $[2.29, 3.57]$, over which nonlinear dynamics are typically observed. We fix θ and $x_{2.29}$ at their marginal posterior means found by Whitaker et al. (2016*a*), that is, at $\theta = (1.45, 0.0009)'$ and $x_{2.29} = (347.55, 398.94)'$. We generate various end-point conditioned scenarios by taking $y_{3.57}$ to be either the 5%, 50% or 95% quantile of $Y_{3.57}|X_{2.29} = (347.55, 398.94)', \sigma$. To investigate the effect of measurement error, we further take $\sigma \in \{5, 10, 50\}$. The resulting quantiles are shown in Table 3.9. For (3.43) we have that

$$H_t = \begin{pmatrix} \theta_1 - \theta_2\eta_{C,t} & -\theta_2\eta_{N,t} \\ \theta_1 & 0 \end{pmatrix},$$

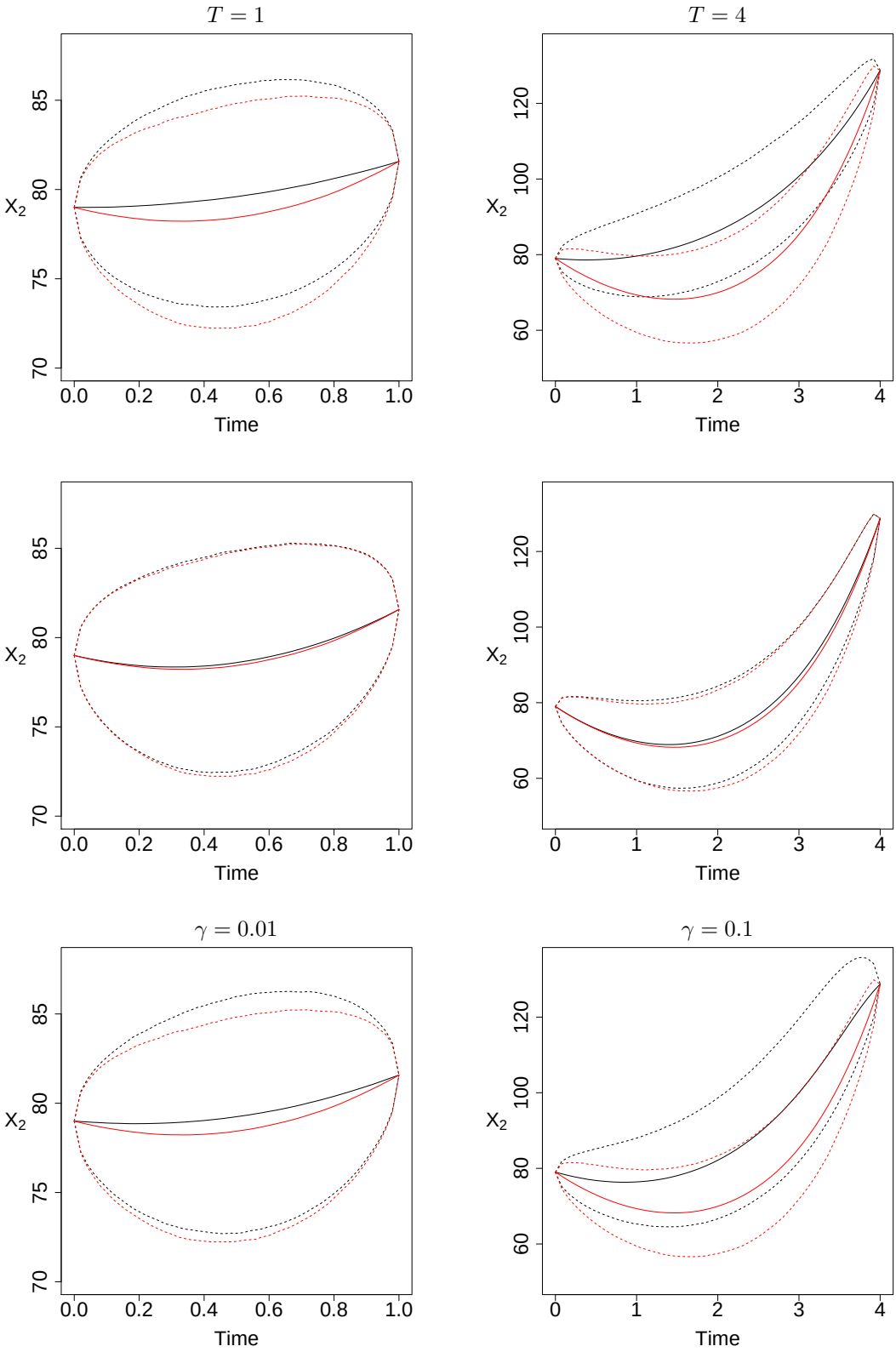

Figure 3.7: Lotka-Volterra model. 95% credible region (dashed line) and mean (solid line) of the true conditioned predator component $X_{2,t}|x_0, x_T$ (red) and various bridge constructs (black) using $x_T = x_{T,(95)}$ for RB (1st row), RB$^-$ (2nd row) and LB (3rd row).

Bridge	$x_{1,(5)}$	$x_{1,(50)}$	$x_{4,(5)}$	$x_{4,(50)}$
MDB	20.414	23.782	0.079	0.043
RB	38.319	56.266	0.579	10.907
RB$^-$	56.505	54.713	10.603	14.427
GP	0.679	0.644	0.614	0.571
GP-MDB	2.148	2.203	1.700	1.599
LB	16.903	37.975	0.534	8.036

Table 3.8: Lotka-Volterra model. Minimum ESS/sec at time $T/2$ for selective end point conditions. The results are based on 100K iterations of a Metropolis-Hastings independence sampler.

	$\sigma = 5$	$\sigma = 10$	$\sigma = 50$
$y_{3.57,(5)}$	726.75	724.57	762.36
$y_{3.57,(50)}$	786.09	815.51	774.41
$y_{3.57,(95)}$	841.82	856.36	910.86

Table 3.9: Aphid growth model. Quantiles of $Y_{3.57}|X_{2.29} = (347.55, 398.94)'$ found by repeatedly simulating from the Euler-Maruyama approximation of (3.43) with $\theta = (1.45, 0.0009)'$, and corrupting $N_{3.57}$ with additive $N(0, \sigma^2)$ noise.

and thus the ODE system governing the linear noise approximation of (3.43) is

$$\frac{d\eta_t}{dt} = \begin{pmatrix} \theta_1\eta_{N,t} - \theta_2\eta_{N,t}\eta_{C,t} \\ \theta_1\eta_{N,t} \end{pmatrix},$$

$$\frac{dP_t}{dt} = \begin{pmatrix} \theta_1 - \theta_2\eta_{C,t} & -\theta_2\eta_{N,t} \\ \theta_1 & 0 \end{pmatrix} P_t,$$

$$\frac{d\psi_t}{dt} = P_t^{-1} \begin{pmatrix} \theta_1\eta_{N,t} + \theta_2\eta_{N,t}\eta_{C,t} & \theta_1\eta_{N,t} \\ \theta_1\eta_{N,t} & \theta_1\eta_{N,t} \end{pmatrix} \left(P_t^{-1}\right)'.$$

As with the previous example, the ODE system is intractable and we again use the `lsoda` package to numerically solve the system when necessary.

Figure 3.8 shows empirical acceptance probabilities against σ for EM, RB, RB$^-$, GP and GP-MDB. The associated values are given in Table 3.10. Figure 3.9 compares 95% credible regions for a selection of bridges with the true conditioned process (obtained from the output of the independence sampler). All results are based on $m = 50$ (but note that no discernible difference in output was obtained for finer discretisations). As illustrated by both figures, the myopic sampler (EM) performs poorly (in terms of statistical efficiency, as measured by empirical acceptance probability) when the measurement error variance is relatively small ($\sigma = 5$). For $\sigma = 50$, the performance of EM is comparable with the other bridge constructs. In fact, as σ increases, the bridge constructs coincide with the

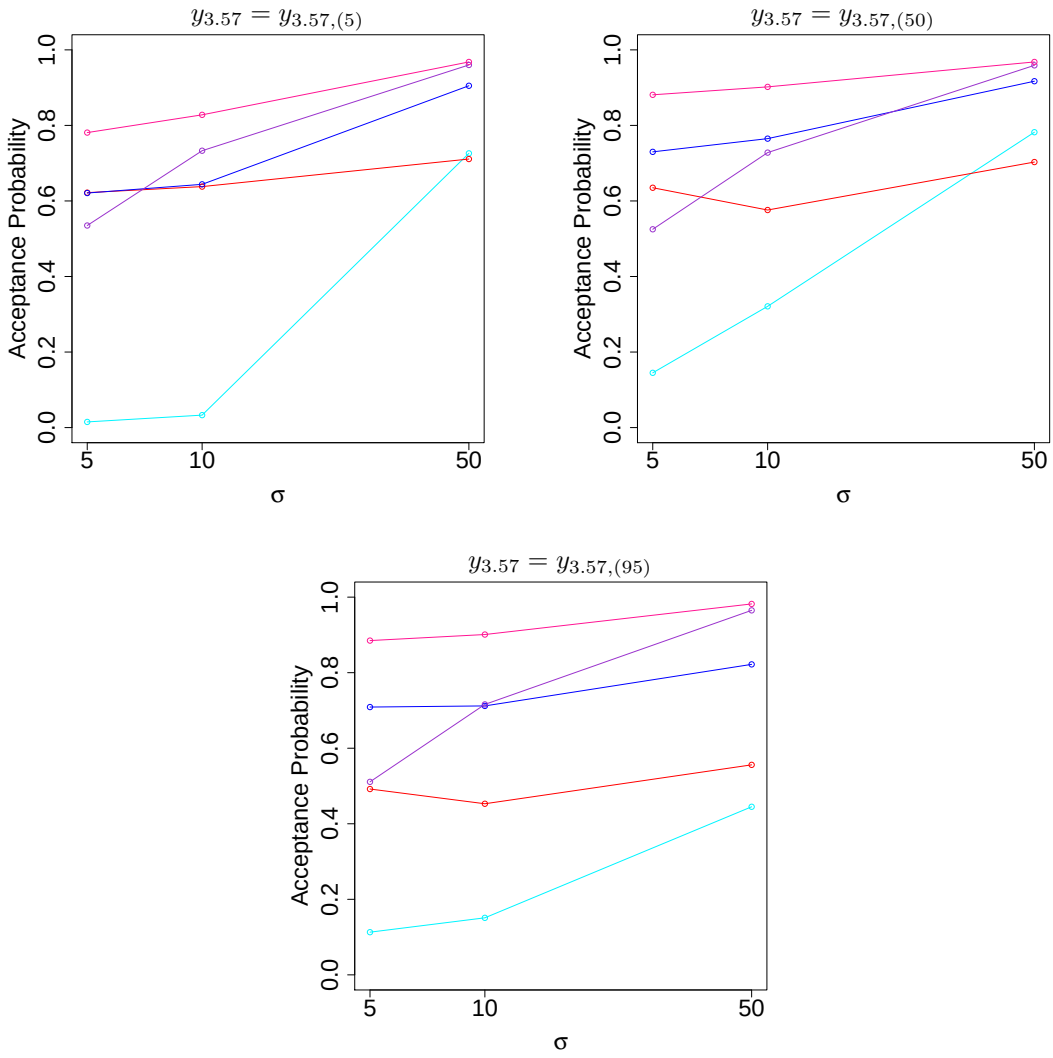

Figure 3.8: Aphid growth model. Empirical acceptance probabilities against σ. The results are based on 100K iterations of a Metropolis-Hastings independence sampler. Turquoise: EM. Red: RB. Blue: RB$^-$. Purple: GP. Pink: GP-MDB.

Euler-Maruyama approximation of the target process. The gain in statistical performance of RB$^-$ over RB is clear. Likewise, GP-MDB outperforms GP, although the difference is very small for $\sigma = 50$ and again we note that as σ increases, the variance under GP-MDB, $\Psi_{\text{MDB}}(x_{\tau_k})$, approaches the Euler-Maruyama variance, as used in GP.

The relative computational cost of each scheme can be found in Table 3.3. EM is particularly cheap to implement, given the simple form of the construct and the Metropolis-Hastings acceptance probability. However, this approach cannot be recommended in this example for $\sigma < 10$ due to its dire statistical efficiency. The computational cost of RB, RB$^-$, GP and GP-M is roughly the same, since for the guided proposals, we found that

Bridge	$\sigma = 5$	$\sigma = 10$	$\sigma = 50$
	(0.015)	(0.033)	(0.726)
EM	0.145	0.321	0.782
	[0.113]	[0.151]	[0.445]
	(0.622)	(0.638)	(0.711)
RB	0.635	0.576	0.703
	[0.492]	[0.453]	[0.556]
	(0.621)	(0.644)	(0.905)
RB^-	0.730	0.765	0.917
	[0.709]	[0.712]	[0.822]
	(0.535)	(0.733)	(0.960)
GP	0.525	0.728	0.959
	[0.511]	[0.716]	[0.965]
	(0.781)	(0.828)	(0.968)
GP-MDB	0.881	0.902	0.968
	[0.885]	[0.901]	[0.982]

Table 3.10: Aphid growth model. Empirical acceptance probabilities against σ, with $(y_{3.57} = y_{3.57,(5)})$, $y_{3.57} = y_{3.57,(50)}$ and $[y_{3.57} = y_{3.57,(95)}]$. The results are based on 100K iterations of a Metropolis-Hastings independence sampler.

a naive implementation that only solves the LNA ODEs once, gave no appreciable difference in empirical acceptance probability as obtained when repeatedly solving the ODE system for each sub-interval $[\tau_k, T]$ (as is required in the case of no measurement error). Consequently, in this example, GP-MDB outperforms RB^- in terms of overall efficiency.

3.6 Summary

Within this chapter we have considered the problem of designing bridge constructs for irreducible, nonlinear, multivariate diffusions. We presented a novel class of bridge constructs which are computationally and statistically efficient (as measured via empirical acceptance probabilities), and can readily be applied in scenarios where only noisy and partial observations are available. Our approach was to partition the process into a deterministic part accounting for forward (nonlinear) dynamics and a residual stochastic process. We then approximated the intractable end-point conditioned SDE through the use of the modified diffusion bridge. Moreover, our approach is straightforward to implement. We considered two variations of the residual SDE:

1. subtraction of a deterministic process based on the drift governing the target process (RB);

2. further subtraction of the expected conditioned residual process via the linear noise approximation (RB^-).

Our examples included a scenario in which the LNA system is tractable, and another where the system must be solved numerically. An example that considers partial and noisy observation of the process at a future time was also presented.

3.6.1 Choice of residual bridge

We find that for all examples considered, the residual bridge that further subtracts the LNA mean results in improved statistical efficiency (over the simple implementation based on the drift subtraction only) at the expense of having to solve a larger ODE system consisting of order d^2 equations (as opposed to just d when using the simpler variant). For a known initial time-point x_0, the ODE system need only be solved once, irrespective of the number of skeleton bridges required. Taking the Lotka-Volterra diffusion (described in Section 3.5.2) as an example, overall efficiency (as measured by minimum effective sample size per second, minESS/sec, at time $T/2$) of RB^- is 1.5 times that of RB when $T = 1$ and x_T is either the 5% or 95% quantile of $X_T|x_0$. This factor increases to 17 when $T = 4$. However, for unknown x_0, as would typically be the case when performing parameter inference, the ODE solution will be required for each skeleton bridge, and the difference in computational cost between the two approaches is likely to be important, especially as the dimension of the state space increases. For the Lotka-Volterra example, the computational cost for solving the ODE system *for each bridge* scales as $1 : 2.8$ for $\text{RB} : \text{RB}^-$. Therefore, the relative difference in minESS/sec would reduce to a factor of roughly 0.5 when $T = 1$ (so that RB would be preferred) and 6 when $T = 4$. We therefore anticipate that in problems where x_0 is unknown, the simple residual bridge is to be preferred, unless the ODE system governing the LNA is tractable, or the dimension d of X_t is relatively small, say $d < 5$.

3.6.2 Residual bridge or guided proposal?

We have compared the performance of our approach to several existing bridge constructs (adapting where necessary to the case of noisy and partial observation). These include the modified diffusion bridge (Durham and Gallant, 2002), Lindström bridge (Lindström, 2012) and guided proposal (Schauer et al., 2016). Our implementation of the latter uses the LNA to guide the proposal. We find that a further modification that replaces the Euler-Maruyama variance with the MDB variance gives a particularly effective bridge, outperforming all others considered here, in terms of statistical efficiency. We find that for fixed x_0 and noisy observation of x_T, an efficient implementation of the guided proposal is

possible, where the ODE system governing the LNA need only be solved once. In this case, the guided proposal outperforms both implementations of the residual bridge in terms of overall efficiency. However, we found that in the case of no measurement error (so that x_T is known exactly), the guided proposal required that the ODEs governing the LNA be re-integrated for each intermediate time-point and for each skeleton bridge required. Unless the ODE system can be solved analytically, we find that when combining statistical and computational efficiency, the guided proposal is outperformed by both implementations of the residual bridge.

Having discussed how to generate a *diffusion bridge*, we now look to implement them within a Bayesian inference scheme.

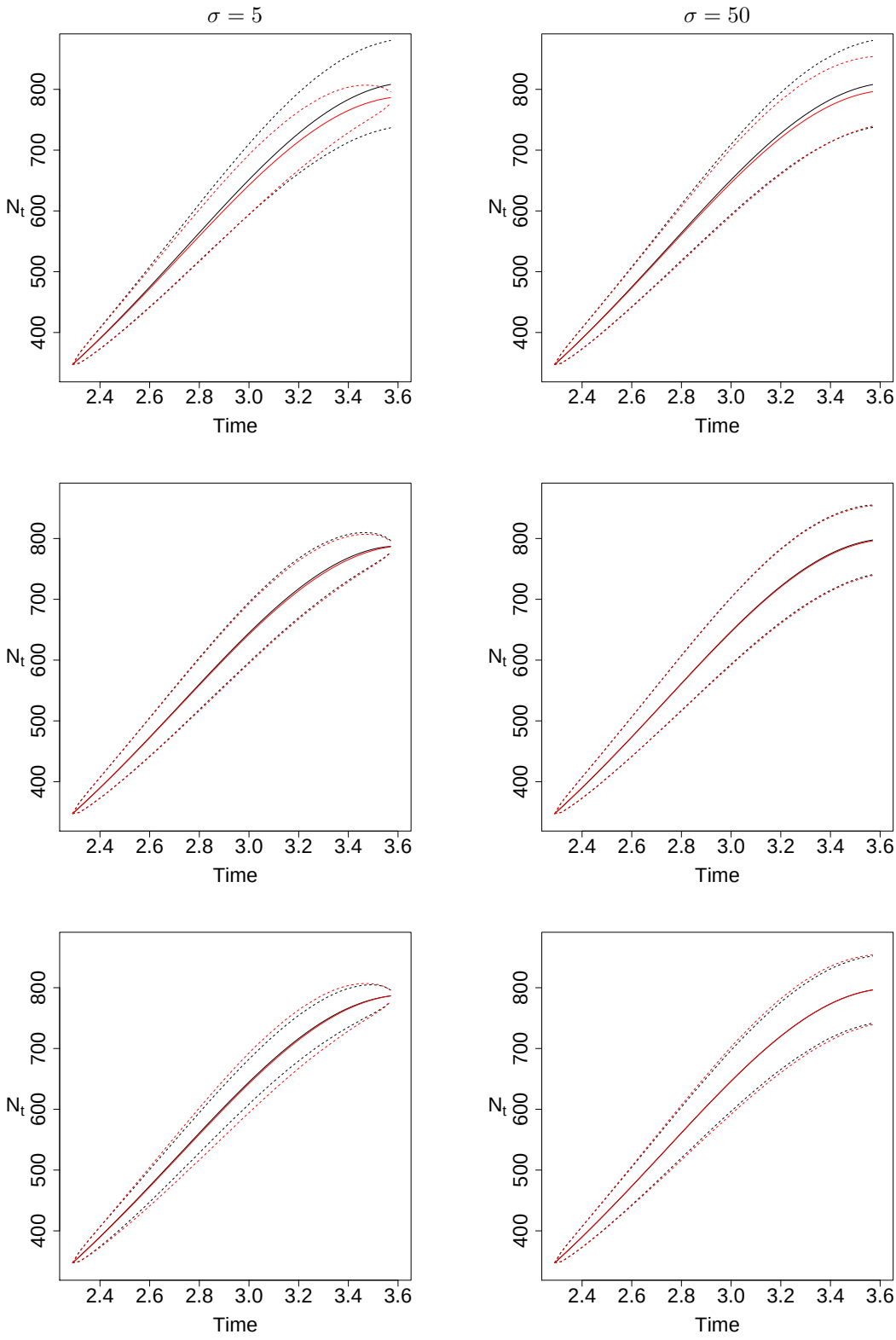

Figure 3.9: Aphid growth model. 95% credible region (dashed line) and mean (solid line) of the true conditioned aphid population component $N_t|x_{2.29}, y_{3.57}$ (red) and various bridge constructs (black) using $y_{3.57} = y_{3.57,(50)}$ for EM (1^{st} row), GP-MDB (2^{nd} row) and RB$^-$ (3^{rd} row).

Chapter 4

Bayesian inference for stochastic differential mixed-effects models

In the previous chapter we discussed various methods to generate a realisation of the sample path of an SDE, known as a diffusion bridge. In this chapter we give details of how to implement these diffusion bridges for a stochastic differential mixed-effects model (SDMEM) within a Bayesian inference scheme. Furthermore we consider the problems surrounding parameter inference; namely that naive schemes result in intolerable mixing as the number of intermediate time points (m) between observation times is increased. The intolerable mixing is due to the dependence between the latent process and the parameters entering the diffusion coefficient. To break said dependence we adapt the modified innovation scheme of Golightly and Wilkinson (2008, 2010) (see also Stramer and Bognar (2011); Fuchs (2013); Papaspiliopoulos et al. (2013)) for SDMEMs. We conclude the chapter by outlining an inference scheme based on the linear noise approximation of Chapter 2.

4.1 Mixed-effects modelling

A mixed-effects model is a model incorporating both fixed effects and random effects. Typically the random effects are assumed to be different draws from a common population profile. It is advantageous to use this type of model in situations when repeated measurements are made on the same experimental units, or when measurements are made on closely related units (usually with the proviso that all units follow the same underlying model). McCulloch and Searle (2004) and Pinheiro and Bates (2009) discuss the importance of mixed-effects models, where each subject is assumed to follow the same model (in our case the same SDE). McCulloch and Searle (2004) discuss linear mixed models before expanding to generalised linear mixed models and in both cases outline the problems one

can encounter when performing inference.

It is only relatively recently that SDEs have been combined with mixed-effects models to give rise to the stochastic differential mixed-effects model. Overgaard et al. (2005) discuss a scenario where the random effects are independent and distributed according to a multivariate normal distribution with zero mean. Parameter estimation is performed via an approximation to the likelihood function. Tornøe et al. (2005) assume the same structure for the random effects and use an extended Kalman filter (EKF) to estimate the (constant) diffusion coefficient. Ditlevsen and De Gaetano (2005) derive the likelihood function for a simple pharmacokinetic example. However, in general, the likelihood function is not available in closed form. Maximum likelihood estimation is examined in Donnet and Samson (2008) using the Stochastic Approximation EM algorithm (SAEM), where an MCMC scheme is used in the simulation step. Donnet et al. (2010) discuss inference for SDMEMs in a Bayesian framework. They implement a Gibbs sampler when the SDE (for each subject) has an explicit solution, as is the case in their chicken growth example. When no explicit solution exists, they propose to approximate the diffusion process using the Euler-Maruyama approximation.

Picchini et al. (2010) propose a procedure to obtain approximate maximum likelihood estimates for SDMEM parameters based on a two step approach. Firstly they use a closed-form Hermite expansion (Aït-Sahalia, 2002, 2008) to approximate the transition density, before using Gaussian quadrature to numerically integrate the conditional likelihood with respect to the random parameters. Picchini and Ditlevsen (2011) note that this approach is, in practice, limited to a scalar random effect parameter since Gaussian quadrature is increasingly computationally inefficient as the dimension of the random effect parameter grows. They extend this methodology to deal with multiple random effects. Using this method has its limitations: for one it may be difficult to gain the transition density using a closed-form Hermite expansion for SDMEMs where the diffusion is irreducible. Another drawback is that the method cannot account for measurement error.

Berglund et al. (2011) compare the use of ODEs and SDEs for describing the kinetics of leucine in blood plasma. Inference is carried out on the SDE parameterisation of the model using the EKF. The `R` package `PSM` (Klim et al., 2009) uses the EKF to estimate SDMEMs. A related approach can be found in Hey et al. (2015) (see also Featherstone et al. (2016)) who build a hierarchical model driven by the linear noise approximation, and apply it for single cell imaging data. Unfortunately, a quantification of the effect of using an approximate inferential model appears to be missing from the literature. For a detailed discussion of hierarchical models, which umbrella mixed-effects models, we refer the reader to Gelman et al. (2013).

In the rest of this chapter we provide a method that permits (simulation-based) Bayesian

inference for a large class of multivariate SDMEMs. As a starting point, we consider a data augmentation approach that adopts an Euler-Maruyama approximation of unavailable transition densities and augments low frequency data with additional time points over which the approximation is satisfactory. Although a discretisation bias is introduced, this can be made arbitrarily small (at greater computational expense). Moreover, the approach is flexible and is not restricted to reducible diffusions. A Bayesian approach then aims to construct the joint posterior distribution for the parameters and the components of the latent process. The intractability of the posterior density necessitates simulation techniques such as Markov chain Monte Carlo. Finally, we detail a competing inference scheme based on the linear noise approximation.

4.2 Stochastic differential mixed-effects models (SDMEMs)

Let us now consider the case where we have N experimental units randomly chosen from some population of units, and associated with each unit i is a continuous-time d-dimensional Itô process $\{X_t^i, t \geq 0\}$ governed by the SDE

$$dX_t^i = \alpha\left(X_t^i, \theta, b^i\right)\, dt + \sqrt{\beta\left(X_t^i, \theta, b^i\right)}\, dW_t^i, \quad X_0^i = x_0^i, \quad i = 1, \ldots, N. \tag{4.1}$$

Here (as previously), α is a d-vector of drift functions, the diffusion coefficient β is a $d \times d$ positive definite matrix with a square root representation $\sqrt{\beta}$ such that $\sqrt{\beta}\sqrt{\beta}' = \beta$ and W_t^i is a d-vector of (uncorrelated) standard Brownian motion processes. The p-vector parameter $\theta = (\theta_1, \ldots, \theta_p)'$ is common to all units whereas the q-vectors $b^i = (b_1^i, \ldots, b_q^i)'$, $i = 1, \ldots, N$, are unit-specific effects, which may be fixed or random. In the most general random effects scenario we let $\pi(b^i|\psi)$ denote the joint distribution of b^i, parameterised by the r-vector $\psi = (\psi_1, \ldots, \psi_r)'$. The model defined by (4.1) allows for differences between experimental units through different realisations of the Brownian motion paths W_t^i and the random effects b^i, accounting for inherent stochasticity within a unit, and variation between experimental units.

It is habitual to assume that each experimental unit $\{X_t^i, t \geq 0\}$ cannot be observed exactly, but observations $y^i = (y_{t_0}^i, y_{t_1}^i, \ldots, y_{t_n}^i)'$ are available, and these are conditionally independent (given the latent process). Note that we observe the process at times $t = t_0, t_1, \ldots, t_n$. We link the observations to the latent process via

$$Y_t^i = F' X_t^i + \epsilon_t, \qquad \epsilon_t|\Sigma \overset{indep}{\sim} N(0, \Sigma), \tag{4.2}$$

where Y_t^i is a d_o-vector, F is a constant $d \times d_o$ matrix and ϵ_t is a random d_o-vector. Note that this setup allows for only observing a subset of components ($d_o < d$), where d_o is

the dimension of the observed components. The aspect of partial observations (subject to error) is explored further in Section 5.2.

Together (4.1) and (4.2) completely specify the stochastic differential mixed-effects model. As alluded to previously, for most problems of interest the form of the SDE associated with each unit will not permit an analytic solution, precluding straightforward inference for the unknown parameters. We therefore work with the Euler-Maruyama approximation

$$\Delta X_t^i \equiv X_{t+\Delta t}^i - X_t^i = \alpha\left(X_t^i, \theta, b^i\right)\,\Delta t + \sqrt{\beta\left(X_t^i, \theta, b^i\right)}\,\Delta W_t^i, \tag{4.3}$$

where $\Delta W_t^i \sim N(0, I_d \Delta t)$ and Δt is the length of time between observations, assumed equally spaced for notational simplicity. It is, of course, unlikely that this approximation will be sufficiently accurate over the intervals between observation times and so we adopt a data augmentation scheme. Partitioning $[t_j, t_{j+1}]$ as

$$t_j = \tau_{j,0} < \tau_{j,1} < \tau_{j,2} < \cdots < \tau_{j,m-1} < \tau_{j,m} = t_{j+1} \tag{4.4}$$

introduces $m-1$ intermediate time points with interval widths of length

$$\Delta\tau \equiv \tau_{j,k+1} - \tau_{j,k} = \frac{t_{j+1} - t_j}{m}, \tag{4.5}$$

cf. equation (3.3). The Euler-Maruyama approximation (4.3) can then be applied over each interval of width $\Delta\tau$, and the associated discretisation bias can be made arbitrarily small at the expense of having to impute $\{X_t^i\}$ at the intermediate times. We adopt the shorthand notation

$$x_{[t_j,t_{j+1}]}^i \equiv x_{[j,j+1]}^i = \left(x_{\tau_{j,0}}^i, x_{\tau_{j,1}}^i, \ldots, x_{\tau_{j,m}}^i\right)'$$

for the latent process between observation times, associated with unit i. Hence, the complete latent trajectory associated with unit i is given by

$$\left(x^i\right)' = \left(\left(x_{[0,1]}^i\right)', \left(x_{(1,2]}^i\right)', \ldots, \left(x_{(n-1,n]}^i\right)'\right)$$

and we stack all unit-specific trajectories into a matrix $x = (x^1, \ldots, x^N)$. Likewise the matrix $y = (y^1, \ldots, y^N)$ denotes the entire set of observations. Next we focus on how to perform Bayesian inference for the model quantities x, θ, $b = (b^1, \ldots, b^N)'$, ψ and Σ.

4.3 Outlining a Bayesian inference scheme

The joint posterior for the common parameters θ, fixed/random effects b, hyperparameters ψ, measurement error variance Σ and latent values x is given by

$$\pi(\theta, \psi, \Sigma, b, x|y) \propto \pi(\theta)\pi(\psi)\pi(\Sigma)\pi(b|\psi)\pi(x|\theta, b)\pi(y|x, \Sigma), \tag{4.6}$$

where $\pi(\theta)\pi(\psi)\pi(\Sigma)$ is the joint prior density ascribed to θ, ψ and Σ. In addition we have that

$$\pi(x|\theta, b) = \prod_{i=1}^{N} \prod_{j=0}^{n-1} \prod_{k=0}^{m-1} \pi\left(x^i_{\tau_{j,k+1}} | x^i_{\tau_{j,k}}, \theta, b^i\right), \tag{4.7}$$

where

$$\pi\left(x^i_{\tau_{j,k+1}} | x^i_{\tau_{j,k}}, \theta, b^i\right) = N\left\{x^i_{\tau_{j,k+1}} \,;\, x^i_{\tau_{j,k}} + \alpha\left(x^i_{\tau_{j,k}}\, \theta, b^i\right)\Delta\tau,\ \beta\left(x^i_{\tau_{j,k}}, \theta, b^i\right)\Delta\tau\right\}$$

and $N(\cdot\,;\, m, V)$ denotes the multivariate Gaussian density with mean m and variance V. Similarly

$$\pi(y|x, \Sigma) = \prod_{i=1}^{N} \prod_{j=0}^{n} \pi\left(y^i_{t_j} | x^i_{t_j}, \Sigma\right),$$

where $\pi(y^i_{t_j}|x^i_{t_j}, \Sigma) = N(y^i_{t_j}\,;\, x^i_{t_j}, \Sigma)$. Given the intractability of the joint posterior distribution in (4.6) we aim to construct a Markov chain Monte Carlo scheme which generates realisations from this posterior (see Chapter 2). The form of the SDMEM admits a Gibbs sampling strategy with blocking that sequentially takes draws from the full conditionals

1. $\pi(x|\theta, \psi, \Sigma, b, y) = \pi(x|\theta, \Sigma, b, y)$,
2. $\pi(\Sigma|\theta, \psi, b, x, y) = \pi(\Sigma|x, y)$,
3. $\pi(\theta|\psi, \Sigma, b, x, y) = \pi(\theta|b, x)$,
4. $\pi(b|\theta, \psi, \Sigma, x, y) = \pi(b|\theta, \psi, x)$,
5. $\pi(\psi|\theta, \Sigma, b, x, y) = \pi(\psi|b)$.

The above scheme can be seen as a data augmentation approach (Tanner and Wong, 1987). Inference may be performed by alternating steps in which the latent trajectories are simulated conditional on the observations and current values of the parameters, and simulation of the parameters given the augmented data. Further blocking strategies that exploit the conditional dependencies between the model parameters and latent trajectories can be used. For example, in step 1 the latent trajectories can be updated separately for each experimental unit. Likewise, the unit-specific random effects can be updated separately.

Where necessary, Metropolis-within-Gibbs updates can be used (see Algorithm 2). We note that as written, this scheme will mix intolerably poorly as the degree of augmentation m is increased due to dependence between the latent values x and the parameters entering the diffusion coefficient (namely θ and b). We refer the reader to Roberts and Stramer (2001) for a detailed discussion of this problem. A simple mechanism for overcoming this issue is to update the parameters and latent trajectories jointly (and this has been considered for SDE models by Stramer and Bognar (2011) and Golightly and Wilkinson (2011)). For SDMEMs a joint update of θ, b and x is likely to result in a sampler with low acceptance rates. We therefore wish to preserve the blocking structure described above and instead adapt the reparameterisation of Golightly and Wilkinson (2008) to our problem. In what follows, we describe in detail each step of the Gibbs sampler.

4.4 Path updates

The full conditional density of the latent paths for all experimental units is given by

$$\pi(x|\theta,\Sigma,b,y) \propto \pi(x|\theta,b)\pi(y|x,\Sigma) = \prod_{i=1}^{N} \pi\left(x^i|\theta,b^i\right)\pi\left(y^i|x^i,\Sigma\right),$$

which suggests a scheme where unit-specific paths are updated separately. In what follows, therefore, we focus on an updating scheme for a single path, and drop i from the notation, writing x in place of x^i and $x_{[j,j+1]}$ in place of $x^i_{[j,j+1]}$. Since the parameters are fixed throughout this updating step, we also drop them from the notation for the duration of this section.

Eraker (2001) suggested that realisations of the latent trajectory are obtained by updating each column of x in turn, conditional on the proceeding and subsequent columns, that is an update which updates x 'one time-point at a time.' Elerian et al. (2001) showed how updating via this method leads to poor mixing within the scheme, caused by high correlation amongst the latent process. They go on to recommend an updating procedure in which the paths are updated in blocks of random size. It is a related version of this strategy which we pursue here.

Following Golightly and Wilkinson (2008) we update x in overlapping blocks of size $2m+1$. Consider times t_j and t_{j+2} at which the current values of the latent process are x_{t_j} and $x_{t_{j+2}}$. The full conditional density of the latent process over the interval (t_j, t_{j+2}) is given by

$$\pi(x_{(j,j+2)}|x_{t_j}, y_{t_{j+1}}, x_{t_{j+2}}) \propto \pi(y_{t_{j+1}}|x_{t_{j+1}}) \prod_{l=j}^{j+1}\prod_{k=0}^{m-1} \pi(x_{\tau_{l,k+1}}|x_{\tau_{l,k}}), \tag{4.8}$$

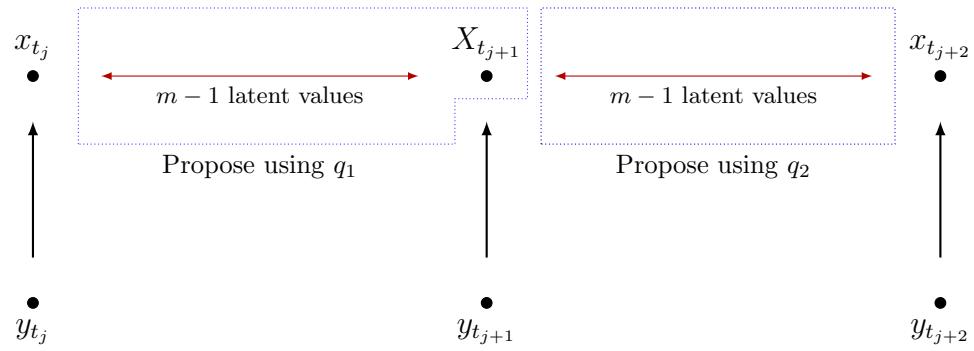

Figure 4.1: Path update illustration over a block of size $2m+1$.

which is analogous in nature to (3.5). Under the nonlinear structure of the diffusion process, this full conditional is intractable and so we use a Metropolis-Hastings step to generate draws from (4.8). We use an independence sampler with a proposal density of the form

$$q\left(x_{(j,j+2)}|x_{t_j}, y_{t_{j+1}}, x_{t_{j+2}}\right) = q_1\left(x_{(j,j+1]}|x_{t_j}, y_{t_{j+1}}\right)\, q_2\left(x_{(j+1,j+2)}|x_{t_{j+1}}, x_{t_{j+2}}\right). \tag{4.9}$$

Figure 4.1 gives an illustration of the updating procedure which can be applied over intervals (t_j, t_{j+2}), $j = 0, 1, \ldots, n-2$. Determining appropriate forms for q_1 and q_2 requires the ability to (approximately) generate a discrete-time realisation of a diffusion process between two time points at which the process is either observed exactly or subject to Gaussian noise (the subject of Chapter 3).

Hence, $q(x_{(j,j+2)}|x_{t_j}, y_{t_{j+1}}, x_{t_{j+2}})$ can be formed by taking any bridge construct discussed in Chapter 3 and simulating a path where

$$q_1(x_{(j,j+1]}|x_{t_j}, y_{t_{j+1}}) = \prod_{k=0}^{m-1} \hat{\pi}(x_{\tau_{j,k+1}}|x_{\tau_{j,k}}, y_{t_{j+1}}) \tag{4.10}$$

and

$$q_2(x_{(j+1,j+2)}|x_{t_{j+1}}, x_{t_{j+2}}) = \prod_{k=0}^{m-2} \hat{\pi}(x_{\tau_{j+1,k+1}}|x_{\tau_{j+1,k}}, x_{t_{j+2}}). \tag{4.11}$$

To elucidate the point, consider the residual bridge (RB) of Section 3.2.1 where we subtract only the drift. We define the proposal mechanism in (4.9) for generating $\{X_t, t \in [t_j, t_{j+2}]\}$ by taking

$$\hat{\pi}(x_{\tau_{j,k+1}}|x_{\tau_{j,k}}, y_{t_{j+1}}) = N\left\{x_{\tau_{j,k+1}}\,;\, x_{\tau_{j,k}} + \mu_{\mathrm{RB}}(x_{\tau_{j,k}})\Delta\tau\,,\, \Psi_{\mathrm{RB}}(x_{\tau_{j,k}})\Delta\tau\right\},$$

where $\mu_{\mathrm{RB}}(x_{\tau_{j,k}})$ and $\Psi_{\mathrm{RB}}(x_{\tau_{j,k}})$ take the form of (3.26) and (3.13) respectively, with $(x_{\tau_k}, \eta_{\tau_k}, r_{\tau_k})$ replaced by $(x_{\tau_{j,k}}, \eta_{\tau_{j,k}}, r_{\tau_{j,k}})$ and (y_T, η_T) replaced by $(y_{t_{j+1}}, \eta_{t_{j+1}})$. Subsequently $\widehat{\pi}(x_{\tau_{j+1,k+1}}|x_{\tau_{j+1,k}}, x_{t_{j+2}})$ can be sampled using

$$\hat{\pi}(x_{\tau_{j+1,k+1}}|x_{\tau_{j+1,k}}, x_{t_{j+2}}) = N\left\{x_{\tau_{j+1,k+1}}\,;\, x_{\tau_{j+1,k}} + \mu^*_{\mathrm{RB}}(x_{\tau_{j+1,k}})\Delta\tau\,,\, \Psi^*_{\mathrm{RB}}(x_{\tau_{j+1,k}})\Delta\tau\right\},$$

with μ^*_{RB} and Ψ^*_{RB} taking the forms expressed in Section 3.2.1, where T is replaced by t_{j+2} and τ_k by $\tau_{j+1,k}$.

A proposed move from $x_{(j,j+2)}$ to $x^*_{(j,j+2)}$ is then accepted with the Metropolis-Hastings acceptance probability $\min(1,A)$, where

$$A = \frac{\pi\Big(y_{t_{j+1}}|x^*_{t_{j+1}}\Big)\prod_{l=j}^{j+1}\prod_{k=0}^{m-1}\pi\Big(x^*_{\tau_{l,k+1}}|x^*_{\tau_{l,k}}\Big)}{\pi\Big(y_{t_{j+1}}|x_{t_{j+1}}\Big)\prod_{l=j}^{j+1}\prod_{k=0}^{m-1}\pi\Big(x_{\tau_{l,k+1}}|x_{\tau_{l,k}}\Big)} \times \frac{q\Big(x_{(j,j+2)}|x_{t_j}, y_{t_{j+1}}, x_{t_{j+2}}\Big)}{q\Big(x^*_{(j,j+2)}|x_{t_j}, y_{t_{j+1}}, x_{t_{j+2}}\Big)}.$$

Note that in the above scheme, x_{t_0} and x_{t_n} remain fixed. We therefore require two additional Metropolis-Hastings steps (such as those described in Golightly and Wilkinson (2006)) that allow for updating X_{t_0} and X_{t_n}. Let us initially consider the problem of updating X_{t_n}.

Take the interval $[t_{n-1}, t_n]$, and partition as in (4.4). We make the assumption that $x_{t_{n-1}}$ is known and fixed for this update, that is, $X_{t_{n-1}} = x_{t_{n-1}}$. Our aim is then to update $x_{(n-1,n]}$ conditional on $x_{t_{n-1}}$ and y_{t_n}. Thus, the proposal density takes the form of q_1 (4.10). The updating procedure is illustrated in Figure 4.2 (and is simply the single interval update extensively discussed in Chapter 3). A move from $x_{(n-1,n]}$ to $x^*_{(n-1,n]}$ has the acceptance probability $\min(1,A)$, where

$$A = \frac{\pi\Big(y_{t_n}|x^*_{t_n}\Big)\prod_{k=0}^{m-1}\pi\Big(x^*_{\tau_{n-1,k+1}}|x^*_{\tau_{n-1,k}}\Big)}{\pi\Big(y_{t_n}|x_{t_n}\Big)\prod_{k=0}^{m-1}\pi\Big(x_{\tau_{n-1,k+1}}|x_{\tau_{n-1,k}}\Big)} \times \frac{\prod_{k=0}^{m-1}\hat{\pi}\Big(x_{\tau_{n-1,k+1}}|x_{\tau_{n-1,k}}, y_{t_n}\Big)}{\prod_{k=0}^{m-1}\hat{\pi}\Big(x^*_{\tau_{n-1,k+1}}|x^*_{\tau_{n-1,k}}, y_{t_n}\Big)}.$$

To update the start of the path, that is, to update X_{t_0}, let us consider the interval $[t_0, t_1]$, again partitioned as in (4.4). In contrast to the above where we looked to update X_{t_n}, we now assume that the right-hand end of the interval is known and fixed, that is, $X_{t_1} = x_{t_1}$. One possible proposal strategy is to draw a candidate $x^*_{t_0}$ from the prior distribution, and propose the path $x^*_{(0,1)}$ using q_2 (4.11) conditional on $x^*_{t_0}$. However, unless the prior distribution is particularly informative, this approach is likely to be highly inefficient. Instead we propose to update X_{t_0} using a Metropolis-Hastings random walk sampler.

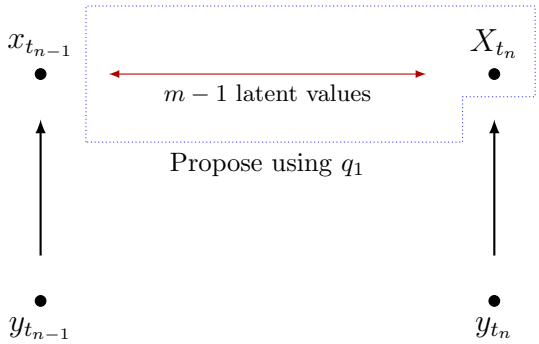

Figure 4.2: Path update illustration to update X_{t_n}.

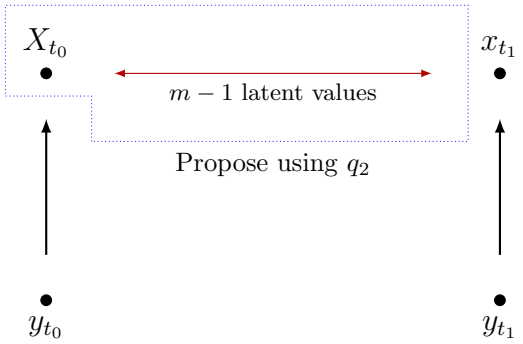

Figure 4.3: Path update illustration to update X_{t_0}.

Hence, we have that

$$x^*_{t_0} = x_{t_0} + \omega, \qquad \omega \sim N(0, \Sigma_\omega).$$

We then propose $x^*_{(0,1)}$ via q_2 (4.11) conditional on $x^*_{t_0}$. See Figure 4.3 for an illustration. A move from $x_{[0,1)}$ to $x^*_{[0,1)}$ is then accepted with probability $\min(1,A)$, where

$$A = \frac{\pi\left(x^*_{t_0}\right)\pi\left(y_{t_0}|x^*_{t_0}\right)\prod_{k=0}^{m-1}\pi\left(x^*_{\tau_{0,k+1}}|x^*_{\tau_{0,k}}\right)}{\pi\left(x_{t_0}\right)\pi\left(y_{t_0}|x_{t_0}\right)\prod_{k=0}^{m-1}\pi\left(x_{\tau_{0,k+1}}|x_{\tau_{0,k}}\right)} \times \frac{\prod_{k=0}^{m-2}\hat{\pi}\left(x_{\tau_{0,k+1}}|x_{\tau_{0,k}}, x_{t_1}\right)}{\prod_{k=0}^{m-2}\hat{\pi}\left(x^*_{\tau_{0,k+1}}|x^*_{\tau_{0,k}}, x_{t_1}\right)}.$$

For the special case of no measurement error and full observation of all components the above scheme simplifies dramatically. We no longer require the use of overlapping blocks of size $2m+1$, and instead we update x using standard blocks of size $m+1$. As x_{t_j}, $j = 0, \ldots, n$ is known and fixed, the left and right endpoints of any interval are known and fixed. Therefore we may update any given block, $x_{(j,j+1)}$, through the use of q_2 (4.11); see Figure 4.4. Note that a separate update for x_{t_0} and x_{t_n} is no longer necessary. In this special case, we sample x using the diffusion bridges characterised by the $\mu^*(x_{\tau_{j,k}})$ and $\Psi^*(x_{\tau_{j,k}})$ of Chapter 3. A proposed move from $x_{(j,j+1)}$ to $x^*_{(j,j+1)}$ has the associated

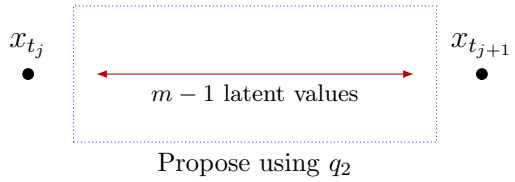

Figure 4.4: Path update illustration over a block of size $m+1$.

acceptance probability $\min(1, A)$, where

$$A = \frac{\prod_{k=0}^{m-1} \pi\left(x^*_{\tau_{j,k+1}} | x^*_{\tau_{j,k}}\right)}{\prod_{k=0}^{m-1} \pi\left(x_{\tau_{j,k+1}} | x_{\tau_{j,k}}\right)} \times \frac{\prod_{k=0}^{m-2} \hat{\pi}\left(x_{\tau_{j,k+1}} | x_{\tau_{j,k}}, x_{t_{j+1}}\right)}{\prod_{k=0}^{m-2} \hat{\pi}\left(x^*_{\tau_{j,k+1}} | x^*_{\tau_{j,k}}, x_{t_{j+1}}\right)}.$$

4.5 Parameter updates

In the last section we outlined a strategy to update the latent process x, fulfilling step 1 in the Gibbs sampling strategy of Section 4.3. We now turn our attention to updating the parameters within the model, that is, steps 2–5 of the presented Gibbs sampling strategy.

The full conditional densities of Σ and ψ are

$$\pi(\Sigma | x, y) \propto \pi(\Sigma)\pi(y|\Sigma) \qquad \text{and} \qquad \pi(\psi | b) \propto \pi(\psi)\pi(b|\psi).$$

Often, semi-conjugate priors can be specified for Σ and ψ negating the need for Metropolis-within-Gibbs steps, and updates proceed from their full conditional distribution.

Unfortunately, for the remaining parameters of interest, θ and $b = (b^1, \ldots, b^N)'$, direct sampling is generally impossible. Therefore we must again appeal to Metropolis-within-Gibbs. For θ and b we have

$$\pi(\theta | b, x) \propto \pi(\theta)\pi(x|\theta, b) \tag{4.12}$$

and

$$\pi(b | \theta, \psi, x) \propto \pi(b|\psi)\pi(x|\theta, b) = \prod_{i=1}^{N} \pi\left(b^i | \psi\right) \pi\left(x^i | \theta, b^i\right), \tag{4.13}$$

where the last expression suggests unit-specific updates of the components of b. The method to update θ or a unit-specific b^i is identical, and thus, we only detail the update for θ here.

We update θ via a random walk proposal with Gaussian innovations (see Chapter 2). Here, a new value θ^* is obtained using

$$\theta^* = \theta + \omega, \qquad \omega \sim N(0, \Sigma_\omega),$$

where Σ_ω is the covariance matrix of tuning parameters. A proposed θ^* is then accepted with the Metropolis-Hastings acceptance probability $\min(1, A)$, where

$$\begin{aligned} A &= \frac{\pi(\theta^*)\pi(x|\theta^*, b)}{\pi(\theta)\pi(x|\theta, b)} \times \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \\ &= \frac{\pi(\theta^*)\pi(x|\theta^*, b)}{\pi(\theta)\pi(x|\theta, b)}. \end{aligned}$$

The above simplification to the acceptance probability is induced by the symmetric nature of the proposal distribution, that is

$$q(\theta^*|\theta) = q(\theta|\theta^*) \propto \exp\left\{-\frac{1}{2}(\theta^* - \theta)'\Sigma_\omega^{-1}(\theta^* - \theta)\right\}.$$

However, for many SDMEMs of interest, we may require that all the components of θ be non-negative, such that $\theta \in \mathbb{R}_+^p$. In this case, we adopt the transformation

$$\xi = \log\theta = (\log\theta_1, \ldots, \log\theta_p)'.$$

For ease of exposition, let us suppose that our prior beliefs about θ can be represented by independent distributions for each component θ_i, $i = 1, \ldots, p$. Whence, the prior distribution for ξ is

$$\pi(\xi) = \prod_{i=1}^{p} \pi_{\theta_i}\left(e^{\xi_i}\right) e^{\xi_i},$$

so that the full conditional density for ξ is given by

$$\pi(\xi|b, x) \propto \pi\left(\xi\right) \pi\left(x|e^{\xi}, b\right). \tag{4.14}$$

We sample (4.14) using a Metropolis-Hastings step with proposal

$$\xi^* = \xi + \omega, \qquad \omega \sim N(0, \Sigma_\omega).$$

A move to ξ^* is then accepted with probability $\min(1,A)$, where

$$\begin{aligned} A &= \frac{\pi\left(\xi^*\right)\pi\left(x|e^{\xi^*},b\right)}{\pi\left(\xi\right)\pi\left(x|e^{\xi},b\right)} \\ &= \frac{\prod_{i=1}^{p}\pi_{\theta_i}\left(e^{\xi_i^*}\right)\theta_i^*}{\prod_{i=1}^{p}\pi_{\theta_i}\left(e^{\xi_i}\right)\theta_i} \times \frac{\pi\left(x|e^{\xi^*},b\right)}{\pi\left(x|e^{\xi},b\right)}. \end{aligned}$$

The method outlined above updates all components of θ (or ξ) in a single block update. As mentioned in Chapter 2, this approach is likely to be inefficient. In practice, single components of θ are updated in turn; or block updates are applied to certain components of θ, with the blocks chosen to break the correlations within θ. Each block is then updated subject to the strategy outlined above.

4.5.1 Convergence problems

As mentioned previously, since θ and the components of b enter into the diffusion coefficient of (4.1), sampling the full conditionals of $\theta|b,x$ and $b|\theta,\psi,x$ as part of a Gibbs sampler (as outlined in the MCMC scheme above) will result in a reducible Markov chain as $m\to\infty$ (or $\Delta\tau\to 0$). Typically we might expect to see intolerably poor mixing of the scheme for $m>5$ (Eraker, 2001). This issue was first discussed in Roberts and Stramer (2001), who highlight the dependence between the quadratic variation of the latent process and the diffusion coefficient. For a specific interval $[t_j,t_{j+1}]$ partitioned as in (4.4) the quadratic variation (for a given experimental unit i) is given by

$$\left\langle x^i,x^i\right\rangle_{[j,j+1]} = \lim_{m\to\infty}\sum_{k=0}^{m-1}\left(x^i_{\tau_{j,k+1}}-x^i_{\tau_{j,k}}\right)\left(x^i_{\tau_{j,k+1}}-x^i_{\tau_{j,k}}\right)' = \int_{t_j}^{t_{j+1}}\beta\left(x^i_\tau,\theta,b^i\right)d\tau. \tag{4.15}$$

The form of (4.15) shows that there is an inherent link between the quadratic variation and the diffusion coefficient: as soon as $x^i_{[j,j+1]}$ is known, θ and b^i can be deduced directly from it using (4.15), and contrariwise, a fixed $\beta(X^i_t,\theta,b^i)$ determines x^i. Therefore a scheme which alternates between updates of x, θ and b will be reducible, as x produces θ and b, and θ and b determine the quadratic variation of x. Thus, the scheme will not converge for $m\to\infty$. In practice however, m is finite and we see intolerable mixing which worsens with m.

Shephard and Pitt (1997) found that the use of random blocks to update x (such as those discussed above in Section 4.4) can aid convergence. However, the dependence between the latent process and parameters still exists, causing poor mixing as $m\to\infty$ (or as the

number of observations increases).

Roberts and Stramer (2001) overcome the dependence (in a univariate diffusion setting) by performing a transformation which leads to an irreducible scheme (even as $m \to \infty$). The transformation is a form of the Lamperti transformation, which is discussed for multivariate diffusions by Picchini and Ditlevsen (2011). If the diffusion in question is reducible then the Lamperti transformation (for a specific experimental unit, where we again suppress i to ease notation) is given by

$$\upsilon\left(x, \theta, b\right) = \left(\upsilon_1\left(x, \theta, b\right), \ldots, \upsilon_d\left(x, \theta, b\right)\right)',$$

such that

$$\nabla\upsilon\left(x, \theta, b\right) = \beta\left(x, \theta, b\right)^{-\frac{1}{2}},$$

where

$$\nabla\upsilon\left(x, \theta, b\right) = \begin{pmatrix} \dfrac{\partial\upsilon_1}{\partial x_1} & \cdots & \dfrac{\partial\upsilon_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial\upsilon_d}{\partial x_1} & \cdots & \dfrac{\partial\upsilon_d}{\partial x_d} \end{pmatrix}.$$

Through the use of the multivariate Itô formula (2.22) with $Y_t = \upsilon\left(X_t, \theta, b\right)$, we obtain a new SDE as

$$dY_t = \alpha^*\left(Y_t, \theta, b\right) dt + dW_t,$$

which clearly has unit diffusion coefficient.

Example: Birth-death model

Recall the SDE for a simple birth-death process given in Section 3.5.1

$$dX_t = \left(\theta_1 - \theta_2\right) X_t\, dt + \sqrt{\left(\theta_1 + \theta_2\right) X_t}\, dW_t, \quad X_0 = x_0. \tag{4.16}$$

The Lamperti transform can be applied to (4.16) by taking

$$\nabla\upsilon = \frac{1}{\sqrt{\left(\theta_1 + \theta_2\right) x}},$$

which gives that

$$\upsilon = \frac{2\sqrt{x}}{\sqrt{\theta_1 + \theta_2}}.$$

Using Itô's formula (2.15) with

$$Y_t = f(X_t, \theta) = \frac{2\sqrt{X_t}}{\sqrt{\theta_1 + \theta_2}},$$

where

$$f_t = 0, \quad f_x = \frac{1}{\sqrt{(\theta_1 + \theta_2)\, x}} \quad \text{and} \quad f_{xx} = \frac{-1}{2\sqrt{\theta_1 + \theta_2}\, x^{3/2}},$$

we obtain

$$dY_t = \left\{ \frac{(\theta_1 - \theta_2)\sqrt{X_t}}{\sqrt{\theta_1 + \theta_2}} - \frac{\sqrt{\theta_1 + \theta_2}}{4\sqrt{X_t}} \right\} dt + dW_t.$$

We note that

$$\sqrt{X_t} = \frac{\sqrt{\theta_1 + \theta_2}\, Y_t}{2},$$

and hence

$$dY_t = \left\{ \frac{(\theta_1 - \theta_2)\, Y_t^2 - 1}{2Y_t} \right\} dt + dW_t.$$

In practice however, the above transformation is almost always impossible to apply for nonlinear multivariate diffusions (Papaspiliopoulos et al., 2003). We also note that time change transformations and the implementation of particle filters can be used to make improvements to the convergence of the scheme (see Fuchs (2013) and the references therein). However, within this thesis, to overcome this problem of dependence we use a reparameterisation which is outlined in the next section.

4.5.2 Modified innovation scheme

The following reparameterisation (and subsequent updating scheme) should be applicable for any multivariate diffusion, subject to fairly general regularity conditions. The associated MCMC scheme is computationally efficient and importantly, does not lead to a reducible scheme as $m \to \infty$. The main assumption is that $\beta(X_t^i, \theta, b^i)$ is invertible. Note that in the case when $\beta(X_t^i, \theta, b^i)$ is rank-degenerate, the inverse may be found using the Moore-Penrose generalised inverse, so that the scheme outlined in this section is still applicable.

The innovation scheme was first outlined in Chib et al. (2004) and exploits the fact that, given θ and b, under the Euler-Maruyama approximation (4.3) there is a one-to-one relationship between ΔX_t and ΔW_t: the increments of the process and the increments of the driving Brownian motion respectively. Here

$$\Delta X_t^i = \alpha\left(X_t^i, \theta, b^i\right) \Delta t + \sqrt{\beta\left(X_t^i, \theta, b^i\right)}\, \Delta W_t^i$$

and hence

$$\Delta W_t^i = \beta\left(X_t^i, \theta, b^i\right)^{-1/2}\left\{\Delta X_t^i - \alpha\left(X_t^i, \theta, b^i\right)\Delta t\right\}.$$

Moreover, whilst the quadratic variation of X determines θ and b (as $m \to \infty$), the quadratic variation of the Brownian process is independent of θ and b *a priori*. Therefore conditioning on the Brownian increment innovations in a Gibbs update should be effective in overcoming the dependence problem. The resulting algorithm is known as the innovation scheme. Unfortunately, combining an updated parameter value with the Brownian increments will not necessarily give an imputed path that is consistent with the observations. Therefore, Golightly and Wilkinson (2008, 2010) suggest that a *diffusion bridge* (such as the modified diffusion bridge of Durham and Gallant (2002), see Section 3.1.2) be used to determine the innovation process, leading to a modified innovation scheme.

Fuchs (2013) considers the modified innovation scheme in a continuous-time framework. Adapting the innovation process used by Fuchs to an SDMEM, we have for an interval $[t_j, t_{j+1}]$, an innovation process $\{Z_t^i, t \in [t_j, t_{j+1}]\}$ satisfying

$$\begin{aligned} dZ_t^i &= \beta\left(X_t^i, \theta, b^i\right)^{-1/2}\left(dX_t^i - \frac{x_{t_{j+1}}^i - X_t^i}{t_{j+1}-t}\,dt\right) \\ &= \beta\left(X_t^i, \theta, b^i\right)^{-1/2}\left\{\alpha\left(X_t^i, \theta, b^i\right) - \frac{x_{t_{j+1}}^i - X_t^i}{t_{j+1}-t}\right\}dt + dW_t^i, \end{aligned} \tag{4.17}$$

with $Z_{t_j}^i = 0$. Clearly, each process Z^i has unit diffusion coefficient and whilst not Brownian motion processes, the probability measures induced by each Z^i are absolutely continuous with respect to Wiener measure. A proof of this result can be found in Fuchs (2013) as well as a justification for using this form of innovation process as the effective component in a Gibbs sampler.

The aim is to apply a discretisation of (4.17) between observation times. We therefore define $x_o^i = (x_{t_0}^i, \ldots, x_{t_n}^i)'$ to be the current values of the (unit-specific) latent process at the observation times, and stack all x_o^i values into the matrix x_o. Following Golightly and Wilkinson (2008), we have for $k = 0, \ldots, m-1$

$$Z_{\tau_{j,k+1}}^i - Z_{\tau_{j,k}}^i = \beta^*\left(X_{\tau_{j,k}}^i, \theta, b^i\right)^{-1/2}\left(X_{\tau_{j,k+1}}^i - X_{\tau_{j,k}}^i - \frac{x_{t_{j+1}}^i - X_{\tau_{j,k}}^i}{t_{j+1}-\tau_{j,k}}\Delta\tau\right),$$

where $Z_{\tau_{j,0}} = 0$ and

$$\beta^*\left(X_{\tau_{j,k}}^i, \theta, b^i\right) = \frac{t_{j+1}-\tau_{j,k+1}}{t_{j+1}-\tau_{j,k}}\beta\left(X_{\tau_{j,k}}^i, \theta, b^i\right).$$

Now define a function f so that $X^i_{\tau_{j,k}} = f(Z^i_{\tau_{j,k}}, \theta, b^i)$ and $Z^i_{\tau_{j,k}} = f^{-1}(X^i_{\tau_{j,k}}, \theta, b^i)$. Let z^i_{imp} denote the (unit-specific) innovation values over $[t_0, t_n]$ and stack all z^i_{imp} values into the matrix z_{imp}. Define x^i_{imp} and x_{imp} similarly. The modified innovation scheme samples $\theta|b, z_{imp}, x_o$ and $b^i|\theta, \psi, z^i_{imp}, x^i_o$, $i = 1, \ldots, N$. Note that for an updated value of b^i, say b^{i*}, a new x^{i*}_{imp} is updated deterministically through $x^{i*}_{imp} = f(z^{i*}_{imp}, \theta, b^{i*})$. Likewise, for a new θ^*, a new x^*_{imp} is updated deterministically through $x^{i*}_{imp} = f(z^{i*}_{imp}, \theta^*, b^i)$, $i = 1, \ldots, N$. The full conditional density of θ is

$$\pi(\theta|b, z_{imp}, x_o) \propto \pi(\theta) \prod_{i=1}^{N} \prod_{j=1}^{n-1} \left[\prod_{k=0}^{m-1} \pi\left(x^i_{\tau_{j,k+1}} | x^i_{\tau_{j,k}}, \theta, b^i\right) \prod_{k=0}^{m-2} J\left\{ f\left(z^i_{\tau_{j,k+1}}, \theta, b^i\right)\right\}\right], \tag{4.18}$$

where

$$J\left\{ f\left(z^i_{\tau_{j,k+1}}, \theta, b^i\right)\right\} = \left|\beta^*\left(x^i_{\tau_{j,k}}, \theta, b^i\right)\right|^{-1/2}$$

is the Jacobian determinant of f. Similarly, the full conditional density of b^i, $i = 1, \ldots, N$ is

$$\pi\left(b^i|\theta, \psi, z^i_{imp}, x^i_o\right) \propto \pi\left(b^i|\psi\right) \prod_{j=1}^{n-1} \left[\prod_{k=0}^{m-1} \pi\left(x^i_{\tau_{j,k+1}} | x^i_{\tau_{j,k}}, \theta, b^i\right) \prod_{k=0}^{m-2} J\left\{ f\left(z^i_{\tau_{j,k+1}}, \theta, b^i\right)\right\}\right]. \tag{4.19}$$

Naturally, the full conditionals in (4.18) and (4.19) will typically be intractable, requiring the use of Metropolis-within-Gibbs updates. Therefore a proposed move from θ to θ^* is accepted with probability min(1,A), where

$$A = \frac{\pi(\theta^*)\pi\{f(z^*_{imp}, \theta^*, b)|\theta^*, b\}J\{f(z^*_{imp}, \theta^*, b)\}}{\pi(\theta)\pi\{f(z_{imp}, \theta, b)|\theta, b\}J\{f(z_{imp}, \theta, b)\}}.$$

Similarly, a proposed value b^{i*} is accepted with probability min(1,A), where

$$A = \frac{\pi\left(b^{i*}|\psi\right)\pi\left\{f\left(z^{i*}_{imp}, \theta, b^{i*}\right)|\theta, b^{i*}\right\} J\left\{f\left(z^{i*}_{imp}, \theta, b^{i*}\right)\right\}}{\pi\left(b^i|\psi\right)\pi\left\{f\left(z^i_{imp}, \theta, b^i\right)|\theta, b^i\right\} J\left\{f\left(z^i_{imp}, \theta, b^i\right)\right\}}.$$

We now explicitly outline the MCMC scheme invoked under the modified innovation scheme. Recall from Section 4.3, the Gibbs sampling steps

1. $\pi(x|\theta, \psi, \Sigma, b, y) = \pi(x|\theta, \Sigma, b, y)$,
2. $\pi(\Sigma|\theta, \psi, b, x, y) = \pi(\Sigma|x, y)$,
3. $\pi(\theta|\psi, \Sigma, b, x, y) = \pi(\theta|b, x)$,

4. $\pi(b|\theta, \psi, \Sigma, x, y) = \pi(b|\theta, \psi, x)$,

5. $\pi(\psi|\theta, \Sigma, b, x, y) = \pi(\psi|b)$.

We have discussed above how the implementation of this scheme leads to a reducible Markov chain. Therefore, we utilise the following Gibbs sampling steps

1. $\pi(z|\theta, \psi, \Sigma, b, y) = \pi(z|\theta, \Sigma, b, y)$,

2. $\pi(\Sigma|\theta, \psi, b, x, y) = \pi(\Sigma|x, y)$,

3. $\pi(\theta|\psi, \Sigma, b, z_{imp}, y) = \pi(\theta|b, z_{imp}, x_o)$,

4. $\pi(b|\theta, \psi, \Sigma, z_{imp}, y) = \pi(b|\theta, \psi, z_{imp}, x_o)$,

5. $\pi(\psi|\theta, \Sigma, b, z_{imp}, y) = \pi(\psi|b)$.

Note that step 2 of the algorithm remains unchanged as Σ does not feature in the diffusion coefficient, and it is only those parameters which feature in the diffusion coefficient that cause the dependence problem to occur.

The updates for Σ and ψ are identical to those presented at the beginning of Section 4.5. We update θ and b using a Metropolis-Hastings algorithm for the full conditional distributions given above, noting that the use of block updates may again be profitable. This leaves only the update for z. Given the relationship between z and x, it is sufficient to update x using the techniques outlined in Section 4.4, before constructing z from x deterministically. A scheme taking this form will be successful in overcoming the dependence between the latent process and the parameters entering the diffusion coefficient, and hence, be irreducible as $m \to \infty$. A rigorous justification of this scheme is beyond the scope of this thesis, however we refer the reader to (Fuchs, 2013, Chapter 7) for further details.

4.6 The linear noise approximation to SDMEMs

We now return to a concept initially introduced in Chapter 2, namely a tractable approximation to an SDE. Here we extend the LNA of Section 2.7 to SDMEMs before outlining an inference scheme based upon this approximation. Recall that for an SDE of the form

$$dX_t = \alpha(X_t, \theta)\, dt + \sqrt{\beta(X_t, \theta)}\, dW_t,$$

we partition X_t as $X_t = \eta_t + R_t$. Replacing R_t with the approximate residual process $\hat{R}_t$, given by

$$d\hat{R}_t = H_t \hat{R}_t\, dt + \sqrt{\beta(\eta_t, \theta)}\, dW_t, \tag{4.20}$$

and noting that $\hat{r}_0 = 0$, we obtain $X_t|X_0 = x_0 \sim N(\eta_t, V_t)$, where

$$\begin{aligned}\frac{d\eta_t}{dt} &= \alpha(\eta_t, \theta), \qquad \eta_0 = x_0,\\ \frac{dV_t}{dt} &= V_t H_t' + \beta(\eta_t, \theta) + H_t V_t, \qquad V_0 = 0,\end{aligned}$$

and H_t is the Jacobian matrix with (i,j)th element

$$(H_t)_{i,j} = \frac{\partial \alpha_i(\eta_t, \theta)}{\partial \eta_{j,t}}.$$

The extension of the above to SDMEMs is straighforward. Noting that each $\{X_t^i, t \geq 0\}$, $i = 1, \ldots, N$, follows the same underlying SDE, we partition X_t^i as $X_t^i = \eta_t^i + R_t^i$, where η_t^i and R_t^i have the same definitions as η_t and R_t given previously. Thus, a solution for each i, is gained through the ODE system

$$\frac{d\eta_t^i}{dt} = \alpha\left(\eta_t^i, \theta, b^i\right), \qquad \eta_0^i = x_0^i, \tag{4.21}$$

$$\frac{dV_t^i}{dt} = V_t^i \left(H_t^i\right)' + \beta\left(\eta_t^i, \theta, b^i\right) + H_t^i V_t^i, \qquad V_0^i = 0, \tag{4.22}$$

where

$$(H_t^i)_{j,k} = \frac{\partial \alpha_j\left(\eta_t^i, \theta, b^i\right)}{\partial \eta_{k,t}^i}.$$

If (4.21) and (4.22) is a tractable system (such that analytic expressions for η_t^i and V_t^i can be obtained), we note that the forms of each η_t^i and V_t^i will be the same, however they will be determined by the differing η_t^i and b^i in each instance.

Fearnhead et al. (2014) describe a filtering algorithm for computing the marginal likelihood $\pi(y|\theta, b, \Sigma)$ for the Gaussian observation regime (4.2). Here we exploit the computational efficiency of their approach. The performance of the linear noise approximation of the SDMEM (4.1) is examined in Chapter 5.

It is worth noting here that the linear form of the SDE (4.20) satisfied by the approximate residual process $(\hat{R}_t)$ coupled with the additive Gaussian observation regime admits a closed form expression for densities of the form $\widehat{\pi}(\hat{r}_{\tau_{j,k+1}}|\hat{r}_{\tau_{j,k}}, y_{t_{j+1}})$, suggesting use of the LNA as a proposal mechanism inside the Bayesian imputation approach of Section 4.4. Whilst the LNA could in principle be used to directly approximate the conditioned residual process governed by the SDE in (3.23), we note that the SDEs in (3.23) and (4.20) have different diffusion coefficients. Consequently, the probability law governing $\hat{R}_t$ is not absolutely continuous with respect to the law of R_t. We therefore do not advocate use of the LNA in this way.

4.6.1 Application to SDMEMs

Under the linear noise approximation of (4.1) the marginal posterior for all parameters is given by

$$\begin{aligned}\pi(\theta, \psi, \Sigma, b|y) &\propto \pi(\theta)\pi(\psi)\pi(\Sigma)\pi(b|\psi)\pi(y|\theta, \Sigma, b) \\ &\propto \pi(\theta)\pi(\psi)\pi(\Sigma)\prod_{i=1}^{N}\pi\left(b^i|\psi\right)\pi\left(y^i|\theta, \Sigma, b^i\right). \end{aligned} \tag{4.23}$$

This factorisation suggests a Gibbs sampler with blocking that sequentially takes draws from the full conditionals

1. $\pi(\Sigma|\theta, \psi, b, y) = \pi(\Sigma|y)$,
2. $\pi(\theta|\psi, \Sigma, b, y) = \pi(\theta|b, y)$,
3. $\pi(b|\theta, \psi, \Sigma, y) = \pi(b|\theta, \psi, y)$,
4. $\pi(\psi|\theta, \Sigma, b, y) = \pi(\psi|b)$,

and uses a Metropolis-Hastings step when a full conditional density is intractable. Interest may also lie in the joint posterior $\pi(\theta, \psi, \Sigma, b, x|y)$ where, since no imputation is required for the LNA, $x^i = (x_{t_0}, \ldots, x_{t_n})'$ and $x = (x^1, \ldots, x^N)$. Realisations from this posterior can be obtained using the above Gibbs sampler with an extra step that draws from $\pi(x^i|\theta, \psi, \Sigma, b^i, y^i) = \pi(x^i|\theta, \Sigma, b^i, y^i)$ for $i = 1, \ldots, N$. The method uses a forward filter, backward sampling (FFBS) algorithm; see West and Harrison (1997) for full details of a FFBS. Note that the backward sweep requires $\text{Cov}(X^i_{t_{j+1}}, X^i_{t_j})$, which from Chapter 3 we know to be

$$\text{Cov}\left(X^i_{t_{j+1}}, X^i_{t_j}\right) = \text{Cov}\left(\hat{R}^i_{t_{j+1}}, \hat{R}^i_{t_j}\right) = P^i_{t_{j+1}}\text{Var}\left(\hat{R}^i_{t_j}\right).$$

Here P^i_t is a $d \times d$ matrix that can be shown to satisfy the ODE

$$\frac{dP^i_t}{dt} = H^i_t P^i_t, \qquad P^i_0 = I_d, \tag{4.24}$$

cf. equation (2.35).

We now describe how to compute the marginal likelihood $\pi(y^i|\theta, \Sigma, b^i)$ for each experimental unit. To ease the notation, consider a single experimental unit and drop i from the notation. Since the parameters θ, ψ, b and Σ remain fixed throughout the calculation, we also drop them from the notation where possible. Define $y_{0:j} = (y_{t_0}, \ldots, y_{t_j})'$. Now suppose that $X_0 \sim N(a, C)$ *a priori*. The marginal likelihood $\pi(y|\theta, \Sigma, b)$ under the LNA

Algorithm 7 LNA FFBS

1. **Forward filter:** Initialisation. Compute $\pi(y_{t_0}) = N(y_{t_0}\,;\, F'a\,,\, F'CF + \Sigma)$. The posterior at time $t_0 = 0$ is therefore $X_{t_0}|y_{t_0} \sim N(a_0, C_0)$, where

$$a_0 = a + CF\left(F'CF + \Sigma\right)^{-1}\left(y_{t_0} - F'a\right)$$
$$C_0 = C - CF\left(F'CF + \Sigma\right)^{-1}F'C\,.$$

   Store the values of a_0 and C_0.

2. For $j = 0, 1, \ldots, n-1$,

   (a) Prior at t_{j+1}. Initialise the LNA with $\eta_{t_j} = a_{t_j}$, $V_{t_j} = C_{t_j}$ and $P_{t_j} = I_d$. Integrate the ODEs (4.21), (4.22) and (4.24) forward to t_{j+1} to obtain $\eta_{t_{j+1}}$, $V_{t_{j+1}}$ and $P_{t_{j+1}}$. Hence $X_{t_{j+1}}|y_{0:j+1} \sim N(\eta_{t_{j+1}}, V_{t_{j+1}})$.

   (b) One step forecast. Using the observation equation (4.2), we have that

$$Y_{t_{j+1}}|y_{0:j} \sim N\left(F'\eta_{t_{j+1}}, F'V_{t_{j+1}}F + \Sigma\right).$$

   Compute the updated marginal likelihood

$$\begin{aligned}\pi(y_{0:j+1}) &= \pi(y_{0:j})\pi(y_{t_{j+1}}|y_{0:j}) \\ &= \pi(y_{0:j})N\left(y_{t_{j+1}}\,;\, F'\eta_{t_{j+1}}\,,\, F'V_{t_{j+1}}F + \Sigma\right).\end{aligned}$$

   (c) Posterior at t_{j+1}. Combining the distributions in (a) and (b) gives the joint distribution of $X_{t_{j+1}}$ and $Y_{t_{j+1}}$ (conditional on $y_{0:j}$) as

$$\begin{pmatrix} X_{t_{j+1}} \\ Y_{t_{j+1}} \end{pmatrix} \sim N\left\{\begin{pmatrix} \eta_{t_{j+1}} \\ F'\eta_{t_{j+1}} \end{pmatrix}, \begin{pmatrix} V_{t_{j+1}} & V_{t_{j+1}}F \\ F'V_{t_{j+1}} & F'V_{t_{j+1}}F + \Sigma \end{pmatrix}\right\}$$

   and therefore $X_{t_{j+1}}|y_{0:j+1} \sim N(a_{t_{j+1}}, C_{t_{j+1}})$, where

$$a_{t_{j+1}} = \eta_{t_{j+1}} + V_{t_{j+1}}F\left(F'V_{t_{j+1}}F + \Sigma\right)^{-1}\left(y_{t_{j+1}} - F'\eta_{t_{j+1}}\right)$$
$$C_{t_{j+1}} = V_{t_{j+1}} - V_{t_{j+1}}F\left(F'V_{t_{j+1}}F + \Sigma\right)^{-1}F'V_{t_{j+1}}\,.$$

   Store the values of $a_{t_{j+1}}$, $C_{t_{j+1}}$, $\eta_{t_{j+1}}$, $V_{t_{j+1}}$ and $P_{t_{j+1}}$.

(and, if desired, realisations from $\pi(\theta, \psi, \Sigma, b, x|y)$) can be obtained using Algorithm 7. If the ODE system governing the LNA is not restarted, the forward filter of Algorithm 7 can be seen as a standard Kalman filter (Kalman, 1960). Note that if no interest lies in the marginal posterior density of the latent states $\pi(x|\theta, \psi, \Sigma, b, y)$, then (4.24) need not be solved and no storage of values is necessary.

Algorithm 7 continued

Sample $\pi(x|y)$ using the following backward sampler.

1. **Backward sampler:** First draw x_{t_n} from $X_{t_n}|y \sim N(a_{t_n}, C_{t_n})$.
2. For $j = n-1, n-2, \ldots, 0$,
   (a) Joint distribution of X_{t_j} and $X_{t_{j+1}}$. Note that $X_{t_j}|y_{0:j} \sim N(a_{t_j}, C_{t_j})$. The joint distribution of X_{t_j} and $X_{t_{j+1}}$ (conditional on $y_{0:j}$) is

$$\begin{pmatrix} X_{t_j} \\ X_{t_{j+1}} \end{pmatrix} \sim N\left\{ \begin{pmatrix} a_{t_j} \\ \eta_{t_{j+1}} \end{pmatrix}, \begin{pmatrix} C_{t_j} & C_{t_j} P'_{t_{j+1}} \\ P_{t_{j+1}} C_{t_j} & V_{t_{j+1}} \end{pmatrix} \right\}.$$

   (b) Backward distribution. The distribution of $X_{t_j}|X_{t_{j+1}}, y_{0:j}$ is $N(\hat{a}_{t_j}, \hat{C}_{t_j})$, where

$$\begin{aligned} \hat{a}_{t_j} &= a_{t_j} + C_{t_j} P'_{t_{j+1}} V^{-1}_{t_{j+1}} \left(x_{t_{j+1}} - \eta_{t_{j+1}} \right), \\ \hat{C}_{t_j} &= C_{t_j} - C_{t_j} P'_{t_{j+1}} V^{-1}_{t_{j+1}} P_{t_{j+1}} C_{t_j}. \end{aligned}$$

   Draw x_{t_j} from $X_{t_j}|X_{t_{j+1}}, y_{0:j} \sim N(\hat{a}_{t_j}, \hat{C}_{t_j})$.

4.7 Summary

We have introduced the concept of stochastic differential mixed-effects models, before describing a framework that permits (simulation-based) Bayesian inference for a large class of multivariate SDMEMs using discrete-time observations, which may be incomplete and subject to measurement error. We have outlined in detail the steps necessary to implement both path and parameter updates. By adopting a Bayesian imputation approach, we have shown how the modified innovation scheme of Golightly and Wilkinson (2008) can be applied to SDMEMs. This approach overcomes the problematic dependence between the latent process and any parameters that feature in the diffusion coefficient. The computational cost of such a scheme is determined by the number of imputed points $(m-1)$ required between observation times.

We also presented a tractable approximation to the SDMEM, the linear noise approximation, and detailed an inference scheme based upon it, exploiting the computational efficiency of Fearnhead et al. (2014). The computational efficiency of the LNA is however linked to the dimension of the SDE driving the SDMEM; for a d-dimensional SDE system, the LNA requires the solution of a system of order d^2 coupled ODEs. Of course, increases in both computational and overall efficiency can be achieved if the resulting ODE system can be solved analytically. If the ODE system governing the LNA is intractable (so that it must be solved numerically) the computational advantage of using the LNA over

the imputation approach will be reduced. We note that, as the dimension d of the SDE increases, the LNA is likely to become infeasible.

Having now outlined two competing inference schemes—a Bayesian imputation approach incorporating the modified innovation scheme, and a scheme based on the linear noise approximation—we now assess their respective performances. To that end we implement both schemes in two examples in the next chapter and compare their performances.

Chapter 5

Numerical examples

We now compare the accuracy and efficiency of our Bayesian imputation approach (coupled with the modified innovation scheme) with an LNA-based solution. Details of both schemes can be found in the previous chapter. We consider two scenarios: one in which the ODEs governing the LNA are tractable and one in which a numerical solver is required. In the first we use synthetic data generated from a simple univariate SDE description of orange tree growth (Picchini and Ditlevsen, 2011). The second example uses real data taken from Matis et al. (2008) to fit an SDMEM driven by the bivariate diffusion approximation of a stochastic kinetic model of aphid dynamics; see Appendix C for an outline of the diffusion approximation of a stochastic kinetic model. The resulting SDMEM is particularly challenging to fit as both the drift and diffusion functions are nonlinear and also only one component of the model is observed (with error). Furthermore, we compare inferences made under the SDMEM and LNA using synthetic data generated under four data-poor scenarios for the model of aphid dynamics.

5.1 Orange tree growth

The SDMEM developed by Picchini et al. (2010) and Picchini and Ditlevsen (2011) to model orange tree growth describes the dynamics of the circumference X_t^i (in mm) of individual trees by

$$dX_t^i = \frac{1}{\phi_1^i \phi_2^i} X_t^i \left(\phi_1^i - X_t^i\right) dt + \sigma \sqrt{X_t^i}\, dW_t^i, \quad X_0^i = x_0^i, \quad i = 1, \ldots, N,$$

with $\phi_1^i \sim N(\phi_1, \sigma_{\phi_1}^2)$ and $\phi_2^i \sim N(\phi_2, \sigma_{\phi_2}^2)$ independently. Here $\theta = \sigma$ is common to all trees, the random effects are $b^i = (\phi_1^i, \phi_2^i)'$, $i = 1, \ldots, N$ and the parameter vector governing the random effects distributions is $\psi = (\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2})'$. Note that the ϕ_1^i can be in-

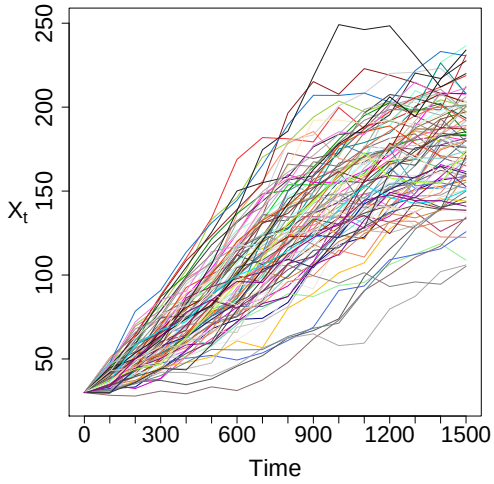

Figure 5.1: Synthetic data for the orange tree growth model for the parameter values $\phi_1 = 195$, $\phi_2 = 350$, $\sigma_{\phi_1} = 25$, $\sigma_{\phi_2} = 52.5$ and $\sigma = 0.08$, with $x_0^i = 30$.

terpreted as asymptotic circumferences and the ϕ_2^i as the time between the inflection point of the model obtained by ignoring stochasticity and the point where $X_t^i = \phi_1^i/(1 + e^{-1})$.

To allow identifiability of all model parameters, we generated 16 observations for the circumference of $N = 100$ trees at intervals of 100 days. Following Picchini and Ditlevsen (2011) we gave each tree the same initial condition ($x_0^i = 30$) and took

$$(\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2}, \sigma) = (195, 350, 25, 52.5, 0.08),$$

which gives random effects distributions $\phi_1^i \sim N(195, 25^2)$ and $\phi_2^i \sim N(350, 52.5^2)$. The synthetic data are shown in Figure 5.1. For our analysis of these data we assumed the parameters to be independent *a priori* with ϕ_1 and ϕ_2 having weak $N(0, 100^2)$ priors, and $1/\sigma_{\phi_1}^2$, $1/\sigma_{\phi_2}^2$ and $1/\sigma^2$ having weak gamma $Ga(1, 0.01)$ priors. In this example we assume there is no measurement error and therefore the target posterior is given by

$$\begin{aligned}\pi(\theta, \psi, b|x) &\propto \pi(\theta)\pi(\psi)\pi(b|\psi)\pi(x|\theta, b) \\ &\propto \pi(\phi_1)\pi(\phi_2)\pi(\sigma_{\phi_1})\pi(\sigma_{\phi_2})\pi(\sigma) \prod_{i=1}^{N} \pi\left(\phi_1^i|\phi_1, \sigma_{\phi_1}\right) \prod_{i=1}^{N} \pi\left(\phi_2^i|\phi_2, \sigma_{\phi_2}\right) \pi(x|\theta, b).\end{aligned}$$

In the Bayesian imputation approach, $\pi(x|\theta, b)$ is as in (4.7) whereas for the LNA–based solution

$$\pi(x|\theta, b) = \prod_{i=1}^{N} \prod_{j=0}^{n-1} N\left(x_{t_{j+1}}^i \,;\, \eta_{t_{j+1}}^i, V_{t_{j+1}}^i\right),$$

where, for each interval $[t_j, t_{j+1}]$ and each tree i, the η_t^i and V_t^i satisfy the ODE system

$$\frac{d\eta_t^i}{dt} = \frac{1}{\phi_1^i \phi_2^i} \eta_t^i \left(\phi_1^i - \eta_t^i\right), \qquad \eta_{t_j}^i = x_{t_j}^i,$$
$$\frac{dV_t^i}{dt} = \frac{2}{\phi_1^i \phi_2^i} \left(\phi_1^i - 2\eta_t^i\right) V_t^i + \sigma^2 \eta_t^i, \qquad V_{t_j}^i = 0.$$

Fortunately this ODE system can be solved analytically giving

$$\eta_t^i = \frac{A\phi_1^i e^{t/\phi_2^i}}{1 + Ae^{t/\phi_2^i}}$$
$$V_t^i = B\left(\frac{1}{2}A^3\phi_2^i e^{2t/\phi_2^i} + 3A^2\phi_2^i e^{t/\phi_2^i} - \phi_2^i e^{-t/\phi_2^i} + 3At - \frac{1}{2}A^3\phi_2^i - 3A^2\phi_2^i + \phi_2^i\right),$$

where

$$A = \frac{x_0^i}{\phi_1^i - x_0^i} \quad \text{and} \quad B = \frac{\sigma^2 A\phi_1^i e^{2t/\phi_2^i}}{(1 + Ae^{t/\phi_2^i})^4}.$$

This solution is derived in Appendix B.

The MCMC scheme can make use of simple semi-conjugate updates for ϕ_1, ϕ_2, σ_{ϕ_1} and σ_{ϕ_2}; see Appendix A.2 for details of these updates. However the remaining parameters (σ and the b^i) require Metropolis-within-Gibbs updates and we have found that componentwise normal random walk updates (so-called random walk Metropolis) on the log scale work particularly well. Also, for the modified innovation scheme, the dynamics of the SDMEM permit the use of the modified diffusion bridge construct (see Section 3.1.2) to update the latent trajectories between observation times: the improved bridge constructs of Section 3.2 are not needed for this example.

The modified innovation scheme requires specification of the level of discretisation m. We performed several short pilot runs of the scheme with $m \in \{5, 10, 20, 40\}$ and found no discernible difference in posterior output for $m \geq 10$. We therefore took $m = 10$. The marginal posterior densities for a selection of parameters using these various levels of discretisation are illustrated in Figure 5.2. The sample output was also used to estimate the marginal posterior variances of σ and the b^i, and thereby provide sensible innovation variances in the random walk Metropolis updates. We note that the computational effort required to find a suitable value of m is typically a small percentage of the total cost of using the Bayesian imputation approach. This is true for both examples considered within this chapter. Both the modified innovation scheme and the LNA–based scheme required a burn in of 500 iterations, a thin of 100 iterates and were run long enough to yield a sample of approximately 10K independent posterior draws. Figure 5.3 shows the marginal posterior densities and autocorrelations for the common parameter σ and the parameters

	ϕ_1	ϕ_2	σ_{ϕ_1}	σ_{ϕ_2}	σ
Imputation	194.229	344.799	24.316	53.219	0.079
	(3.509)	(10.098)	(3.149)	(10.410)	(0.002)
LNA	194.634	347.631	24.207	53.960	0.079
	(4.025)	(10.844)	(3.154)	(10.193)	(0.002)

Table 5.1: Marginal posterior means (standard deviations) of the random effects hyper-parameters (ϕ_1, ϕ_2, σ_{ϕ_1}, σ_{ϕ_2}) and common parameter σ in the orange tree growth SDMEM. The synthetic data used $\phi_1 = 195$, $\phi_2 = 350$, $\sigma_{\phi_1} = 25$, $\sigma_{\phi_2} = 52.5$ and $\sigma = 0.08$.

governing the random effects distributions. The bivariate marginal posterior densities for these parameters are illustrated in Figures 5.4 and 5.5. Figures 5.6 and 5.7 show the marginal posterior densities of five randomly chosen random effects. The marginal posterior means and standard deviations of $(\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2}, \sigma)$ are given in Table 5.1. The figures and table show that both the imputation approach and LNA–based approach generally give similar output and are consistent with the true values from which the data were simulated. There are however slight differences in some of the bivariate marginal posterior densities obtained, see (for example), the density for ϕ_2 and σ_{ϕ_2} in Figure 5.5.

Both schemes were coded in C and run on an Intel Xeon 3.0GHz processor. The modified innovation scheme took 43504.3 seconds to run whilst the LNA inference scheme took 2483.3 seconds. We use the minimum (over each parameter chain) effective sample size (minESS) to measure the statistical efficiency of each scheme. The modified innovation scheme produced a minESS of 7949.0 and the LNA–based approach gave 7820.6. Therefore, in terms of minESS/sec, using the LNA outperforms the imputation approach in this example by a factor of approximately 17. It should be noted, however, that for most nonlinear SDMEMs the ODEs governing the LNA solution will rarely be tractable and the consequent use of a numerical scheme will degrade its performance.

In the next section we consider an example in which the LNA ODEs are intractable.

5.2 Cotton aphid dynamics

5.2.1 Model and data

Aphids (also known as plant lice or greenfly) are small sap sucking insects which live on the leaves of plants. As they suck the sap they also secrete honey-dew which forms a protective cover over the leaf, ultimately resulting in aphid starvation. Matis et al. (2006) describe a model for aphid dynamics in terms of population size (N_t) and cumulative population size (C_t). The model is a stochastic birth-death model with linear birth rate λN_t and death rate $\mu N_t C_t$. The key probabilistic laws governing the time-evolution of the

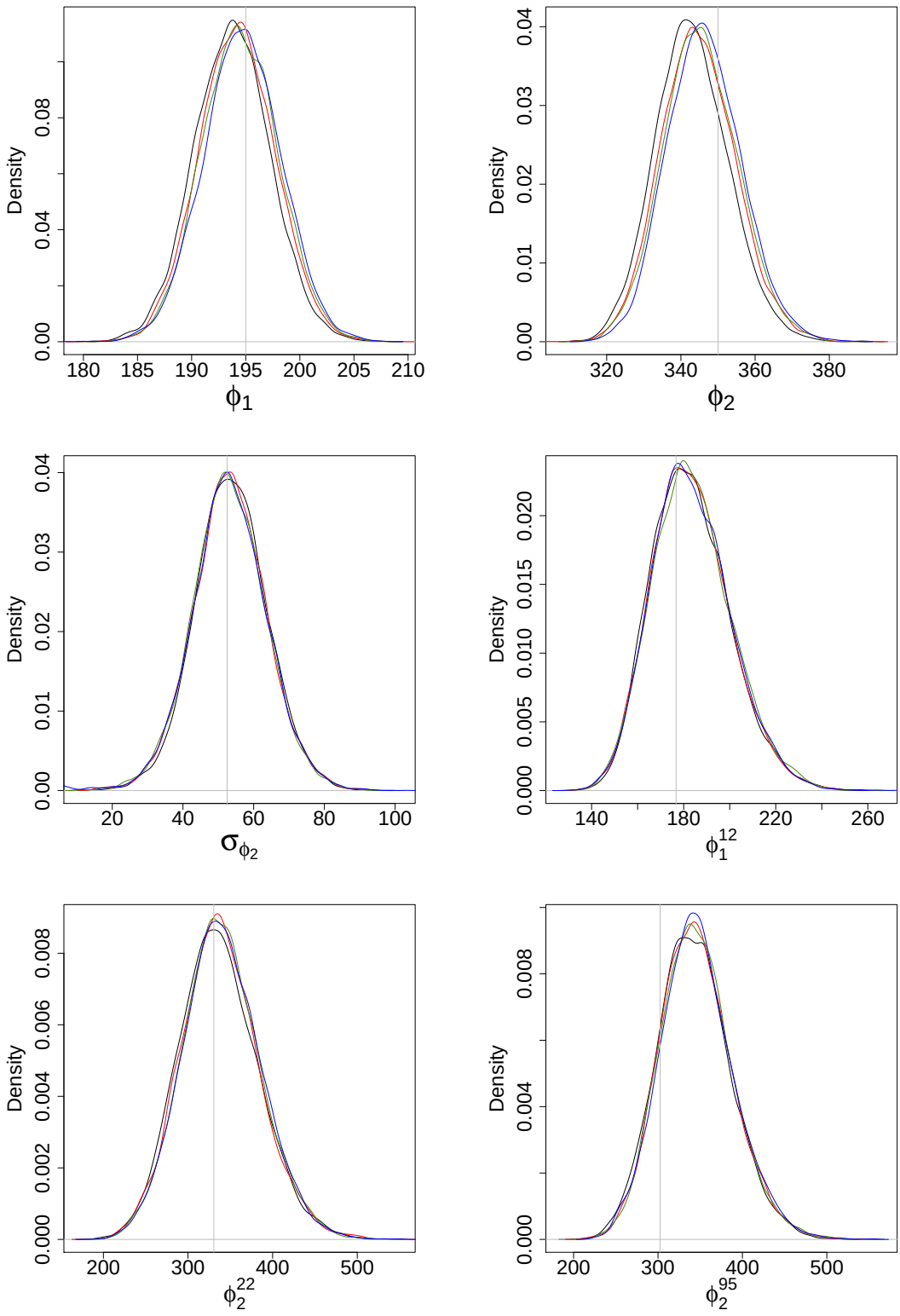

Figure 5.2: Marginal posterior densities for a selection of the orange tree growth model parameters using various levels of discretisation m. Black: $m = 5$. Red: $m = 10$. Green: $m = 20$. Blue: $m = 40$. The grey lines indicate the ground truth.

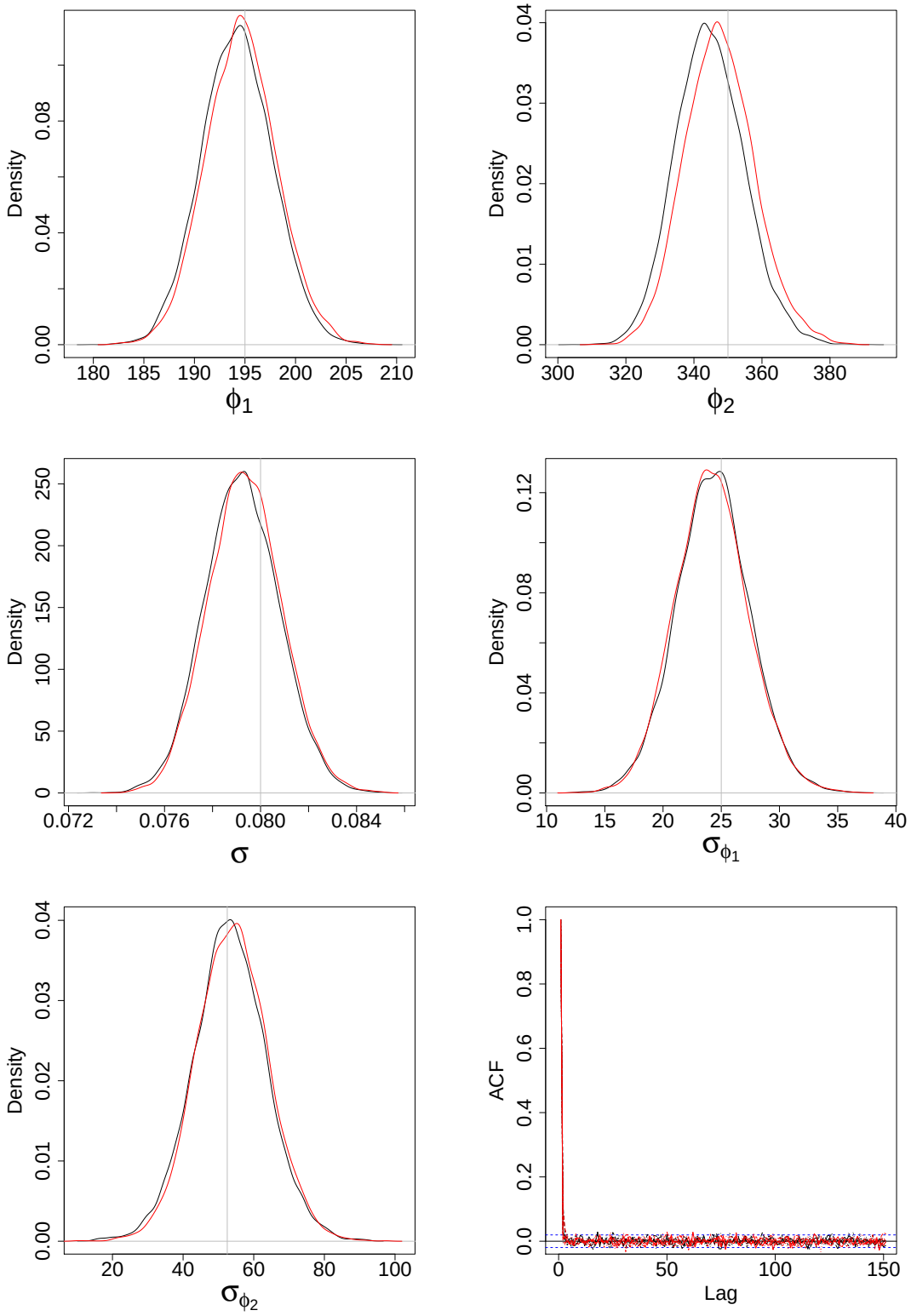

Figure 5.3: Marginal posterior densities for the random effects hyper-parameters (ϕ_1, ϕ_2, σ_{ϕ_1}, σ_{ϕ_2}) and common parameter σ in the orange tree growth SDMEM, together with their (overlayed) autocorrelation functions. Black: Bayesian imputation. Red: LNA. The grey lines indicate the ground truth.

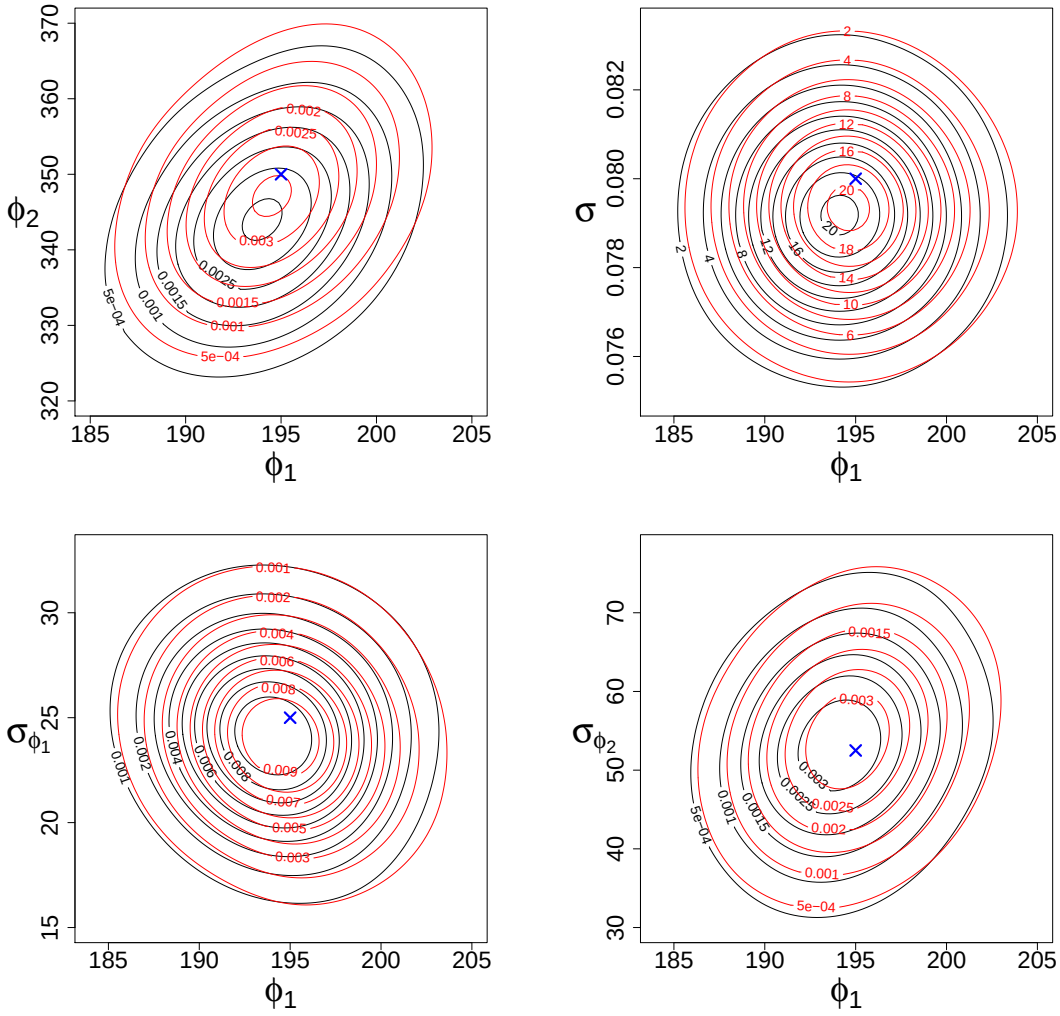

Figure 5.4: Bivariate marginal posterior densities for the random effects hyper-parameters (ϕ_1, ϕ_2, σ_{ϕ_1}, σ_{ϕ_2}) and common parameter σ in the orange tree growth SDMEM. Black: Bayesian imputation. Red: LNA. The blue crosses indicate the ground truth.

process over a small interval $(t, t+dt]$ are

$$\begin{aligned} \Pr(N_{t+dt} = n_t + 1,\, C_{t+dt} = c_t + 1 \,|\, n_t, c_t) &= \lambda n_t \, dt + o(dt), \\ \Pr(N_{t+dt} = n_t - 1,\, C_{t+dt} = c_t \,|\, n_t, c_t) &= \mu n_t c_t \, dt + o(dt). \end{aligned} \tag{5.1}$$

The diffusion approximation of the Markov jump process defined by (5.1) is

$$\begin{pmatrix} dN_t \\ dC_t \end{pmatrix} = \begin{pmatrix} \lambda N_t - \mu N_t C_t \\ \lambda N_t \end{pmatrix} dt + \begin{pmatrix} \lambda N_t + \mu N_t C_t & \lambda N_t \\ \lambda N_t & \lambda N_t \end{pmatrix}^{1/2} dW_t, \quad X_0 = x_0. \tag{5.2}$$

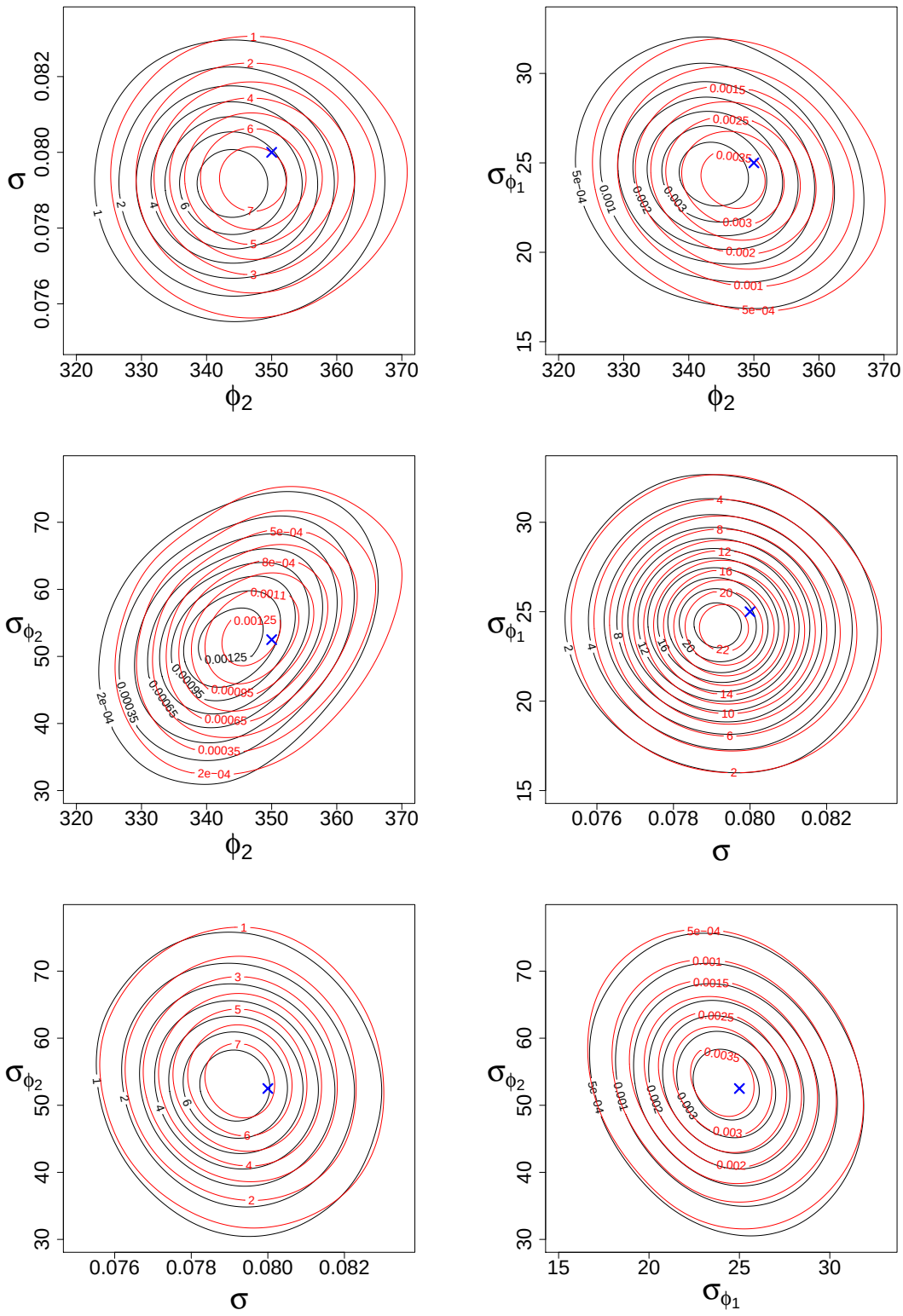

Figure 5.5: Bivariate marginal posterior densities for the random effects hyper-parameters (ϕ_1, ϕ_2, σ_{ϕ_1}, σ_{ϕ_2}) and common parameter σ in the orange tree growth SDMEM. Black: Bayesian imputation. Red: LNA. The blue crosses indicate the ground truth.

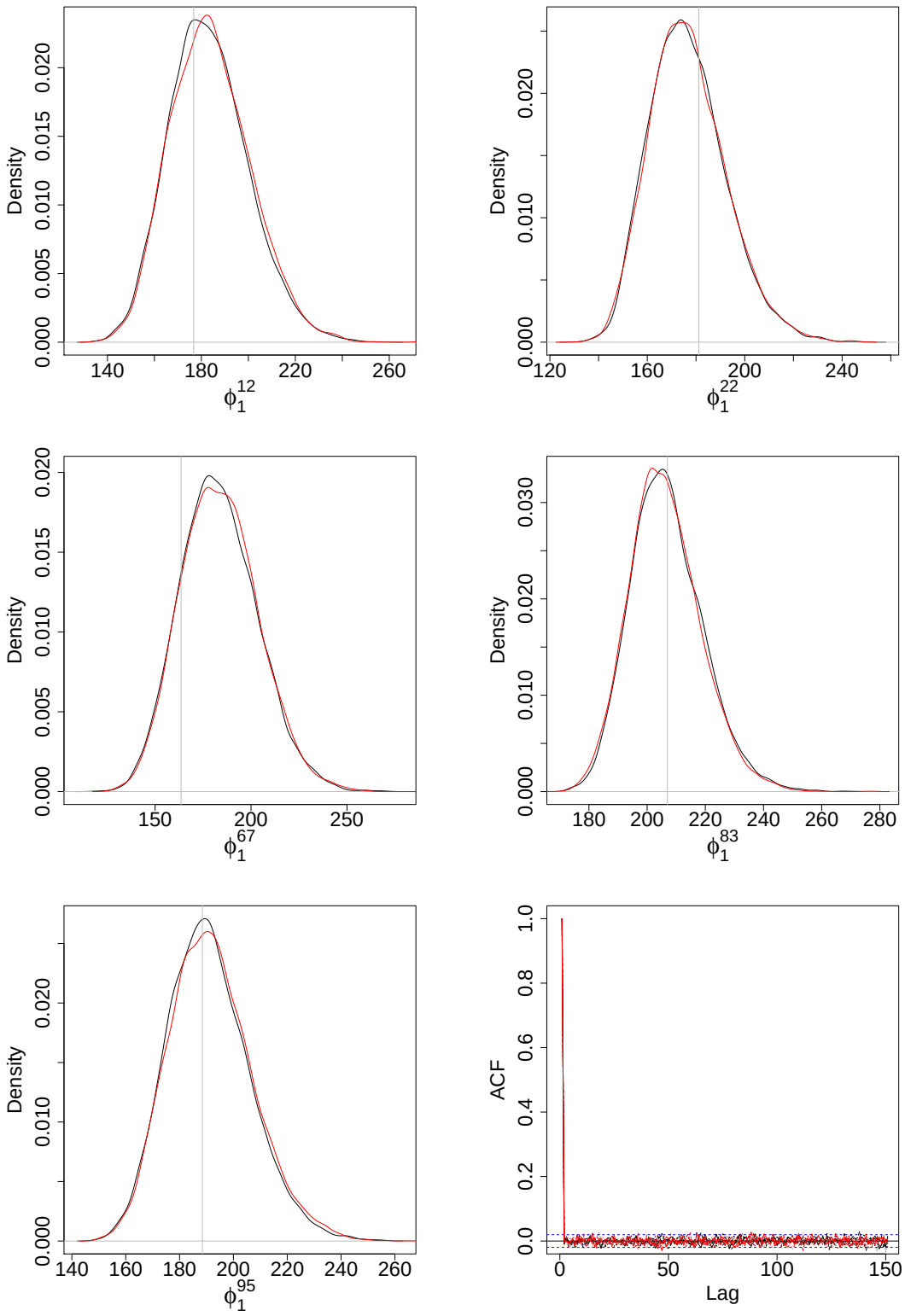

Figure 5.6: Marginal posterior densities for a random selection of ϕ_1^i in the orange tree growth SDMEM, together with their (overlayed) autocorrelation functions. Black: Bayesian imputation. Red: LNA. The grey lines indicate the ground truth.

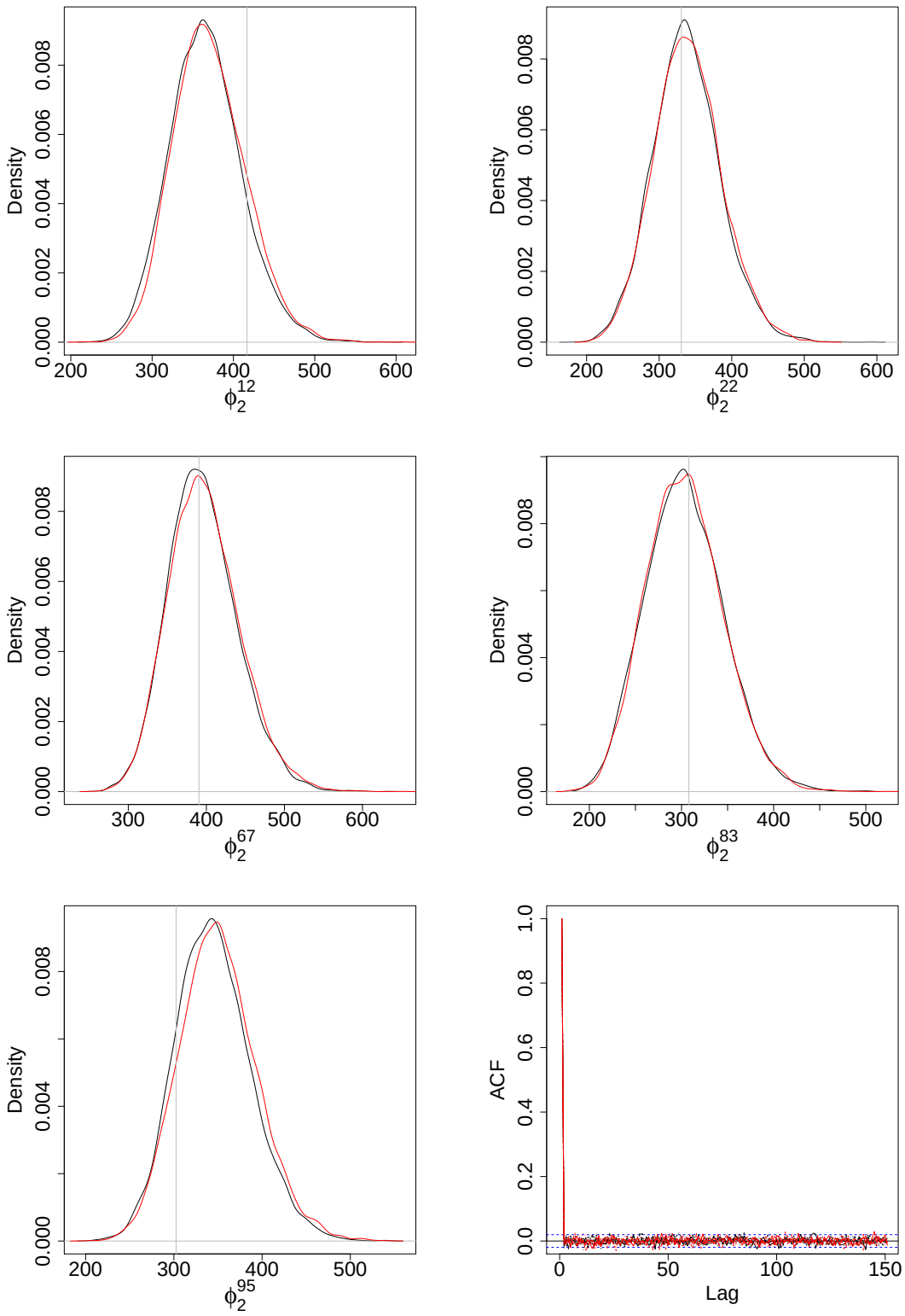

Figure 5.7: Marginal posterior densities for a random selection of ϕ_2^i in the orange tree growth SDMEM, together with their (overlayed) autocorrelation functions. Black: Bayesian imputation. Red: LNA. The grey lines indicate the ground truth.

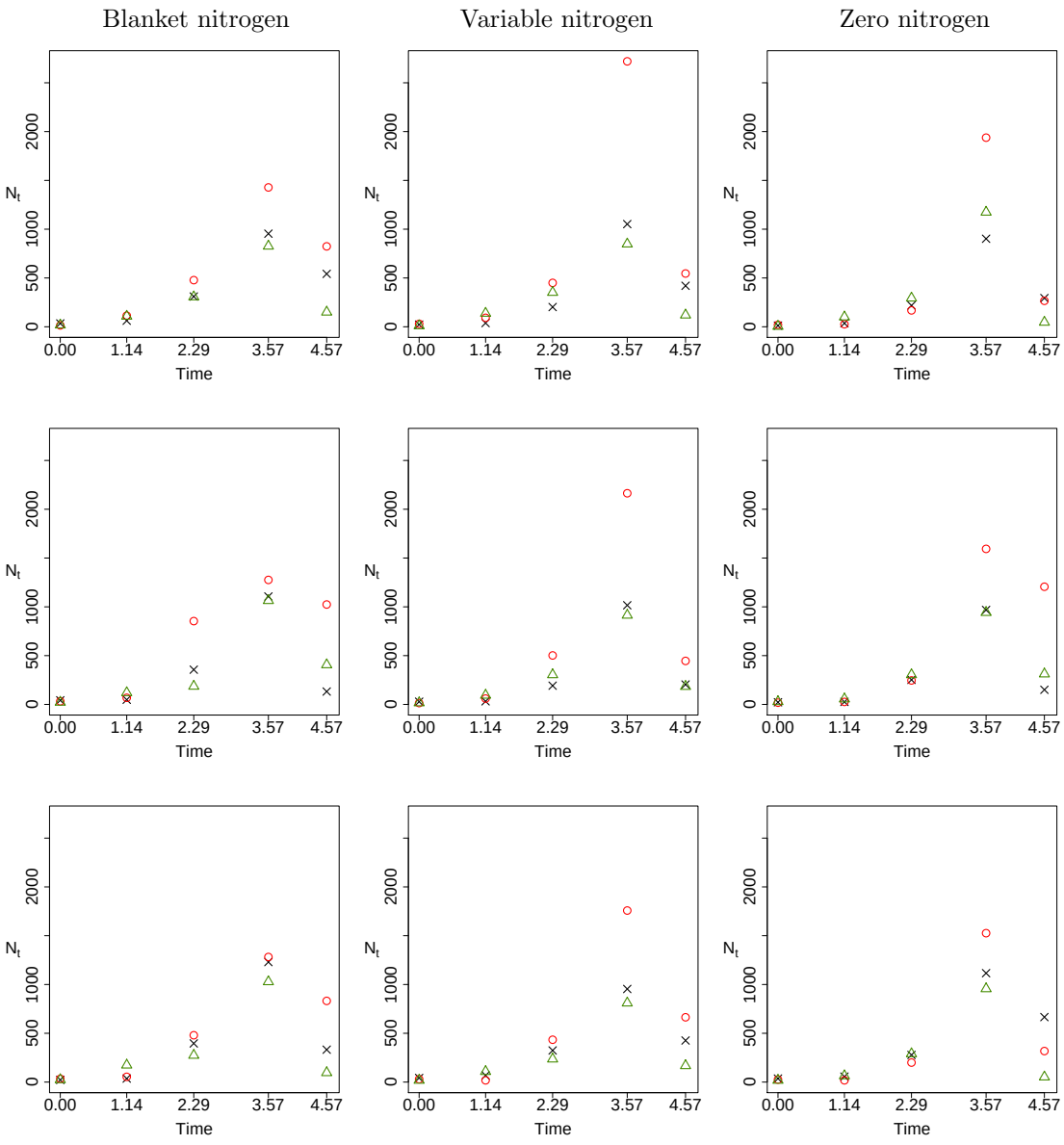

Figure 5.8: Aphid numbers (N_t) against time (in weeks) taken from Matis et al. (2008). Low water (1st row), medium water (2nd row) and high water (3rd row). Black crosses: Block 1. Red circles: Block 2. Green triangles: Block 3.

Matis et al. (2008) also provide a dataset of cotton aphid (*Aphis gossypii*) counts collected from three blocks/plots (1/2/3) and using treatments constructed from two factors: water irrigation (low/medium/high) and nitrogen (blanket/variable/none). The data were collected in July 2004 in Lamesa, Texas and consist of five observations of aphid counts aggregated over twenty randomly chosen leaves in each plot for the twenty-seven treatment-block combinations. The data were recorded at times $t = 0, 1.14, 2.29, 3.57$ and 4.57 weeks

(approximately every 7/8 days). The data are illustrated in Figure 5.8.

We now formulate an appropriate SDMEM model driven by (5.2) for these data and then fit the model. For notational simplicity, let i, j, k denote the level of water, nitrogen and block number respectively with $i, j, k \in \{1, 2, 3\}$, where 1 represents low water/blanket nitrogen, 2 represents medium water/variable nitrogen and 3 represents high water/zero nitrogen. Let N_t^{ijk} denote the number of aphids at time t for combination ijk and C_t^{ijk} denote the corresponding cumulative population size. We write $X_t^{ijk} = (N_t^{ijk}, C_t^{ijk})'$ and consider the SDMEM

$$dX_t^{ijk} = \alpha\left(X_t^{ijk}, b^{ijk}\right) dt + \sqrt{\beta\left(X_t^{ijk}, b^{ijk}\right)}\, dW_t^{ijk}, \quad X_0^{ijk} = x_0^{ijk}, \quad i, j, k \in \{1, 2, 3\},$$

where

$$\alpha\left(X_t^{ijk}, b^{ijk}\right) = \begin{pmatrix} \lambda^{ijk} N_t^{ijk} - \mu^{ijk} N_t^{ijk} C_t^{ijk} \\ \lambda^{ijk} N_t^{ijk} \end{pmatrix},$$
$$\beta\left(X_t^{ijk}, b^{ijk}\right) = \begin{pmatrix} \lambda^{ijk} N_t^{ijk} + \mu^{ijk} N_t^{ijk} C_t^{ijk} & \lambda^{ijk} N_t^{ijk} \\ \lambda^{ijk} N_t^{ijk} & \lambda^{ijk} N_t^{ijk} \end{pmatrix}.$$

The fixed effects $b^{ijk} = (\lambda^{ijk}, \mu^{ijk})'$ have a standard structure which allows for main factor and block effects and single factor-block interactions, with

$$\begin{aligned} \lambda^{ijk} &= \lambda + \lambda_{W_i} + \lambda_{N_j} + \lambda_{B_k} + \lambda_{WN_{ij}} + \lambda_{WB_{ik}} + \lambda_{NB_{jk}} \\ \mu^{ijk} &= \mu + \mu_{W_i} + \mu_{N_j} + \mu_{B_k} + \mu_{WN_{ij}} + \mu_{WB_{ik}} + \mu_{NB_{jk}}. \end{aligned} \tag{5.3}$$

Also for identifiability we use the corner constraints $\lambda_{W_1} = \lambda_{N_1} = \lambda_{B_1} = 0$,

$$\lambda_{WN_{ij}} = \lambda_{WN_{ij}}(1 - \kappa_{ij}), \quad \lambda_{WB_{ik}} = \lambda_{WB_{ik}}(1 - \kappa_{ik}) \quad \text{and} \quad \lambda_{NB_{jk}} = \lambda_{NB_{jk}}(1 - \kappa_{jk}),$$

where $\kappa_{rs} = \max(\delta_{1r}, \delta_{1s})$ and $\delta_{..}$ is the Kronecker delta, with equivalent constraints on the death rates. The interpretation of (5.3) is straightforward. For example, $\lambda^{111} = \lambda$ and $\mu^{111} = \mu$ are the baseline birth and death rates inferred using all $5 \times 3^3 = 135$ observations, and correspond to the treatment combination low water, blanket nitrogen and block 1. Likewise, all $5 \times 3^2 = 45$ observations taken from block 2 inform the main effects of block 2 (λ_{B_2} and μ_{B_2}) relative to the baseline.

A related approach can be found in Gillespie and Golightly (2010), where the diffusion approximation is eschewed in favour of a further approximation via moment closure. Our approach differs further from theirs by allowing for measurement error and leads to a much improved predictive fit. The measurement error model is in part motivated by an over-dispersed Poisson error structure which we then approximate by a Gaussian distribution.

Specifically, we assume that aphid population size N_t is observed with Gaussian error and that the error variance is proportional to the latent aphid numbers, giving

$$Y_t^{ijk}|N_t^{ijk},\sigma \overset{indep}{\sim} N\left(N_t^{ijk},\sigma^2 N_t^{ijk}\right), \quad t=0,1.14,2.29,3.57,4.57. \tag{5.4}$$

5.2.2 Implementation

Our prior beliefs for $1/\sigma^2$ are described by a $Ga(a,a)$ distribution. We found little difference in results for $a \in \{0.01, 0.1, 1\}$ and so here we report results for $a = 1$. The prior for the elements in (5.3) consists of independent components subject to the birth and death rates for each treatment combination $(\lambda^{ijk}, \mu^{ijk})$ being positive. The baseline rates λ and μ must be positive and so, following Gillespie and Golightly (2010), we assign weak $U(-10, 10)$ priors to $\log\lambda$ and $\log\mu$ and also to the remaining parameters. We also take a fairly weak $N(24, 90)$ prior for each $N_{t_0}^{ijk}$ and use a proposal of the form $N(N_{t_0},\sigma^2 N_{t_0})$ for updates. The cumulative population sizes must be at least as large as their equivalent population size. However, we do not expect them to be greatly different *a priori*. We investigated using a truncated distribution of the form $C_{t_0}|N_{t_0} \sim N(N_{t_0}, d_c^2)$, $C_{t_0} > N_{t_0}$ as the prior and found that this led to little difference in posterior output for $d_c \in \{1, 10, 100\}$. We have, therefore, chosen to fix $C_{t_0}^{ijk} = N_{t_0}^{ijk}$ in our analysis.

The nonlinear form of the observation model (5.4) can be problematic for the modified innovation scheme. In particular, the proposal mechanism for the path update requires an observation model that is linear in N_t. Therefore, when proposing from the bridge construct (RB) in Section 3.2.1 (that is the residual bridge where we subtract only the drift), we replace Σ in (3.13) and (3.26) with $\sigma^2\eta_{N,t_{j+1}}$, where $\eta_{t_{j+1}} = (\eta_{N,t_{j+1}}, \eta_{C,t_{j+1}})'$ is the solution of (3.22). Since the proposal mechanism is corrected for via the Metropolis-Hastings step, no additional approximations to the target distribution are needed.

In order to obtain a statistically efficient implementation of the modified innovation scheme, we investigate the performance of the modified diffusion bridge construct of Durham and Gallant (2002) (see Section 3.1.2) and the simple residual bridge construct of Section 3.2.1 (where again we subtract only the drift) in a scenario typical of the real dataset. We consider the simple residual bridge construct, as opposed to the more sophisticated improved bridge constructs of Chapter 3, due to the nature of the data, that is, a partial and noisy observation regime, which has associated with it additional computational considerations as discussed in Chapter 3 and Whitaker et al. (2016*b*). Using the simulation study of Gillespie and Golightly (2010), we take $(\lambda,\mu)' = (1.75, 0.00095)'$, $x_0 = (28, 28)'$ and recursively apply the Euler-Maruyama approximation to give $x_{3.57} = (829.08, 1406.07)'$. We then compare the performance of each bridge construct over the final observation interval $[3.57, 4.57]$ by taking $y_{4.57}$ as the median of (5.4) with $\sigma = 1$. Figure 5.9 shows

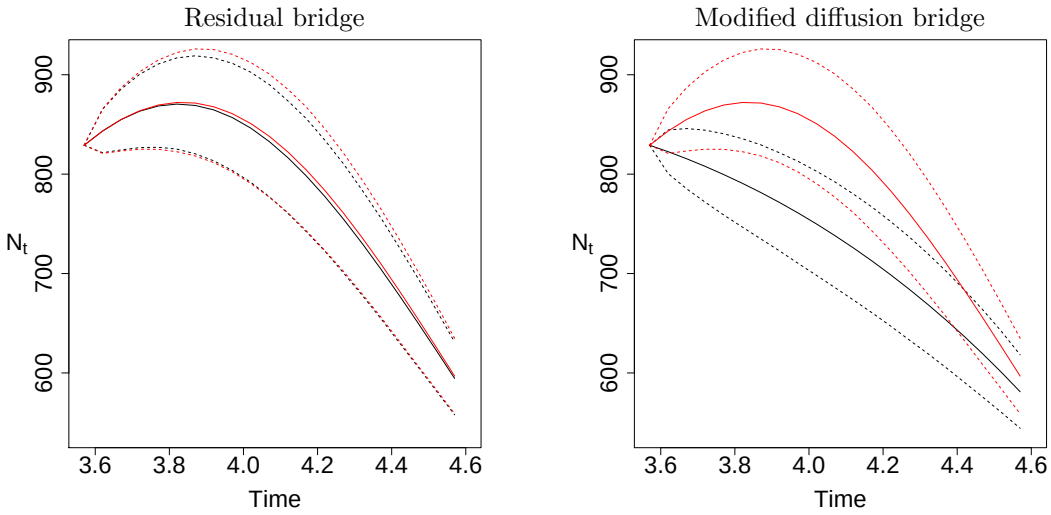

Figure 5.9: 95% credible region (dashed line) and mean (solid line) of the true conditioned aphid population component $N_t|x_{3.57}, y_{4.57}$ (red) and two competing bridge constructs (black).

95% credible regions of the true conditioned process $N_t|x_{3.57}, y_{4.57}$ (found via Monte Carlo simulation) with 95% credible regions obtained by repeatedly simulating from the modified diffusion bridge and our improved construct. It is evident that the modified diffusion bridge fails to adequately account for the nonlinear behaviour of the conditioned process. Use of each construct as a proposal mechanism inside a Metropolis-Hastings independence sampler (100K iterations) results in an estimated acceptance rate of 58% for the residual bridge construct and just 1% for the modified diffusion bridge. It is for these reasons that the modified diffusion bridge is eschewed in favour of our improved bridge construct when applying the Bayesian imputation approach.

Finally, fitting the LNA requires the solution of (4.21) and (4.22) where the Jacobian matrix is

$$H_t^{ijk} = \begin{pmatrix} \lambda^{ijk} - \mu^{ijk}\eta_{C,t}^{ijk} & -\mu^{ijk}\eta_{N,t}^{ijk} \\ \lambda^{ijk} & 0 \end{pmatrix}.$$

Explicitly, the system is given by

$$\begin{aligned} \frac{d\eta_t^{ijk}}{dt} &= \begin{pmatrix} \lambda^{ijk}\eta_{N,t}^{ijk} - \mu^{ijk}\eta_{N,t}^{ijk}\eta_{C,t}^{ijk} \\ \lambda^{ijk}\eta_{N,t}^{ijk} \end{pmatrix}, \qquad \eta_{t_s}^{ijk} = x_{t_s}^{ijk}, \\ \frac{dV_t^{ijk}}{dt} &= \begin{pmatrix} \lambda^{ijk} - \mu^{ijk}\eta_{C,t}^{ijk} & -\mu^{ijk}\eta_{N,t}^{ijk} \\ \lambda^{ijk} & 0 \end{pmatrix} V_t^{ijk} + \begin{pmatrix} \lambda^{ijk}\eta_{N,t}^{ijk} + \mu^{ijk}\eta_{N,t}^{ijk}\eta_{C,t}^{ijk} & \lambda^{ijk}\eta_{N,t}^{ijk} \\ \lambda^{ijk}\eta_{N,t}^{ijk} & \lambda^{ijk}\eta_{N,t}^{ijk} \end{pmatrix} \\ &\quad + V_t^{ijk} \begin{pmatrix} \lambda^{ijk} - \mu^{ijk}\eta_{C,t}^{ijk} & \lambda^{ijk} \\ -\mu^{ijk}\eta_{N,t}^{ijk} & 0 \end{pmatrix}, \qquad V_{t_s}^{ijk} = 0. \end{aligned}$$

In this example, interest also lies in the marginal posterior distribution of the latent states,

and hence (4.24), which here is

$$\frac{dP_t^{ijk}}{dt} = \begin{pmatrix} \lambda^{ijk} - \mu^{ijk}\eta_{C,t}^{ijk} & -\mu^{ijk}\eta_{N,t}^{ijk} \\ \lambda^{ijk} & 0 \end{pmatrix} P_t^{ijk}, \qquad P_{t_s}^{ijk} = I_d,$$

also must be solved.

This ODE system is intractable and so our C implementation uses a standard ODE solver from the GNU scientific library, namely the explicit embedded Runge-Kutta-Fehlberg $(4, 5)$ method. Note that the tractability of the marginal likelihood under the LNA requires a linear Gaussian observation model. Therefore, when applying Algorithm 7 of Chapter 4, we make an approximation to the marginal likelihood calculation by replacing Σ with $\sigma^2\eta_{N,t_{j+1}}$.

The form of the prior for σ gives a semi-conjugate update (see Appendix A.3). The remaining parameters in (5.3) are updated using random walk Metropolis on the pairwise λ, μ component blocks $(\lambda, \mu), (\lambda_{W_2}, \mu_{W_2}), (\lambda_{W_3}, \mu_{W_3}), \ldots, (\lambda_{NB_{33}}, \mu_{NB_{33}})$.

5.2.3 Results

The time between observations is almost but not quite constant and so we have allowed each interval to have its own discretisation level, m. That said, the interval-specific values vary very little, and by at most two for the larger m values. Several short pilot runs of the modified innovation scheme were performed with typical $m \in \{5, 10, 20, 40, 50\}$. These gave no discernible difference in posterior output for $m \geq 20$ and so we took $m = 20$. The sample output was also used to estimate the marginal posterior variances of the λ, μ component blocks of the parameters in (5.3), to be used in the random walk Metropolis updates. Both the modified innovation scheme and MCMC scheme under the LNA were run for 40M iterations with the output thinned by taking every 4Kth iterate to give a final sample of size 10K.

Figure 5.10 shows the marginal posterior densities of the baseline parameters and the parameter σ controlling the observation error variance, with the bivariate marginal posterior densities given in Figure 5.11. A selection of the remaining parameters are displayed in Figure 5.12. As in Gillespie and Golightly (2010) we find that block 2 plays an important role. The 95% credible regions for μ_{B_2}, the main block 2 death rate, and $\lambda_{NB_{22}}$, the birth rate characterising the interaction with nitrogen, are plausibly non-zero. Whilst the imputation approach and LNA generally give consistent output, there are some notable differences. We find, in general, that the LNA tends to underestimate parameter values (and slightly exaggerates the confidence in these estimates) compared to those obtained under the modified innovation scheme.

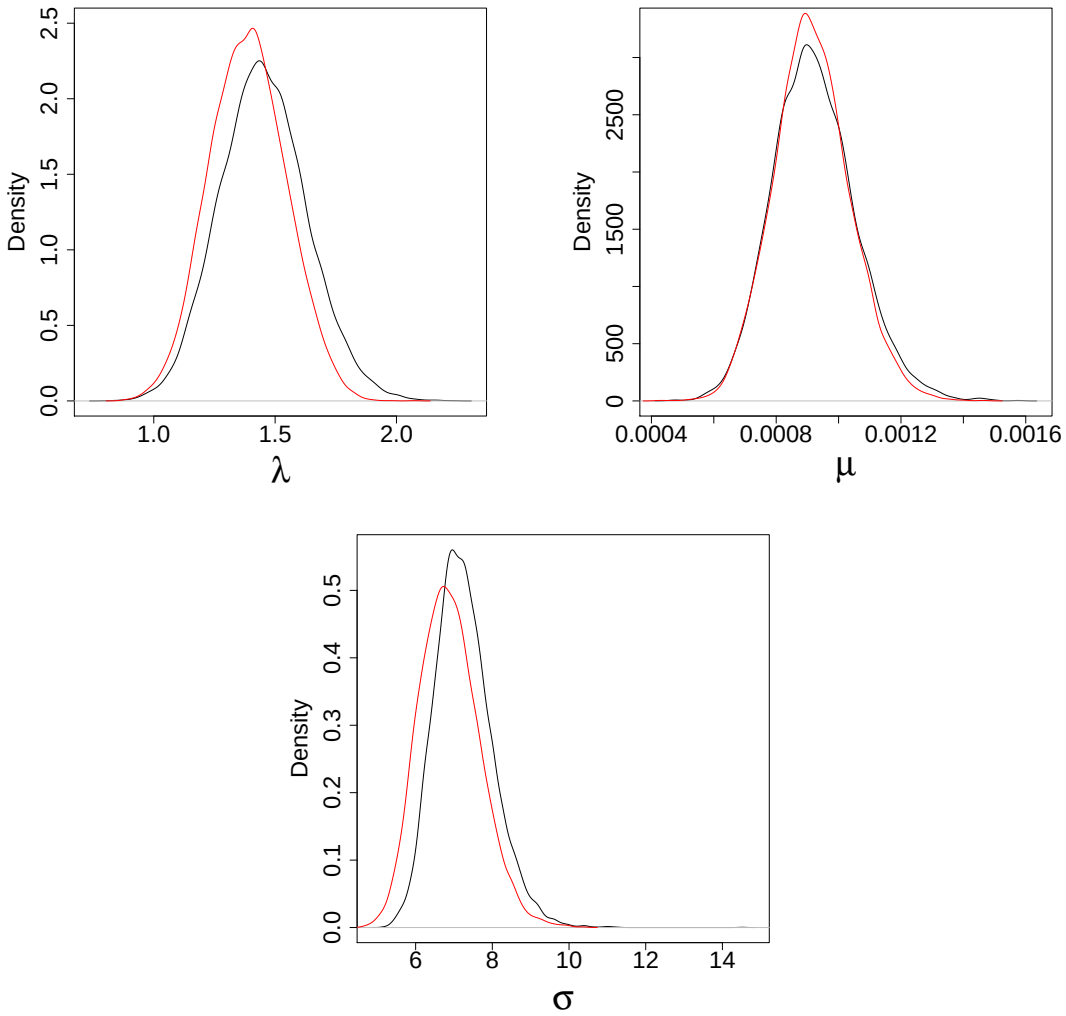

Figure 5.10: Marginal posterior densities for the baseline parameters and the parameter σ controlling the observation error variance of the aphid model. Black: Bayesian imputation. Red: LNA.

We also compared the predictive distributions obtained under each inferential model. The within sample predictive distribution for the observation process $\{Y_t, t = 0, \ldots, 4.57\}$ can be obtained by integrating over the posterior uncertainty of the latent process and parameter values in the observation model (5.4). Specifically, given samples $\{(n_t^{ijk(l)}, \sigma^{(l)}), l = 1, \ldots, L\}$ from the marginal posterior $\pi(n_t^{ijk}, \sigma|y)$, the predictive density at time t can be estimated by

$$\frac{1}{L}\sum_{l=1}^{L} N\left\{y_t\,;\, n_t^{ijk(l)}, \left(\sigma^{(l)}\right)^2 n_t^{ijk(l)}\right\}.$$

Likewise, for a new experiment repeated under the same conditions, the out-of-sample predictive distribution for the aphid population size can be determined for each treat-

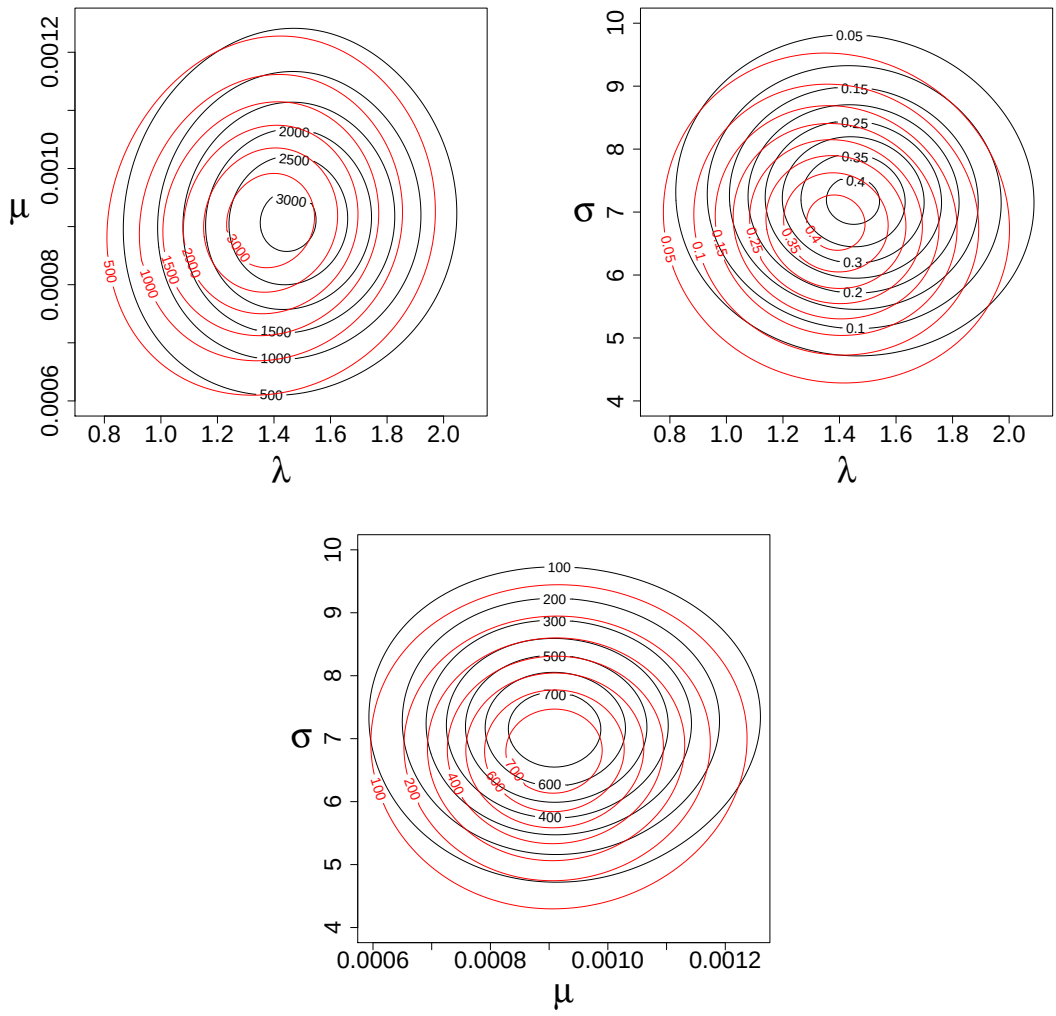

Figure 5.11: Bivariate marginal posterior densities for the baseline parameters and the parameter σ controlling the observation error variance of the aphid model. Black: Bayesian imputation. Red: LNA.

ment combination. This is estimated by averaging realisations of N_t (obtained by applying the Euler-Maruyama approximation to (5.2)) over draws from the marginal posterior $\pi(n_0^{ijk}, b^{ijk}|y)$ obtained using either Bayesian imputation or the LNA. Figures 5.13, 5.14 and 5.15 summarise these predictive distributions for a random selection of treatment combinations. Both the SDMEM and LNA give a satisfactory fit to the observed data, with all observations within or close to the central 50% of the distribution, and no observation outside the equi-tailed 95% credible intervals. As expected, the SDMEM gives a better fit over the LNA, although there is little difference between the two. There are however noticeable differences in the out-of-sample predictives, especially in the lower credible bound (in Figure 5.15) suggesting that in some situations, using the inferences made under the

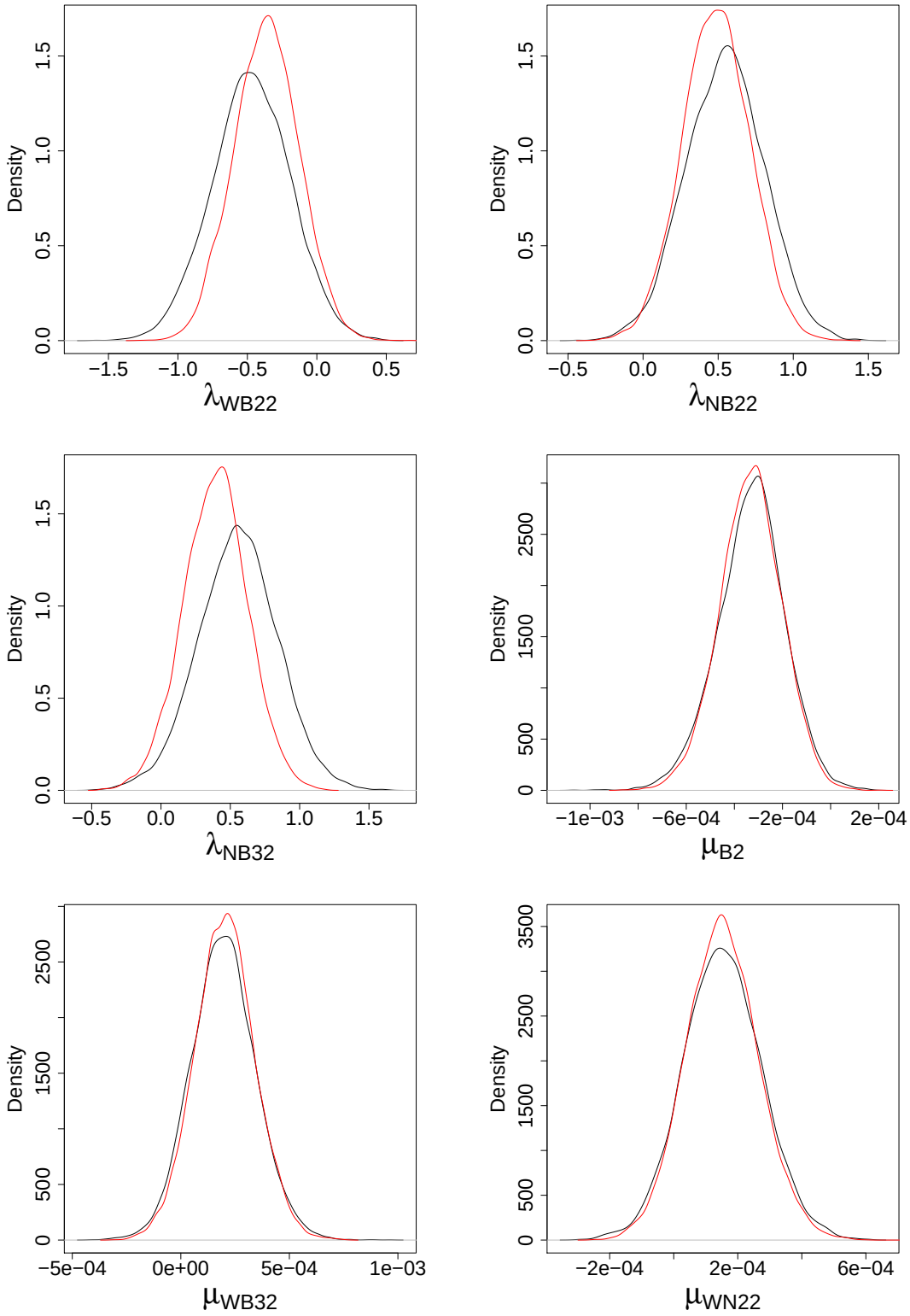

Figure 5.12: Marginal posterior densities for a selection of the aphid model parameters. Black: Bayesian imputation. Red: LNA.

LNA to predict the outcome of future experiments can give misleading results. These differences lead us to examine the marginal posterior densities of the treatment-block specific birth and death rates, λ^{ijk} and μ^{ijk}, over whose uncertainty we average. Samples from these posteriors are straightforward to obtain, using the posterior samples of the constituent parameters in (5.3). Figure 5.16 shows marginal posterior densities of the overall birth rates (λ^{ijk}) associated with the six treatment-block combinations for which predictives were presented in Figure 5.15. We see distinct differences between posteriors obtained under the Bayesian imputation approach and the LNA approach. The posteriors displayed are indicative of those obtained for all treatment combinations. Moreover, similar patterns are evident in the overall death rates (μ^{ijk}).

We obtained a minESS of 1038.6 under the modified innovation scheme. The LNA, however, clearly benefits from analytically integrating out the latent process and gave a minESS of 8907.5. For this example, we found that significant gains in computational efficiency were possible by performing the parameter updates and, for the modified innovation scheme, the path updates, in parallel. For example, updating λ_{B_2} and μ_{B_2} involves calculating a product of likelihoods (or marginal likelihoods for the LNA) over all $3^2 = 9$ treatment combinations that include block 2. These constituent likelihoods can be calculated in parallel. Similarly, for the modified innovation scheme, the treatment specific path updates can be performed in parallel. Both the modified innovation scheme and the LNA–based scheme were again coded in C and run on a high performance computing cluster with 14 cores (made up of Intel Xeon 3.0GHz processors). The modified innovation scheme took approximately 18 days to run whereas the LNA–based scheme required only approximately 4.3 days. Note that here the speed advantage of the LNA–based scheme has reduced, now being roughly 4 times faster than the modified innovation scheme, whereas in Section 5.1, the LNA was approximately 20 times faster. The intractability of the ODEs driving the LNA clearly plays a significant role in computational efficiency. In terms of overall efficiency (as measured by minESS/sec) the LNA–based scheme outperforms the Bayesian imputation approach by a factor of around 36. These computational advantages of the LNA must be weighted against the inaccuracies of the resulting posterior and predictive distributions, inaccuracies which can at times be substantial, as will be demonstrated in the following simulation study.

5.2.4 Simulation study

Following on from the analysis of the real data, we now look further to investigate differences between the Bayesian imputation approach and an inference scheme based on the LNA by considering synthetic data generated from (5.2). For simplicity, we consider a fixed treatment (low water, blanket nitrogen) and three blocks. We therefore

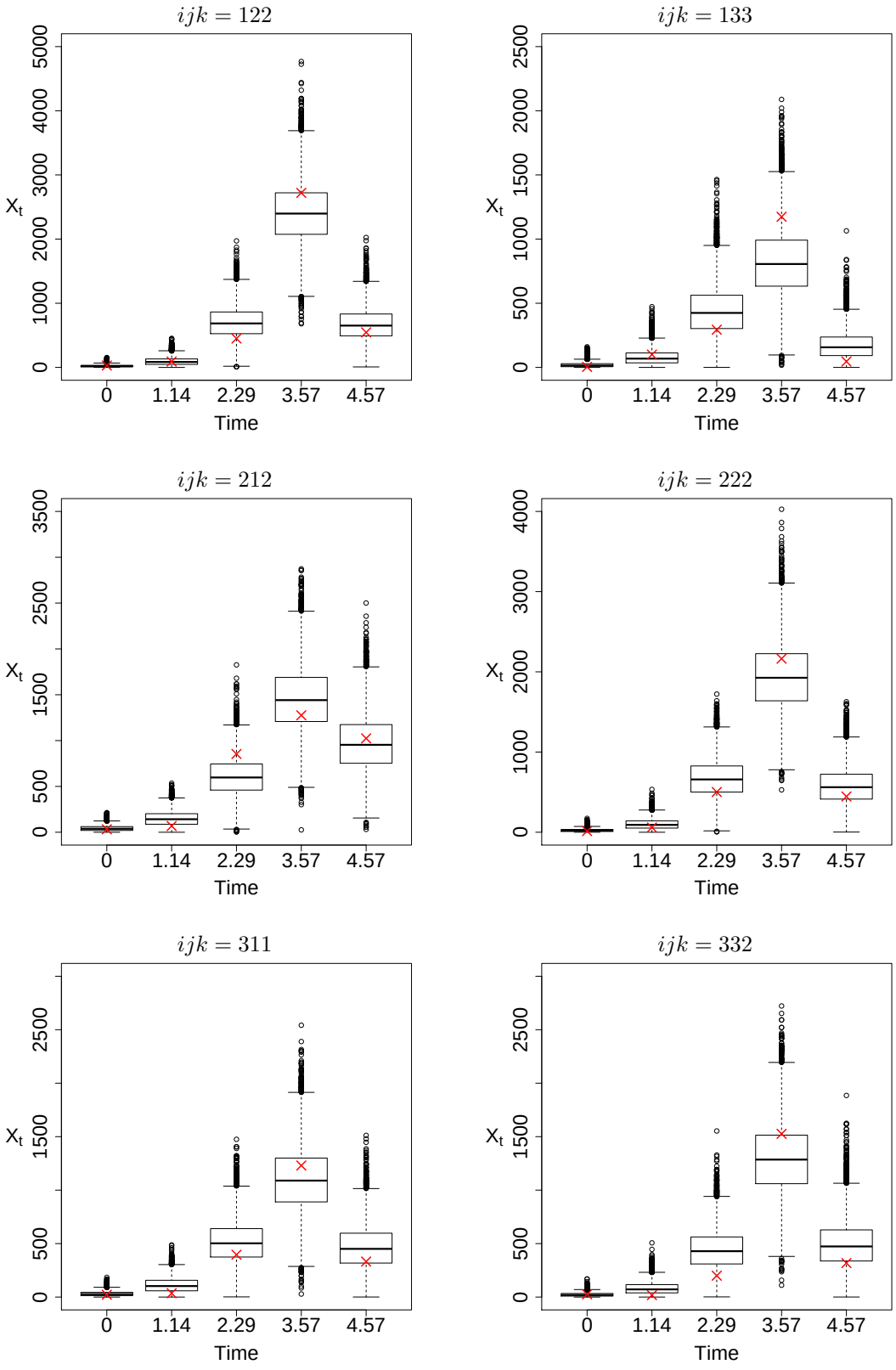

Figure 5.13: Within sample predictive distributions for the Bayesian imputation approach. The red crosses indicate the observed values.

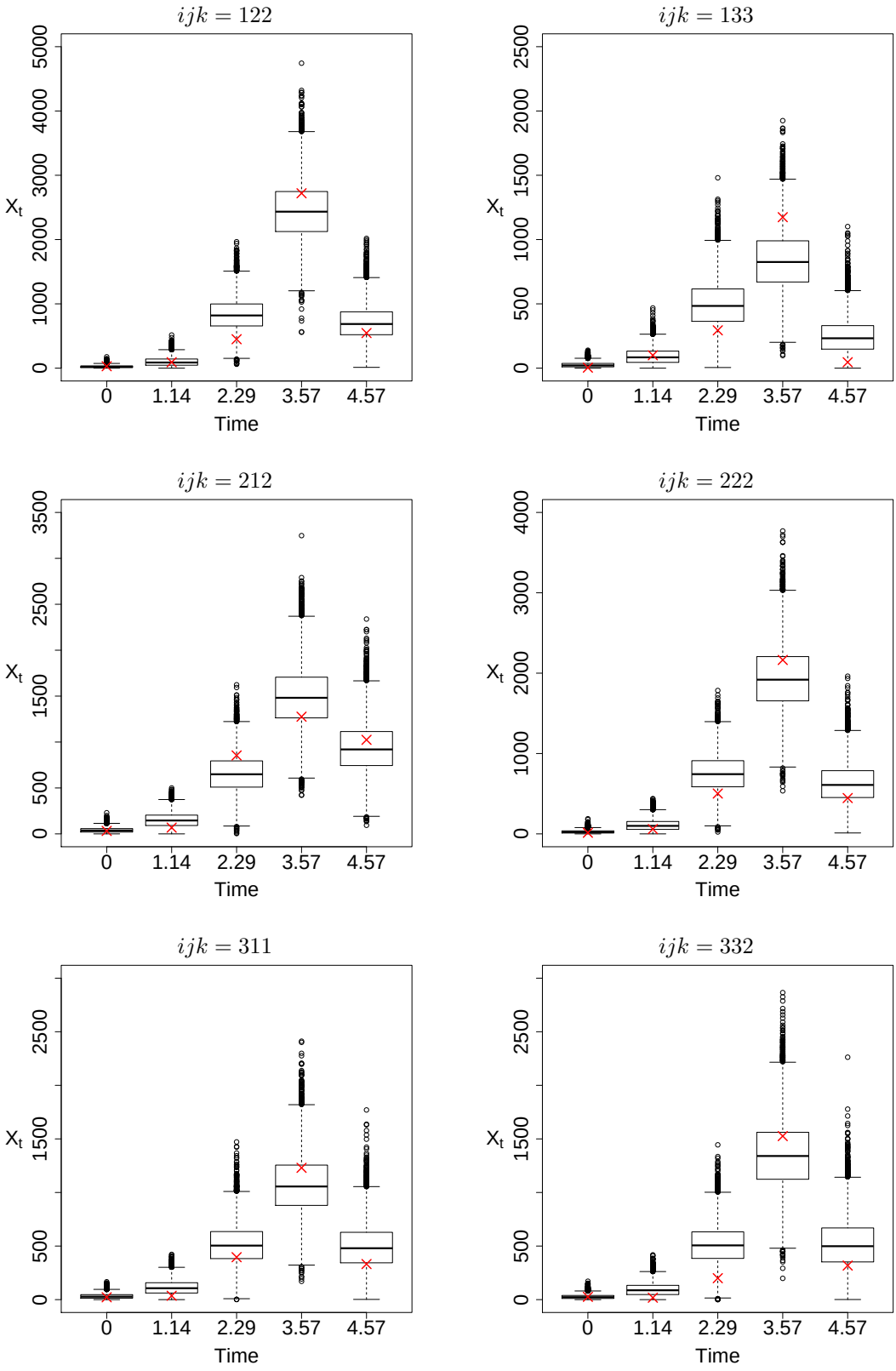

Figure 5.14: Within sample predictive distributions for the LNA. The red crosses indicate the observed values.

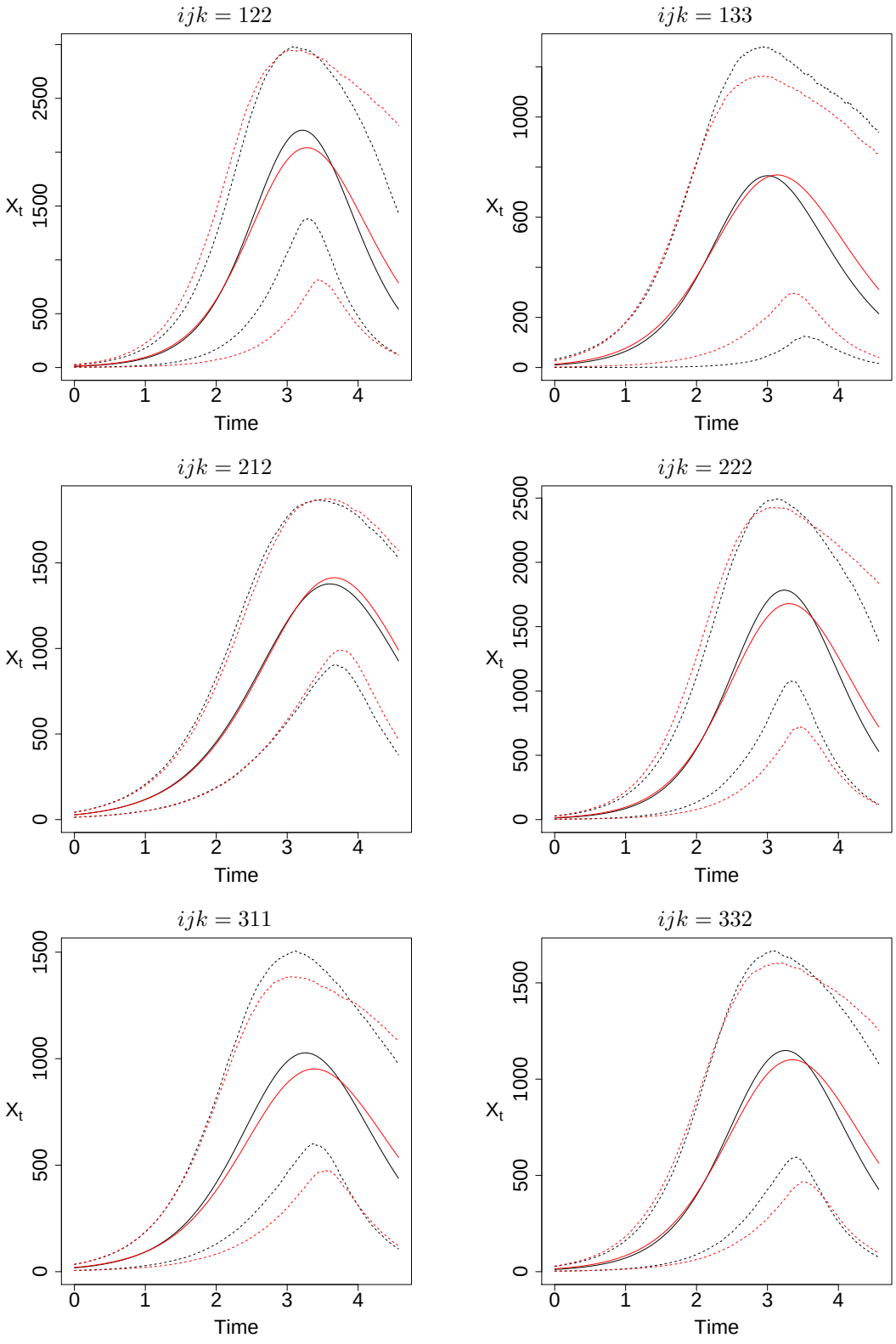

Figure 5.15: Out-of-sample predictive intervals for the aphid population size (N_t^{ijk}) against time for a random selection of treatment combinations. The mean is depicted by the solid line with the dashed representing a 95% credible region. Black: Bayesian imputation. Red: LNA.

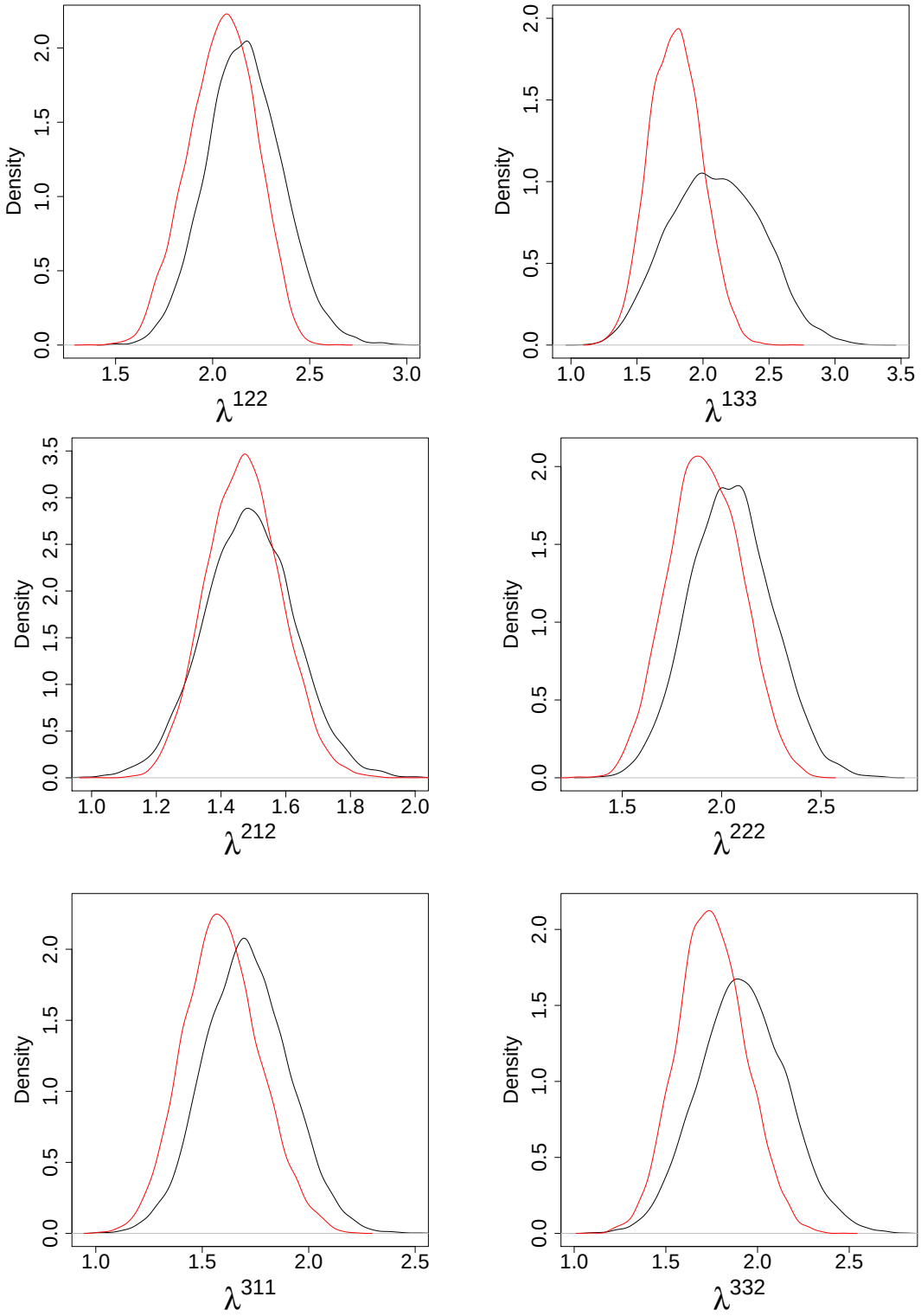

Figure 5.16: Marginal posterior densities for a random selection of the birth rates associated with specific treatment combinations in the aphid model. Black: Bayesian imputation. Red: LNA.

write $X_t^{11k} = (N_t^{11k}, C_t^{11k})'$ and consider the SDMEM

$$dX_t^{11k} = \alpha\left(X_t^{11k}, b^{11k}\right)\, dt + \sqrt{\beta\left(X_t^{11k}, b^{11k}\right)}\, dW_t^{11k}, \quad X_0^{11k} = x_0^{11k}, \quad k \in \{1,2,3\},$$

where

$$\alpha\left(X_t^{11k}, b^{11k}\right) = \begin{pmatrix} \lambda^{11k} N_t^{11k} - \mu^{11k} N_t^{11k} C_t^{11k} \\ \lambda^{11k} N_t^{11k} \end{pmatrix},$$
$$\beta\left(X_t^{11k}, b^{11k}\right) = \begin{pmatrix} \lambda^{11k} N_t^{11k} + \mu^{11k} N_t^{11k} C_t^{11k} & \lambda^{11k} N_t^{11k} \\ \lambda^{11k} N_t^{11k} & \lambda^{11k} N_t^{11k} \end{pmatrix}.$$

The fixed effects $b^{11k} = (\lambda^{11k}, \mu^{11k})'$ have a standard structure to incorporate block effects, with

$$\lambda^{11k} = \lambda + \lambda_{B_k} \quad \text{and} \quad \mu^{11k} = \mu + \mu_{B_k},$$

where we again impose the corner constraints $\lambda_{B_1} = \mu_{B_1} = 0$ for identifiability purposes.

To mimic the real dataset, we took $\lambda = 1.75$, $\mu = 0.00095$, $\lambda_{B2} = -0.1154$, $\lambda_{B3} = -0.0225$, $\mu_{B2} = -0.0004$ and $\mu_{B3} = 0.0002$. For each block, we generated five observations (on a regular grid) by using the Euler-Maruyama approximation with a small time-step ($\Delta t = 0.001$) and an initial condition of $x_0 = (5,5)'$. This gave observations at times $t = 0, 1, 2, 3$ and 4 weeks. To assess the impact of measurement error on the quality of inferences that can be made about each parameter, we corrupted our data via the following observation model:

$$Y_t^{11k} | N_t^{11k}, \sigma \overset{indep}{\sim} N\left(N_t^{11k}, \sigma^2 N_t^{11k}\right), \quad t = 0, 1, 2, 3, 4.$$

We then took $\sigma \in \{0, 0.5, 1, 5\}$ to give four synthetic datasets. We adopt the same prior specification for the unknown parameters as used in the real data application.

Both the modified innovation scheme (again incorporating the improved bridge construct) and the LNA-based inference scheme were run long enough to yield a sample of 10K almost un-autocorrelated posterior draws. For the former, we fixed the discretisation level by taking $m = 20$ and note that $m > 20$ gave little difference in posterior output. Figures 5.17, 5.18 and 5.19 show the marginal posterior densities of the baseline parameters (λ and μ) and the measurement error variance (σ). The joint posterior densities of $(\mu, \lambda)'$ are shown in Figure 5.20. It is evident that when fitting the SDMEM using the Bayesian imputation approach, the posterior samples obtained are consistent with the ground truth. This is true to a lesser extent when using the LNA, with the ground truth found in the tail of the posterior distribution in three out of the four scenarios. In fact, when using synthetic data with $\sigma < 5$, we see substantive differences in posterior output. As was

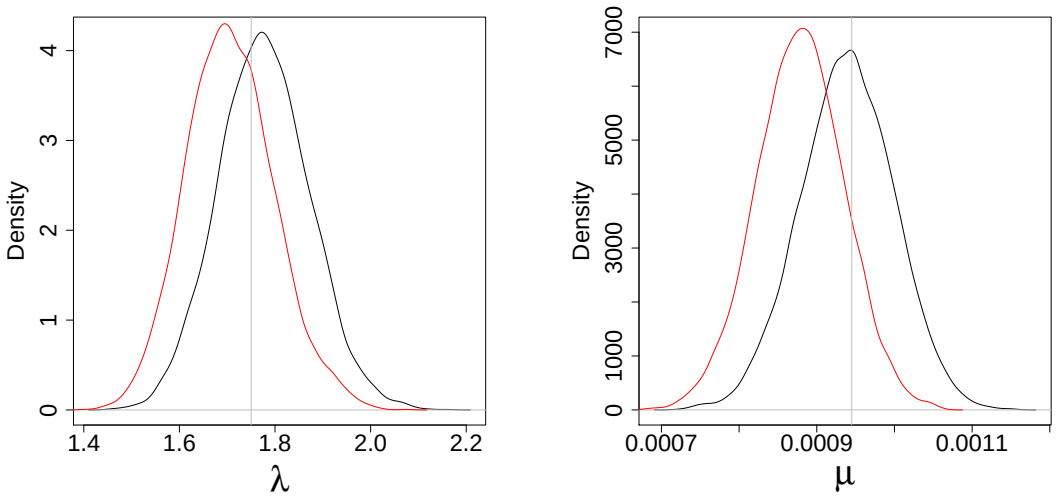

Figure 5.17: Marginal posterior densities for the baseline parameters in the aphid simulation study for the case of no measurement error ($\sigma = 0$). Black: Bayesian imputation. Red: LNA. The grey lines indicate the ground truth.

observed when using real data, the LNA underestimates parameter values compared to those obtained under the Bayesian imputation scheme. In this case, the LNA provides a relatively poor approximation to the true posterior distribution.

Increasing σ to 5 (and beyond) gives output from both schemes which is largely in agreement. This is intuitively reasonable, since, as the variance of the measurement process is increased, the ability of both inference schemes to accurately infer the underlying dynamics is diminished. Essentially, the relative difference between the LNA and SDE is reduced.

5.3 Summary

By considering two applications we have presented a systematic comparison between the Bayesian imputation approach (coupled with the modified innovation scheme) and the LNA-based inference scheme. We note that for the two examples considered within this chapter, we found little difference in the posterior output under the imputation approach for $m \geq 20$.

The ODE system governing the LNA in the orange tree growth example could be solved analytically, resulting in increases to both computational and overall efficiency (as measured by minimum ESS per second) of around a factor of 20. Moreover, in this first application, we observed little difference in the accuracy of inferences under both approaches.

Our second application concerned the diffusion approximation of a Markov jump process

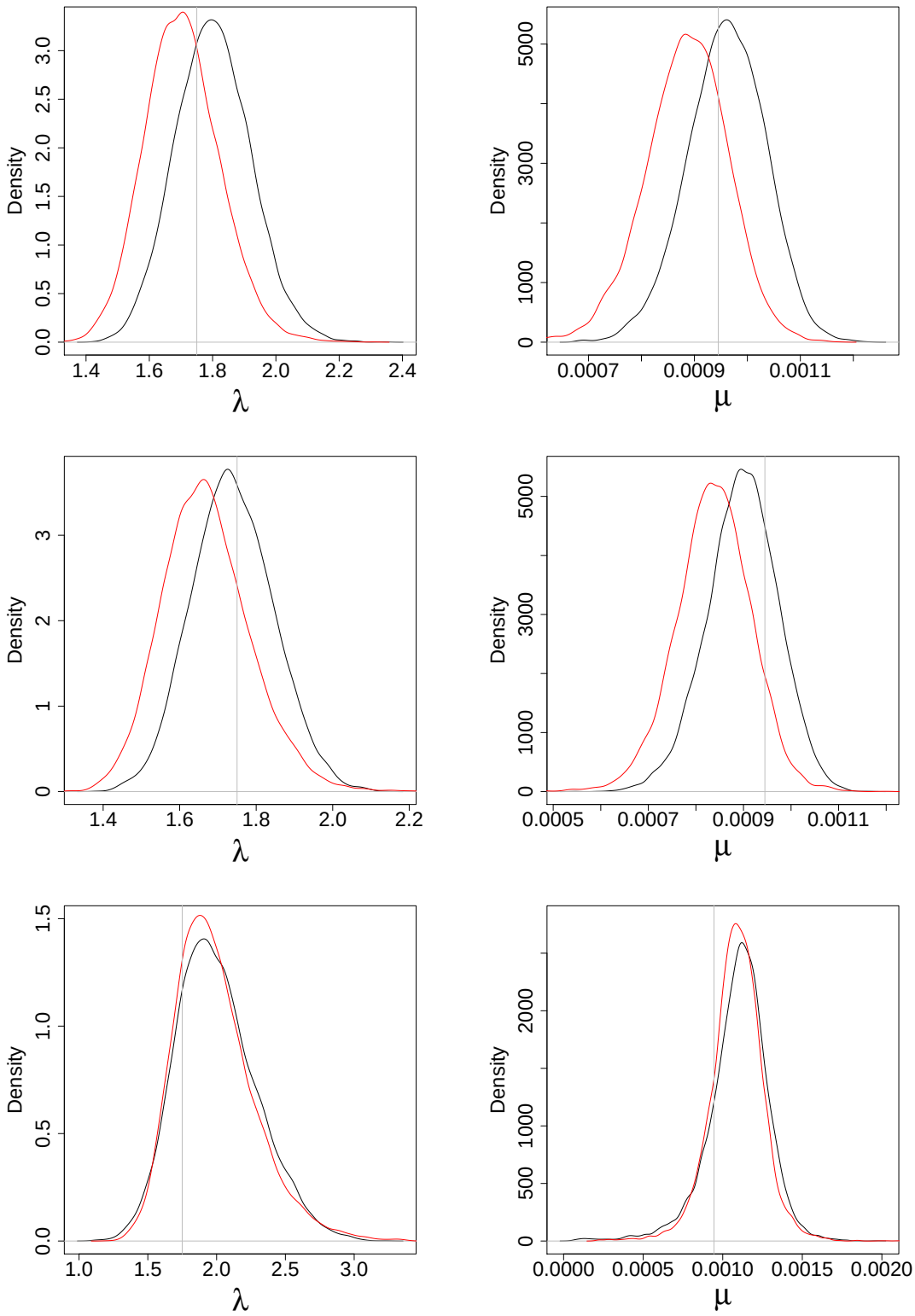

Figure 5.18: Marginal posterior densities for the baseline parameters in the aphid simulation study. $\sigma = 0.5$ (1^{st} row), $\sigma = 1$ (2^{nd} row), $\sigma = 5$ (3^{rd} row). Black: Bayesian imputation. Red: LNA. The grey lines indicate the ground truth.

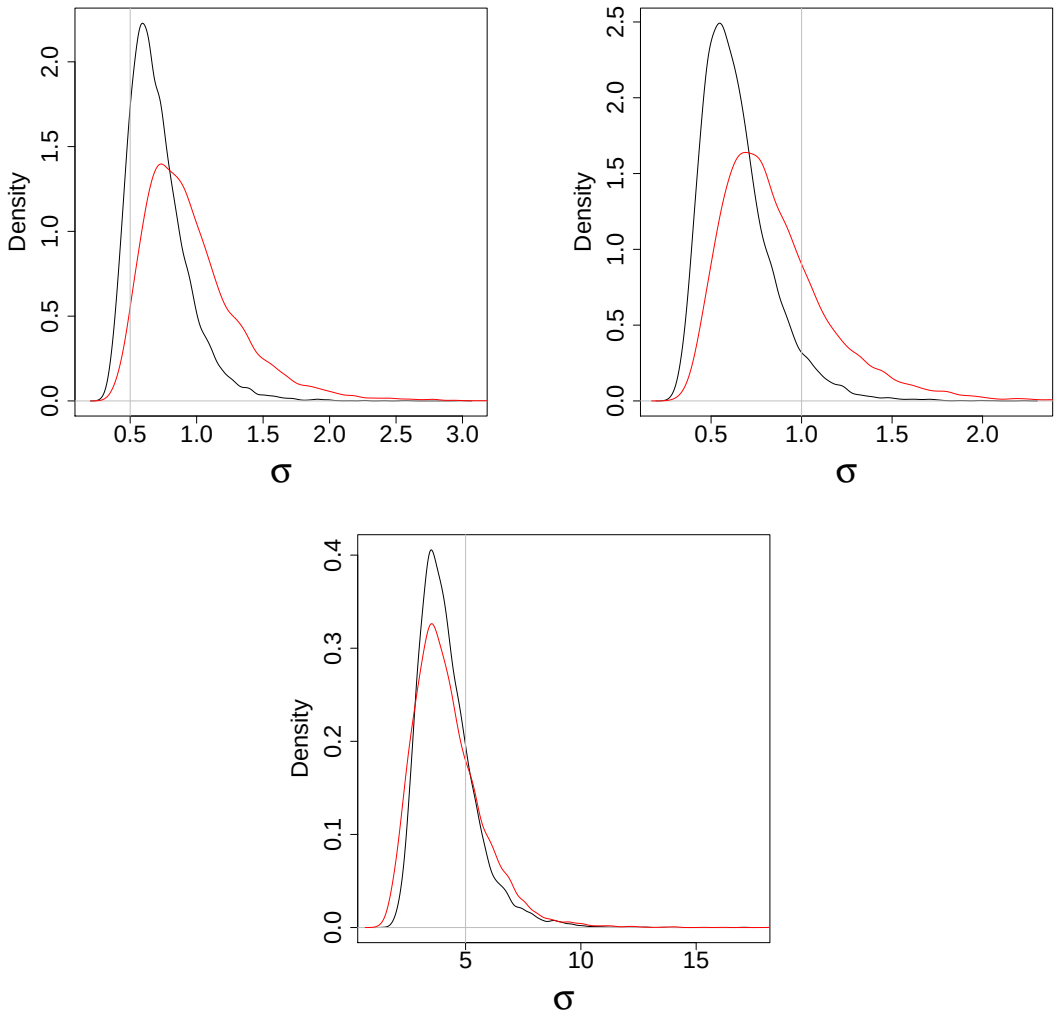

Figure 5.19: Marginal posterior densities for the parameter σ controlling the observation error variance in the aphid simulation study. Black: Bayesian imputation. Red: LNA. The grey lines indicate the ground truth.

description of aphid dynamics. The resulting SDMEM was particularly challenging to fit due to both the drift and diffusion function being nonlinear, and because only one component of the model could be observed (with error). We considered data from Matis et al. (2008) as well as synthetic data generated under four data-poor scenarios. Here, the ODE system was intractable, leading to a reduction in the computational advantage of the LNA over the imputation approach to around a factor of 4. However, there was a clear benefit to analytically integrating over the latent process using the LNA, as we see an overall increase in efficiency of around a factor of 36. Prudence must be advised though, as whilst both the imputation and LNA approaches provided a reasonable fit to the data of Matis et al. (2008), differences were found between the parameter posteriors. These differences lead to

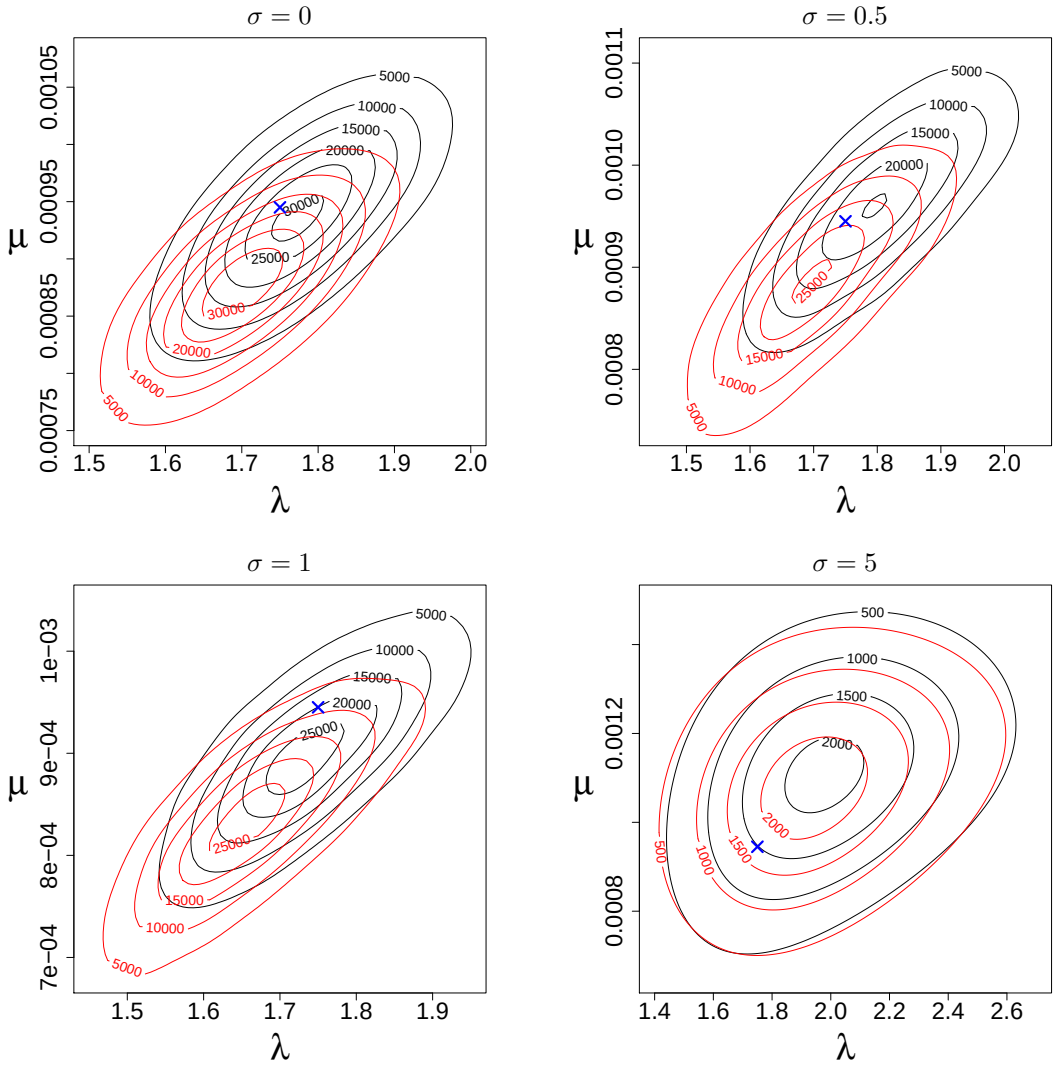

Figure 5.20: Bivariate marginal posterior densities for the baseline parameters in the aphid simulation study Black: Bayesian imputation. Red: LNA. The blue cross indicates the ground truth.

differences in the out-of-sample predictive distributions. Further differences between the LNA and Bayesian imputation approaches were highlighted through the simulation study, where the most exacerbated differences occurred when there was little or no measurement error. We therefore advocate caution when predicting the outcome of future experiments using output from an LNA-based fit.

Chapter 6

Conclusions

The intention of this thesis was to provide a framework for performing (simulation-based) Bayesian inference on a large class of multivariate SDMEMs, using discrete-time observations that may be incomplete and subject to measurement error. Furthermore, we required a method that allowed for observations to be made sparsely in time, and to work well even when the process of interest exhibits nonlinear dynamics between observation times.

Unfortunately, analytic intractability of SDEs governing most nonlinear multivariate diffusion processes precludes a closed-form expression for the transition densities. Consequently, inferring the parameters of the SDE using observations taken at discrete times is a challenging problem. We implemented a widely adopted solution, a Bayesian imputation approach, which replaces an intractable transition density with a first order Euler-Maruyama approximation, and uses data augmentation to limit the discretisation error incurred by the approximation.

The essential idea of the Bayesian imputation approach is to augment low frequency data by introducing intermediate time-points between observation times. An Euler-Maruyama scheme is then applied by approximating the transition densities over the induced discretisation as Gaussian. Computationally intensive algorithms, such as Markov chain Monte Carlo (MCMC), are then used to integrate over the uncertainty associated with the missing/unobserved data. Care must be taken in the design of such schemes due to dependence between the parameters and the latent process, and dependence between values of the latent process itself. The former was first highlighted as a problem by Roberts and Stramer (2001). To overcome this issue we implemented a reparameterisation (Roberts and Stramer, 2001; Golightly and Wilkinson, 2008, 2010) and extended a recently proposed MCMC scheme to include the SDE driven mixed-effects framework.

Key to the development of an efficient inference scheme was the ability to generate discrete-time realisations of the latent process between observation times, typically termed diffusion

bridges. We have presented a novel class of bridge constructs that are both computationally and statistically efficient, and can be readily applied in situations where only noisy and partial observation of the process is possible. Our approach is straightforward to implement and is based on a partition of the process into a deterministic part that accounts for forward dynamics, and a residual stochastic process. The intractable end-point conditioned residual SDE is then approximated using the modified diffusion bridge of Durham and Gallant (2002). Through such an approach we ensure any resulting diffusion bridges inherit the desirable properties of the modified diffusion bridge, along with the limiting (as $\Delta\tau \to 0$) forms of the Metropolis-Hastings acceptance probabilities. Using three examples, we investigated the empirical performance of two variants of the residual bridge. The first constructs the residual SDE by subtraction of a deterministic process based on the drift governing the target process. The second variant further subtracts the linear noise approximation of the expected conditioned residual process. We find that the second approach results in improved statistical efficiency at the expense of having to solve a larger ODE system consisting of order d^2 equations (as opposed to just d when using the simpler variant). For a known initial time-point x_0, the ODE system need only be solved once, irrespective of the number of skeleton bridges required. However, for unknown x_0, the ODE solution will be required for each skeleton bridge, and the difference in computational cost between the two approaches is likely to be important, especially as the dimension of the state space increases.

In Chapter 3 we compared the performance of our approach to several existing bridge constructs (adapting where necessary to the case of noisy and partial observation). These included the modified diffusion bridge (Durham and Gallant, 2002), Lindström bridge (Lindström, 2012) and guided proposal (Schauer et al., 2016). Our implementation of the latter used the LNA to guide the proposal. We find that a further modification that replaces the Euler-Maruyama variance with the modified diffusion bridge variance gave a particularly effective bridge, outperforming all others considered in terms of statistical efficiency. However, the guided proposal requires that the ODEs governing the LNA be reintegrated for each intermediate time-point and for each skeleton bridge required. Unless the ODE system can be solved analytically, we find that when combining statistical and computational efficiency, the guided proposal is outperformed by the residual bridge (that uses the further LNA subtraction).

By adopting a Bayesian imputation approach, we have shown how the modified innovation scheme of Golightly and Wilkinson (2008), which is necessary for overcoming the problematic dependence between the latent process and any parameters that enter the diffusion coefficient, can be applied to SDMEMs. The associated techniques introduce $m-1$ intermediate time points between observations. Chapter 4 gives details of the steps and methods required to construct a (simulation-based) Bayesian inference scheme using data

augmentation. The computational cost of the Bayesian imputation scheme is dictated by the number of imputed points (characterised by m) between observation times. In the examples considered in this thesis we see little difference in posterior output under the Bayesian imputation scheme for $m \geq 20$.

We also considered a tractable approximation to the SDMEM, the linear noise approximation, and provided a systematic comparison in Chapter 5 using two applications. The computational efficiency of the LNA depends on the dimension of the SDE driving the SDMEM. For a d-dimensional SDE system, the LNA requires the solution of a system of order d^2 coupled ODEs. In our first application, the resulting ODE system can be solved analytically, leading to increases in both computational and overall efficiency (as measured by minimum ESS per second) of around a factor of 20. Moreover, we found little difference in the accuracy of inferences made under the LNA and Bayesian imputation approaches. In our second application, we fitted the diffusion approximation of a Markov jump process description of aphid dynamics using data from Matis et al. (2008). In this example, the ODE system governing the LNA was intractable and the computational advantage of using the LNA over an imputation approach reduced to around a factor of 4. However, the benefit of using the LNA to analytically integrate over the latent process was clear, giving an overall increase in efficiency of around a factor of 36. It is important to note that whilst the LNA is preferred in terms of overall efficiency for the examples considered here, as the dimension d of the SDE is increased, the LNA is likely to become infeasible. Moreover, whilst both the imputation and LNA approaches provided a reasonable fit to the aphid data, differences were found between the parameter posteriors, leading to differences in the out-of-sample predictive distributions. A simulation study highlighted further differences between the LNA and Bayesian imputation approaches. Care must therefore be taken in predicting the outcome of future experiments when using output from an LNA–based fit.

Our work can be extended in a number of ways. For example, it may be possible to improve the statistical performance of the residual bridges by replacing the Euler-Maruyama approximation of the variance of $Y_T|X_0$ with that obtained under the LNA. This approach could also be combined with the Lindström sampler to avoid specification of a tuning parameter. Deriving the limiting forms of the Metropolis-Hastings acceptance rates associated with the residual bridges would be problematic due to the time-dependent terms entering the variance of the constructs. Nevertheless, this merits further research.

Applying the recently proposed particle marginal Metropolis-Hastings (PMMH) scheme (Andrieu et al., 2010) to SDMEMs is also of interest. This scheme can be used to target the marginal posterior density of the unit-specific parameters. Essentially a sequential Monte Carlo scheme (also known as a particle filter) is used to unbiasedly estimate the intractable marginal likelihood in each step of a Metropolis-Hastings sampler. We anticipate that an

efficient implementation can be obtained by using the bridge constructs outlined in this thesis inside an auxiliary particle filter (Pitt et al., 2010). Finally, we note that the tractability of the LNA can be further exploited to reduce the computational cost of the PMMH scheme; see recent work on delayed acceptance PMMH (Golightly et al., 2015).

Appendix A

Miscellaneous

A.1 Runge-Kutta-Fehlberg method

The Runge-Kutta-Fehlberg method, or indeed Fehlberg method, is an embedded method from the Runge-Kutta family (a collection of algorithms used to numerically solve a system of ODEs). It was presented in Fehlberg (1969) and is often referred to as the RKF45 method, due to the fact that it is a method of $O(h^4)$ with an error estimator of $O(h^5)$.

Briefly we explain the Runge-Kutta method. For an unknown function of time y_t (which could be a scalar or vector), with $dy/dt = f(y_t, t)$ (where f is a known function) and initial condition $y_{t_0} = y_0$. Take a step-size $h > 0$, then for $i = 0, 1, \ldots$ we have

$$y_{t_i+h} = y_{t_i} + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k4),$$

where

$$\begin{aligned}
k_1 &= f\left(y_{t_i}, t_i\right), \\
k_2 &= f\left(y_{t_i} + \frac{h}{2}k_1, t_i + \frac{h}{2}\right), \\
k_3 &= f\left(y_{t_i} + \frac{h}{2}k_2, t_i + \frac{h}{2}\right), \\
k_4 &= f\left(y_{t_i} + hk_3, t_i + h\right).
\end{aligned}$$

Here y_{t_i+h} is determined by the present value y_{t_i} and a weighted average of four increments (commonly called the RK4 approximation).

The RK4 method can be generalised to a family of explicit Runge-Kutta methods where

$$y_{t_i+h} = y_{t_i} + h\sum_{j=1}^{n} b_j k_j.$$

The forms of k_j are given by

$$\begin{aligned}
k_1 &= f\left(y_{t_i}, t_i\right), \\
k_2 &= f\left(y_{t_i} + h(a_{21}k_1), t_i + c_2 h\right), \\
k_3 &= f\left(y_{t_i} + h(a_{31}k_1 + a_{32}k_2), t_i + c_3 h\right), \\
&\vdots \\
k_n &= f\left(y_{t_i} + h(a_{n1}k_1 + a_{n2}k_2 + \cdots + a_{n,n-1}k_{n-1}), t_i + c_n h\right).
\end{aligned}$$

The a_{jl} are known as the Runge-Kutta matrix, whilst the b_j and c_j are the weights and nodes respectively. These data are usually arranged in a Butcher tableau which takes the form

$$\begin{array}{c|cccc|c}
0 & & & & & \\
c_2 & a_{21} & & & & \\
c_3 & a_{31} & a_{32} & & & \\
\vdots & \vdots & & \ddots & & \\
c_n & a_{n1} & a_{n2} & \cdots & a_{n,n-1} & \\
\hline
 & b_1 & b_2 & \cdots & & b_n
\end{array}$$

The RKF45 method is an adaptive Runge-Kutta method, which are designed to produce an estimate of the local truncation error in each step. This estimate of the error is used to control the step-size h. An adaptive Runge-Kutta method involves the use of two methods, one of order s and one of order $s-1$. Within the lower-order method

$$y^*_{t_i+h} = y_{t_i} + h\sum_{j=1}^{n} b^*_j k_j,$$

where the k_j are as given above. The error is given by

$$e_{t_i+h} = y_{t_i+h} - y^*_{t_i+h} = h\sum_{j=1}^{n}\left(b_j - b^*_j\right)k_j.$$

The error is of $O(s)$. The Butcher tableau is extended to now include the values of b^*_j, and is given by

$$
\begin{array}{c|cccccc}
0 & & & & & & \\
c_2 & a_{21} & & & & & \\
c_3 & a_{31} & a_{32} & & & & \\
\vdots & \vdots & & \ddots & & & \\
c_n & a_{n1} & a_{n2} & \cdots & a_{n,n-1} & & \\
\hline
 & b_1 & b_2 & \cdots & & b_n & \\
 & b_1^* & b_2^* & \cdots & & b_n^* &
\end{array}
$$

The extended Butcher tableau associated with the RKF45 method is

$$
\begin{array}{c|cccccc}
0 & & & & & & \\
\frac{1}{4} & \frac{1}{4} & & & & & \\
\frac{3}{8} & \frac{3}{32} & \frac{9}{32} & & & & \\
\frac{12}{13} & \frac{1932}{2197} & \frac{-7200}{2197} & \frac{7296}{2197} & & & \\
1 & \frac{439}{216} & -8 & \frac{3680}{513} & \frac{-845}{4104} & & \\
\frac{1}{2} & \frac{-8}{27} & 2 & \frac{-3544}{2565} & \frac{1859}{4104} & \frac{-11}{40} & \\
\hline
 & \frac{16}{135} & 0 & \frac{6656}{12825} & \frac{28561}{56430} & \frac{-9}{50} & \frac{2}{55} \\
 & \frac{25}{216} & 0 & \frac{1408}{2565} & \frac{2197}{4104} & \frac{-1}{5} & 0
\end{array}
$$

A.2 Semi-conjugate updates for the parameters in the orange tree growth example

Here we give details of the semi-conjugate updates used for the random effects hyper-parameters $(\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2})$ in the orange tree growth example of Section 5.1. Recall that

$$\pi(\theta, \psi, b|x) \propto \pi(\phi_1)\pi(\phi_2)\pi(\sigma_{\phi_1})\pi(\sigma_{\phi_2})\pi(\sigma) \prod_{i=1}^{N} \pi\left(\phi_1^i|\phi_1, \sigma_{\phi_1}\right) \prod_{i=1}^{N} \pi\left(\phi_2^i|\phi_2, \sigma_{\phi_2}\right) \pi(x|\theta, b),$$

and that $\phi_1^i \sim N(\phi_1, \sigma_{\phi_1}^2)$ and $\phi_2^i \sim N(\phi_2, \sigma_{\phi_2}^2)$ independently.

A.2.1 Updating $\boldsymbol{\sigma_{\phi_1}}$ and $\boldsymbol{\sigma_{\phi_2}}$

We begin by considering the update for σ_{ϕ_1}, which we update by considering an update for $\varsigma_{\phi_1} = 1/\sigma_{\phi_1}^2$, whose prior distribution is assumed to be $\varsigma_{\phi_1} \sim Ga(g, h)$. Thus, the full

conditional for ς_{ϕ_1} is given by

$$\begin{aligned}
\pi\left(\varsigma_{\phi_1}|\phi_1^1,\ldots,\phi_1^N,\phi_1\right) &\propto \pi(\varsigma_{\phi_1})\prod_{i=1}^{100}\pi\left(\phi_1^i|\phi_1,\varsigma_{\phi_1}\right) \\
&\propto \varsigma_{\phi_1}^{g-1}\exp\left(-h\varsigma_{\phi_1}\right)\prod_{i=1}^{100}\left[\varsigma_{\phi_1}^{1/2}\exp\left\{-\frac{\varsigma_{\phi_1}}{2}\left(\phi_1^i-\phi_1\right)^2\right\}\right] \\
&\propto \varsigma_{\phi_1}^{g-1}\exp\left(-h\varsigma_{\phi_1}\right)\varsigma_{\phi_1}^{50}\exp\left\{-\frac{\varsigma_{\phi_1}}{2}\sum_{i=1}^{100}\left(\phi_1^i-\phi_1\right)^2\right\} \\
&\propto \varsigma_{\phi_1}^{g+49}\exp\left[-\varsigma_{\phi_1}\left\{h+\frac{1}{2}\sum_{i=1}^{100}\left(\phi_1^i-\phi_1\right)^2\right\}\right].
\end{aligned}$$

Therefore

$$\varsigma_{\phi_1}|\phi_1^1,\ldots,\phi_1^N,\phi_1 \sim Ga\left\{g+50, h+\frac{1}{2}\sum_{i=1}^{100}\left(\phi_1^i-\phi_1\right)^2\right\}.$$

Hence, if ς_{ϕ_1} has a $Ga(1,0.01)$ prior, we have that

$$\varsigma_{\phi_1}|\phi_1^1,\ldots,\phi_1^N,\phi_1 \sim Ga\left\{51, 0.01+\frac{1}{2}\sum_{i=1}^{100}\left(\phi_1^i-\phi_1\right)^2\right\}.$$

Similarly, if ς_{ϕ_2} has a $Ga(1,0.01)$ prior, then

$$\varsigma_{\phi_2}|\phi_2^1,\ldots,\phi_2^N,\phi_2 \sim Ga\left\{51, 0.01+\frac{1}{2}\sum_{i=1}^{100}\left(\phi_2^i-\phi_2\right)^2\right\}.$$

Realisations from the above full conditionals can easily be transformed to give draws of σ_{ϕ_1} and σ_{ϕ_2}.

A.2.2 Updating ϕ_1 and ϕ_2

We now turn our attention to updating ϕ_1. Let us assume the prior distribution of $\phi_1 \sim N(b,1/d)$. Therefore, the full conditional distribution for ϕ_1 is

$$\begin{aligned}
\pi\left(\phi_1|\phi_1^1,\ldots,\phi_1^N,\varsigma_{\phi_1}\right) &\propto \pi(\phi_1)\prod_{i=1}^{100}\pi\left(\phi_1^i|\phi_1,\varsigma_{\phi_1}\right) \\
&\propto \exp\left\{-\frac{d}{2}(\phi_1-b)^2\right\}\prod_{i=1}^{100}\exp\left\{-\frac{\varsigma_{\phi_1}}{2}\left(\phi_1^i-\phi_1\right)^2\right\} \\
&\propto \exp\left\{-\frac{d}{2}(\phi_1-b)^2-\frac{\varsigma_{\phi_1}}{2}\sum_{i=1}^{100}\left(\phi_1^i-\phi_1\right)^2\right\}
\end{aligned}$$

$$\propto \exp\left\{-\frac{d}{2}\left(\phi_1^2 - 2b\phi_1 + b^2\right) - \frac{\varsigma_{\phi_1}}{2}\left(\sum_{i=1}^{100}\phi_1^{i^2} - 2\phi_1\sum_{i=1}^{100}\phi_1^i + 100\phi_1^2\right)\right\}$$
$$\propto \exp\left\{-\frac{d}{2}\left(\phi_1^2 - 2b\phi_1\right) - \frac{\varsigma_{\phi_1}}{2}\left(100\phi_1^2 - 200\bar{\phi_1}\phi_1\right)\right\}$$

where $\bar{\phi_1} = \frac{1}{100}\sum_{i=1}^{100}\phi_1^i$ is the arithmetic mean of the ϕ_1^i, $i = 1, \ldots, 100$. Hence

$$\pi\left(\phi_1|\phi_1^1, \ldots, \phi_1^N, \varsigma_{\phi_1}\right) \propto \exp\left\{-\frac{1}{2}\left(d\phi_1^2 - 2db\phi_1 + 100\varsigma_{\phi_1}\phi_1^2 - 200\varsigma_{\phi_1}\bar{\phi_1}\phi_1\right)\right\}$$
$$\propto \exp\left[-\frac{1}{2}\left\{\left(d + 100\varsigma_{\phi_1}\right)\phi_1^2 - \left(2db + 200\varsigma_{\phi_1}\bar{\phi_1}\right)\phi_1\right\}\right].$$

On completing the square, this gives

$$\pi\left(\phi_1|\phi_1^1, \ldots, \phi_1^N, \varsigma_{\phi_1}\right) \propto \exp\left[-\frac{d + 100\varsigma_{\phi_1}}{2}\left\{\left(\phi_1 - \frac{db + 100\varsigma_{\phi_1}\bar{\phi_1}}{d + 100\varsigma_{\phi_1}}\right)^2 - \left(\frac{db + 100\varsigma_{\phi_1}\bar{\phi_1}}{d + 100\varsigma_{\phi_1}}\right)^2\right\}\right]$$
$$\propto \exp\left[-\frac{d + 100\varsigma_{\phi_1}}{2}\left\{\left(\phi_1 - \frac{db + 100\varsigma_{\phi_1}\bar{\phi_1}}{d + 100\varsigma_{\phi_1}}\right)^2\right\}\right].$$

Whence

$$\phi_1|\phi_1^1, \ldots, \phi_1^N, \varsigma_{\phi_1} \sim N\left(\frac{db + 100\varsigma_{\phi_1}\bar{\phi_1}}{d + 100\varsigma_{\phi_1}}, \frac{1}{d + 100\varsigma_{\phi_1}}\right).$$

Therefore, if $\phi_1 \sim N(0, 100^2)$ *a priori* (as is the case in Section 5.1), we obtain

$$\phi_1|\phi_1^1, \ldots, \phi_1^N i, \varsigma_{\phi_1} \sim N\left(\frac{100\varsigma_{\phi_1}\bar{\phi_1}}{0.0001 + 100\varsigma_{\phi_1}}, \frac{1}{0.0001 + 100\varsigma_{\phi_1}}\right).$$

Similarly, taking a $N(0, 100^2)$ prior distribution for ϕ_2 gives

$$\phi_2|\phi_2^1, \ldots, \phi_2^N, \varsigma_{\phi_2} \sim N\left(\frac{100\varsigma_{\phi_2}\bar{\phi_2}}{0.0001 + 100\varsigma_{\phi_2}}, \frac{1}{0.0001 + 100\varsigma_{\phi_2}}\right).$$

A.3 Semi-conjugate update for σ in the cotton aphid dynamics example

In this section we outline the steps involved in the semi-conjugate update for σ in the cotton aphid dynamics example of Section 5.2. As above, in Section A.2.1, we update σ by considering the update for ς, where $\varsigma = 1/\sigma^2$. We assume that ς follows the prior distribution, $Ga(g, h)$. Noting that ς enters only in the prior density and the density of

the observations (y) given the latent process (x), we obtain

$$\begin{aligned}
\pi(\varsigma|y,x) &\propto \pi(\varsigma)\prod_{i=1}^{3}\prod_{j=1}^{3}\prod_{k=1}^{3}\prod_{l=0}^{4}\pi\left(y_{t_l}^{ijk}|x_{t_l}^{ijk},\varsigma\right)\\
&\propto \varsigma^{g-1}\exp\left(-h\varsigma\right)\prod_{i=1}^{3}\prod_{j=1}^{3}\prod_{k=1}^{3}\prod_{l=0}^{4}\left[\varsigma^{1/2}\exp\left\{-\frac{\varsigma}{2}\frac{\left(y_{t_l}^{ijk}-n_{t_l}^{ijk}\right)^2}{n_{t_l}^{ijk}}\right\}\right]\\
&\propto \varsigma^{g-1}\exp\left(-h\varsigma\right)\varsigma^{135/2}\exp\left\{-\frac{\varsigma}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=0}^{4}\frac{\left(y_{t_l}^{ijk}-n_{t_l}^{ijk}\right)^2}{n_{t_l}^{ijk}}\right\}\\
&\propto \varsigma^{g+133/2}\exp\left[-\varsigma\left\{h+\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=0}^{4}\frac{\left(y_{t_l}^{ijk}-n_{t_l}^{ijk}\right)^2}{n_{t_l}^{ijk}}\right\}\right].
\end{aligned}$$

Therefore

$$\varsigma|y,x \sim Ga\left\{g+\frac{135}{2}, h+\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=0}^{4}\frac{\left(y_{t_l}^{ijk}-n_{t_l}^{ijk}\right)^2}{n_{t_l}^{ijk}}\right\}.$$

Thus, if ς follows a $Ga(1,1)$ prior (as in the final analysis of Section 5.2) we have

$$\varsigma|y,x \sim Ga\left\{\frac{137}{2}, 1+\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=0}^{4}\frac{\left(y_{t_l}^{ijk}-n_{t_l}^{ijk}\right)^2}{n_{t_l}^{ijk}}\right\}.$$

Finally, transforming ς gives a sample for σ.

Appendix B

LNA solutions for certain tractable systems

Recall that under the LNA we partition X_t as $X_t = \eta_t + R_t$, and approximate the distribution of the residual stochastic process R_t as a Gaussian. This gives

$$X_t | X_0 = x_0 \sim N\left(\eta_t, P_t \psi_t P_t'\right)$$

where η_t, P_t and ψ_t satisfy the ODE system

$$\begin{aligned}
\frac{d\eta_t}{dt} &= \alpha\left(\eta_t, \theta, b\right), \qquad \eta_0 = x_0, \\
\frac{dP_t}{dt} &= H_t P_t, \quad P_0 = I_d, \\
\frac{d\psi_t}{dt} &= P_t^{-1} \beta\left(\eta_t, \theta, b\right) \left(P_t^{-1}\right)', \quad \psi_0 = 0.
\end{aligned}$$

Here H_t is the Jacobian matrix with (i,j)th element

$$(H_t)_{i,j} = \frac{\partial \alpha_i\left(\eta_t, \theta, b\right)}{\partial \eta_{j,t}}.$$

Furthermore, recall that this ODE system may be equivalently expressed as

$$\begin{aligned}
\frac{d\eta_t}{dt} &= \alpha\left(\eta_t, \theta, b\right), \qquad \eta_0 = x_0, \\
\frac{dV_t}{dt} &= H_t V_t + \beta\left(\eta_t, \theta, b\right) + V_t H_t', \quad V_0 = 0,
\end{aligned}$$

where $X_t | X_0 = x_0 \sim N(\eta_t, V_t)$.

B.1 Birth-death model

A simple birth-death process is characterised by the SDE

$$dX_t = (\theta_1 - \theta_2)\, X_t\, dt + \sqrt{(\theta_1 + \theta_2)\, X_t}\, dW_t, \quad X_0 = x_0,$$

and the ODE system ((3.22), (3.28) and (3.29)) governing the linear noise approximation of the birth-death process is given by

$$\frac{d\eta_t}{dt} = (\theta_1 - \theta_2)\, \eta_t, \qquad \eta_0 = x_0, \tag{B.1}$$

$$\frac{dP_t}{dt} = (\theta_1 - \theta_2)\, P_t, \qquad P_0 = I_d = 1, \tag{B.2}$$

$$\frac{d\psi_t}{dt} = P_t^{-1}\, (\theta_1 + \theta_2)\, \eta_t \left(P_t^{-1}\right)', \qquad \psi_0 = 0. \tag{B.3}$$

Let us first consider (B.1). Rearranging we obtain

$$\frac{d\eta_t}{\eta_t} = (\theta_1 - \theta_2)\, dt.$$

Thus, it follows that $\log \eta_t = (\theta_1 - \theta_2)\, t + A$, where A is a constant, and so

$$\eta_t = \tilde{A} \exp\{(\theta_1 - \theta_2)\, t\}.$$

Noting that at $t = 0$, $\eta_0 = x_0$ gives $\tilde{A} = x_0$. Hence

$$\eta_t = x_0 \exp\{(\theta_1 - \theta_2)\, t\}. \tag{B.4}$$

Similarly, solving (B.2) gives

$$P_t = \tilde{A} \exp\{(\theta_1 - \theta_2)\, t\}.$$

At $t = 0$, $P_0 = 1$ and consequently $\tilde{A} = 1$. Therefore

$$P_t = \exp\{(\theta_1 - \theta_2)\, t\}. \tag{B.5}$$

Substituting (B.4) and (B.5) into (B.3) we get

$$\begin{aligned}
\frac{d\psi_t}{dt} &= \exp\{-(\theta_1 - \theta_2)\, t\}\, (\theta_1 + \theta_2)\, \eta_t \exp\{-(\theta_1 - \theta_2)\, t\} \\
&= \exp\{-2\,(\theta_1 - \theta_2)\, t\}\, (\theta_1 + \theta_2)\, x_0 \exp\{(\theta_1 - \theta_2)\, t\} \\
&= (\theta_1 + \theta_2)\, x_0 \exp\{-(\theta_1 - \theta_2)\, t\}.
\end{aligned}$$

Hence

$$\psi_t = \frac{-(\theta_1+\theta_2)\,x_0}{\theta_1-\theta_2}\exp\{-(\theta_1-\theta_2)\,t\} + A,$$

where A is a constant. At $t=0$, $\psi_0 = 0$ which implies that

$$A = \frac{(\theta_1+\theta_2)\,x_0}{\theta_1-\theta_2}.$$

It then follows that

$$\psi_t = \frac{(\theta_1+\theta_2)\,x_0}{\theta_1-\theta_2}\left[1-\exp\{-(\theta_1-\theta_2)\,t\}\right]. \tag{B.6}$$

Therefore the solution to the LNA for the birth-death model is given by (B.4)–(B.6).

B.2 Orange tree growth model

Recall that each tree in the orange tree growth model is represented by the SDMEM

$$dX_t^i = \frac{1}{\phi_1^i\phi_2^i}X_t^i\left(\phi_1^i - X_t^i\right)\,dt + \sigma\sqrt{X_t^i}\,dW_t^i, \quad X_0^i = x_0^i, \quad i=1,\ldots,N.$$

For each tree i, the ODE system (4.21) and (4.22) governing the linear noise approximation for this model is

$$\frac{d\eta_t^i}{dt} = \frac{1}{\phi_1^i\phi_2^i}\eta_t^i\left(\phi_1^i-\eta_t^i\right), \qquad \eta_{t_j}^i = x_{t_j}^i, \tag{B.7}$$

$$\frac{dV_t^i}{dt} = \frac{2}{\phi_1^i\phi_2^i}\left(\phi_1^i-2\eta_t^i\right)V_t^i + \sigma^2\eta_t^i, \qquad V_{t_j}^i = 0. \tag{B.8}$$

As the solution of the ODE system is identical in form for every tree, we henceforth suppress i from the notation to aid simplicity. Let us begin by solving (B.7). Rearranging gives

$$\frac{d\eta_t}{\eta_t(\phi_1-\eta_t)} = \frac{dt}{\phi_1\phi_2} \quad\Longrightarrow\quad \int\frac{d\eta_t}{\eta_t(\phi_1-\eta_t)} = \frac{t}{\phi_1\phi_2} + A,$$

where A is a constant. Proceeding using partial fractions we obtain

$$\int\left(\frac{1}{\phi_1\eta_t} + \frac{1}{\phi_1^2-\phi_1\eta_t}\right)d\eta_t = \frac{t}{\phi_1\phi_2} + A$$

$$\Longrightarrow \qquad \frac{\log\eta_t - \log\left(\phi_1^2-\phi_1\eta_t\right)}{\phi_1} = \frac{t}{\phi_1\phi_2} + A$$

$$
\begin{aligned}
\implies \qquad \log \eta_t - \log\left(\phi_1^2 - \phi_1 \eta_t\right) &= \frac{t}{\phi_2} + \tilde{A} \\
\implies \qquad \frac{\eta_t}{\phi_1^2 - \phi_1 \eta_t} &= \hat{A} e^{t/\phi_2} \\
\implies \qquad \eta_t &= \bar{A}\phi_1 e^{t/\phi_2} - \bar{A}\eta_t e^{t/\phi_2} \\
\implies \qquad \eta_t + \bar{A}\eta_t e^{t/\phi_2} &= \bar{A}\phi_1 e^{t/\phi_2} \\
\implies \qquad \eta_t\left(1 + \bar{A} e^{t/\phi_2}\right) &= \bar{A}\phi_1 e^{t/\phi_2} \\
\implies \qquad \eta_t &= \frac{\bar{A}\phi_1 e^{t/\phi_2}}{1 + \bar{A} e^{t/\phi_2}},
\end{aligned} \tag{B.9}
$$

where $\tilde{A}$, $\hat{A}$ and $\bar{A}$ are constants. For this model, at $t = 0$, $\eta_0 = x_0$. Thus it follows that

$$
\begin{aligned}
x_0 &= \frac{\bar{A}\phi_1}{1 + \bar{A}} \\
\implies \qquad x_0 + x_0\bar{A} &= \bar{A}\phi_1 \\
\implies \qquad x_0 &= \bar{A}(\phi_1 - x_0),
\end{aligned}
$$

and so

$$
\bar{A} = \frac{x_0}{\phi_1 - x_0}. \tag{B.10}
$$

We now turn our attention to solving (B.8). We begin by substituting (B.9) into (B.8) to obtain

$$
\begin{aligned}
\frac{dV_t}{dt} &= \frac{2}{\phi_1\phi_2}\left(\phi_1 - \frac{2\bar{A}\phi_1 e^{t/\phi_2}}{1 + \bar{A}e^{t/\phi_2}}\right) V_t + \frac{\sigma^2 \bar{A}\phi_1 e^{t/\phi_2}}{1 + \bar{A}e^{t/\phi_2}} \\
\implies \qquad \frac{dV_t}{dt} &- \frac{2}{\phi_1\phi_2}\left(\phi_1 - \frac{2\bar{A}\phi_1 e^{t/\phi_2}}{1 + \bar{A}e^{t/\phi_2}}\right) V_t = \frac{\sigma^2 \bar{A}\phi_1 e^{t/\phi_2}}{1 + \bar{A}e^{t/\phi_2}}.
\end{aligned}
$$

A solution to V_t may be obtained through the use of an integrating factor (IF), where

$$
IF = \exp\left\{\int -\frac{2}{\phi_1\phi_2}\left(\phi_1 - \frac{2\bar{A}\phi_1 e^{t/\phi_2}}{1 + \bar{A}e^{t/\phi_2}}\right)\right\}.
$$

Now

$$
\begin{aligned}
\int -\frac{2}{\phi_1\phi_2}\left(\phi_1 - \frac{2\bar{A}\phi_1 e^{t/\phi_2}}{1 + \bar{A}e^{t/\phi_2}}\right) &= -\int \frac{2}{\phi_2} dt + \int \frac{4\bar{A}e^{t/\phi_2}}{\phi_2\left(1 + \bar{A}e^{t/\phi_2}\right)} dt \\
&= -\frac{2t}{\phi_2} + \frac{4\bar{A}}{\phi_2}\int \frac{e^{t/\phi_2}}{1 + \bar{A}e^{t/\phi_2}} dt.
\end{aligned}
$$

Defining $q = t/\phi_2$, gives $dq/dt = 1/\phi_2$, with $dt = \phi_2 dq$. Hence

$$\int -\frac{2}{\phi_1\phi_2}\left(\phi_1 - \frac{2\bar{A}\phi_1 e^{t/\phi_2}}{1+\bar{A}e^{t/\phi_2}}\right) = -\frac{2t}{\phi_2} + 4\bar{A}\int \frac{e^q}{1+\bar{A}e^q}dq.$$

Furthermore, setting $r = 1 + \bar{A}e^q$, gives $dr/dq = \bar{A}e^q$ and $dq = dr/\bar{A}e^q$. Thus

$$\begin{aligned}\int -\frac{2}{\phi_1\phi_2}\left(\phi_1 - \frac{2\bar{A}\phi_1 e^{t/\phi_2}}{1+\bar{A}e^{t/\phi_2}}\right) &= -\frac{2t}{\phi_2} + 4\int \frac{dr}{r} \\ &= -\frac{2t}{\phi_2} + 4\log r \\ &= -\frac{2t}{\phi_2} + 4\log\left(1+\bar{A}e^{t/\phi_2}\right).\end{aligned}$$

Therefore

$$IF = \exp\left\{-\frac{2t}{\phi_2} + 4\log\left(1+\bar{A}e^{t/\phi_2}\right)\right\} = e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4.$$

Whence

$$\begin{aligned}\frac{d}{dt}\left\{e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4 V_t\right\} &= e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4 \times \frac{\sigma^2\bar{A}\phi_1 e^{t/\phi_2}}{1+\bar{A}e^{t/\phi_2}} \\ \implies \qquad e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4 V_t &= \sigma^2\bar{A}\phi_1\int \frac{\left(1+\bar{A}e^{t/\phi_2}\right)^3}{e^{t/\phi_2}}dt.\end{aligned}$$

If $q = e^{t/\phi_2}$, then $dq/dt = e^{t/\phi_2}/\phi_2$ and $dt = \phi_2 dq/e^{t/\phi_2}$. Hence

$$e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4 V_t = \sigma^2\bar{A}\phi_1\phi_2\int \frac{\left(1+\bar{A}q\right)^3}{q^2}dq.$$

Now taking $r = 1 + \bar{A}q$, gives $dr/dq = \bar{A}$, $dq = dr/\bar{A}$. Thus

$$e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4 V_t = \sigma^2\bar{A}^2\phi_1\phi_2\int \frac{r^3}{(r-1)^2}dr.$$

Alternatively

$$\frac{r^3}{(r-1)^2} = r + 2 + \frac{3}{r} + \frac{4}{r^2} + \frac{5}{r^3} + \frac{6}{r^4} + \cdots .$$

Noting the Laurent expansions

$$\frac{1}{x-1} = \frac{1}{x} + \frac{1}{x^2} + \frac{1}{x^3} + \frac{1}{x^4} + \cdots$$

and

$$\frac{1}{(x-1)^2} = \frac{1}{x^2} + \frac{2}{x^3} + \frac{3}{x^4} + \frac{4}{x^5} + \cdots,$$

it clearly follows that

$$\frac{r^3}{(r-1)^2} = r + 2 + \frac{3}{r-1} + \frac{1}{(r-1)^2}.$$

From this we write

$$\begin{aligned} e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4 V_t &= \sigma^2\bar{A}^2\phi_1\phi_2\int\left\{r+2+\frac{3}{r-1}+\frac{1}{(r-1)^2}\right\}dr \\ &= \sigma^2\bar{A}^2\phi_1\phi_2\left\{\frac{r^2}{2}+2r+3\log(r-1)-\frac{1}{r-1}+B\right\}, \end{aligned}$$

where B is a constant, and so

$$\begin{aligned} e^{-2t/\phi_2}\left(1+\bar{A}e^{t/\phi_2}\right)^4 V_t &= \sigma^2\bar{A}^2\phi_1\phi_2\left\{\frac{\left(1+\bar{A}e^{t/\phi_2}\right)^2}{2}+2\left(1+\bar{A}e^{t/\phi_2}\right)\right. \\ &\qquad \left. +3\log\left(\bar{A}e^{t/\phi_2}\right)-\frac{1}{\bar{A}e^{t/\phi_2}}\right\}+\tilde{B} \\ &= \sigma^2\bar{A}^2\phi_1\phi_2\left(\frac{1+2\bar{A}e^{t/\phi_2}+\bar{A}^2e^{2t/\phi_2}}{2}+2\bar{A}e^{t/\phi_2}+\frac{3t}{\phi_2}-\frac{1}{\bar{A}e^{t/\phi_2}}\right)+\hat{B} \\ &= \sigma^2\bar{A}^2\phi_1\phi_2\left(\frac{1}{2}\bar{A}^2e^{2t/\phi_2}+3\bar{A}e^{t/\phi_2}-\frac{1}{\bar{A}e^{t/\phi_2}}+\frac{3t}{\phi_2}\right)+\bar{B} \\ &= \sigma^2\bar{A}\phi_1\left(\frac{1}{2}\bar{A}^3\phi_2e^{2t/\phi_2}+3\bar{A}^2\phi_2e^{t/\phi_2}-\phi_2e^{-t/\phi_2}+3\bar{A}t\right)+\bar{B}, \end{aligned}$$

where $\tilde{B}$, $\hat{B}$ and $\bar{B}$ are constants. It therefore follows that

$$V_t = \frac{\sigma^2\bar{A}\phi_1e^{2t/\phi_2}\left(\frac{1}{2}\bar{A}^3\phi_2e^{2t/\phi_2}+3\bar{A}^2\phi_2e^{t/\phi_2}-\phi_2e^{-t/\phi_2}+3\bar{A}t\right)}{\left(1+\bar{A}e^{t/\phi_2}\right)^4}+\frac{\bar{B}e^{2t/\phi_2}}{\left(1+\bar{A}e^{t/\phi_2}\right)^4}.$$

At $t=0$, $V_0=0$, and so

$$\bar{B} = -\sigma^2\bar{A}\phi_1\left(\frac{1}{2}\bar{A}^3\phi_2+3\bar{A}^2\phi_2-\phi_2\right).$$

Therefore

$$V_t = C\left(\frac{1}{2}\bar{A}^3\phi_2e^{2t/\phi_2}+3\bar{A}^2\phi_2e^{t/\phi_2}-\phi_2e^{-t/\phi_2}+3\bar{A}t-\frac{1}{2}\bar{A}^3\phi_2-3\bar{A}^2\phi_2+\phi_2\right), \quad \text{(B.11)}$$

where

$$C = \frac{\sigma^2 \bar{A} \phi_1 e^{2t/\phi_2}}{\left(1 + \bar{A} e^{t/\phi_2}\right)^4}. \tag{B.12}$$

Hence the solution of the LNA for a given tree, i, is given by (B.9) and (B.11), via (B.10) and (B.12).

Appendix C

The diffusion approximation of stochastic kinetic models

In this appendix we discuss a Markov jump process representation of a reaction network as an alternative representation of the systems presented in this thesis, and then outline the diffusion approximation to such a system.

C.1 Reaction networks

Let us consider a system of (pseudo-) reactions involving d species $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_d$, and r reactions $R_1, R_2, \ldots, R_r$. Typically, these reactions are written as

$$\begin{aligned}
R_1 &: p_{11}\mathcal{X}_1 + p_{12}\mathcal{X}_2 + \cdots + p_{1d}\mathcal{X}_d \quad \longrightarrow \quad q_{11}\mathcal{X}_1 + q_{12}\mathcal{X}_2 + \cdots + q_{1d}\mathcal{X}_d \\
R_2 &: p_{21}\mathcal{X}_1 + p_{22}\mathcal{X}_2 + \cdots + p_{2d}\mathcal{X}_d \quad \longrightarrow \quad q_{21}\mathcal{X}_1 + q_{22}\mathcal{X}_2 + \cdots + q_{2d}\mathcal{X}_d \\
&\vdots \qquad\qquad\qquad\qquad\quad \vdots \qquad\qquad\qquad\quad \vdots \\
R_r &: p_{r1}\mathcal{X}_1 + p_{r2}\mathcal{X}_2 + \cdots + p_{rd}\mathcal{X}_d \quad \longrightarrow \quad q_{r1}\mathcal{X}_1 + q_{r2}\mathcal{X}_2 + \cdots + q_{rd}\mathcal{X}_d,
\end{aligned}$$

where $X_{i,t}$ is the number of species $\mathcal{X}_i$ at time t and $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})'$. Furthermore, we define P to be the matrix with (i,j)th element p_{ij}, with an analogous definition for Q. The matrix P is often referred to as the matrix of reactants, whilst Q is the matrix of products. This terminology is used as these methods were originally introduced to describe chemical reactions. In addition, let S be the stoichiometry matrix, where $S = (Q - P)'$. The stoichiometry matrix allows for easy identification of the specific species involved in a specific reaction, along with all the reactions that a specific species is involved with. It is clear that if the ith reaction, R_i, occurs, we update the state of the system X_t by adding

the ith column of S to X_t.

Attached to each reaction (under the assumption of mass action kinetics) is a stochastic rate constant, denoted θ_i, $i = 1, \ldots, r$, along with an associated hazard function (or rate law), which we denote $h_i(x, \theta_i)$, $i = 1, \ldots, r$. The hazard function of each reaction represents the probability that the reaction will occur in the interval $(t, t+dt]$ (where dt is small), conditional on X_t. Thus, the probability of reaction R_i over the interval $(t, t+dt]$ is $h_i(x, \theta_i)dt$. Note that, in the absence of any other reactions, the time to reaction R_i is distributed as an exponential, $\text{Exp}\{h_i(x, \theta_i)\}$, random quantity. It should be clear that reaction events occur continuously in time. When a reaction occurs, the effect is to change the system by a discrete amount. The times and types of each reaction event depend only on the current state of the system, this is the Markov property. Hence, the system is most naturally represented as a Markov jump process, that is, a continuous-time, discrete valued stochastic process.

The form of the hazard function, $h_i(x, \theta_i)$, is determined by the order of the reaction R_i. A zeroth-order reaction of the form

$$R_i : \emptyset \xrightarrow{\theta_i} \mathcal{X},$$

leads to the hazard, $h_i(x, \theta_i) = \theta_i$. Of course such a reaction is impossible in 'realistic' circumstances, as things do not just appear out of thin air. However such a reaction can be useful for modelling processes such as immigration into a system. More common are first-order reactions

$$R_i : \mathcal{X}_j \xrightarrow{\theta_i} ?$$

with an associated hazard function $h_i(x, \theta_i) = \theta_i x_j$, or second-order reactions

$$R_i : \mathcal{X}_j + \mathcal{X}_k \xrightarrow{\theta_i} ?,$$

with $h_i(x, \theta_i) = \theta_i x_j x_k$. The above second-order reaction occurs when one of species $\mathcal{X}_j$ encounters one of species $\mathcal{X}_k$ whilst moving around randomly, driven by Brownian motion. Moreover, there is an alternative second-order reaction to the one given above, namely

$$R_i : 2\mathcal{X}_j \xrightarrow{\theta_i} ?$$

In this situation the accompanying hazard is given by $h_i(x, \theta_i) = \theta_i x_j (x_j - 1)/2$.

In general, the hazard function for any reaction is proportional to the product of binomial coefficients, and hence

$$h_i(x, \theta_i) = \theta_i \prod_{j=1}^{d} \binom{x_{j,t}}{p_{ij}}. \tag{C.1}$$

It is (usually) convention to express third-order (and higher) reactions as a combination of first and second-order reactions. Note that the hazard function (C.1) differs slightly from the standard mass action rate laws of continuous deterministic modelling, but can be seen as consistent (up to a constant of proportionality) asymptotically in the high concentration limit.

C.2 The Gillespie algorithm

For most systems of reasonable complexity, the transition probability mass function associated with the jump process will be intractable. However it is possible to generate realisations of the system using the Gillespie algorithm or Gillespie's direct method (Gillespie, 1977). This turns out to be reasonably straightforward. We note that there is an alternative (but equivalent) algorithm known as the first reaction method (see Gillespie (1976)), although Gillespie's direct method is to be preferred as it is typically more efficient.

As discussed above, each reaction R_i has associated with it, the hazard function $h_i(x, \theta_i)$. Hence, it follows that the hazard for any reaction (of any type) occurring is given by

$$h_0(x, \theta) \equiv \sum_{i=1}^{r} h_i(x, \theta_i),$$

where $\theta = (\theta_1, \ldots, \theta_r)'$. Therefore, the time to the next reaction (of any type) is $\text{Exp}\{h_0(x, \theta)\}$. The probability of the next reaction to occur being R_i is $h_i(x, \theta_i)/h_0(x, \theta)$ (see Wilkinson (2011) for a proof). Given the time to the next reaction and the reaction type, we may update X_t, and thus simulation may continue. The specific steps of the Gillespie algorithm are detailed in Algorithm 8.

Whilst it is possible to 'speed up' the Gillespie algorithm, we note that for many systems of interest the algorithm will still be computationally demanding (as discussed in Boys et al. (2008)). We therefore eschew exact simulation, in favour of computational speed, by considering a continuous approximation to the system, namely the diffusion approximation.

C.3 The diffusion approximation (chemical Langevin equation)

Whilst other approximations to the system in question are possible, such as the Poisson time-step method (Gibson and Bruck, 2000) or the τ-leap method (Gillespie, 2001), we

Algorithm 8 The Gillespie Algorithm

1. Set $t = t_0$ and initialise $\theta = (\theta_1, \ldots, \theta_r)'$ and $x_{t_0} = (x_1, \ldots, x_d)'$.
2. For $i = 1, \ldots, r$, calculate the reaction hazard $h_i(x, \theta_i)$ based on the current state x.
3. Calculate the combined reaction hazard, $h_0(x, \theta) \equiv \sum_{i=1}^{r} h_i(x, \theta_i)$.
4. Simulate the time to the next reaction, t^*, as an $\text{Exp}\{h_0(x, \theta)\}$ random quantity and set $t = t + t^*$.
5. Simulate the next reaction type, R_i, as a discrete random quantity with probabilities $h_i(x, \theta_i)/h_0(x, \theta)$, $i = 1, \ldots, r$.
6. Update x according to reaction R_i by setting $x = x + S_{\cdot i}$, where $S_{\cdot i}$ is the ith column of S (the stoichiometry matrix).
7. Output x and t. If $t < T_{\max}$ return to step 2.

continue the theme of this thesis and consider the diffusion approximation (often referred to as the chemical Langevin equation). Let us consider a sufficiently small time interval, $(t, t + dt]$, making the assumption that the hazard functions associated with each reaction are constant over the interval. It is clear that the number of R_i reactions occurring within the interval follows a Poisson distribution, independent of all other reaction types. If we denote the number of occurrences of R_i by O_i, then, providing dt is 'small'

$$O_i \overset{approx}{\sim} Po\left\{h_i\left(x, \theta_i\right) dt\right\}.$$

We let dO_t be the r-vector representing the number of occurrences in the interval of each reaction type, R_i, that is $dO_{t_i} = O_i$. Hence, it follows that the change in X_t (dX_t) is given by SdO_t, providing a sufficiently small time-step. From the above we have that

$$\begin{aligned} \text{E}(dO_t) &= h(x, \theta)dt, \\ \text{Var}(dO_t) &= \text{diag}\{h(x, \theta)\}dt, \end{aligned}$$

with $h(x, \theta) = (h_1(x, \theta_1), \ldots, h_r(x, \theta_r))'$ and $\text{diag}\{h(x, \theta)\}$ being the matrix with the elements of $h(x, \theta)$ along the leading diagonal, and zeros elsewhere. Whence

$$dO_t = h(X_t, \theta)\, dt + \sqrt{\text{diag}\{h(X_t, \theta)\}}\, dW_t, \tag{C.2}$$

where W_t is a r-vector of (uncorrelated) standard Brownian motion processes. The form of (C.2) is that of an Itô SDE, and is analogous to the SDEs considered within this thesis.

Noting that $dX_t = SdO_t$ leads to the diffusion approximation

$$dX_t = Sh(X_t, \theta)\, dt + \sqrt{S\text{diag}\{h(X_t, \theta)\}}\, dW_t, \quad X_0 = x_0. \tag{C.3}$$

As written, (C.3) is slightly unconventional, as W_t is typically of a higher dimension than X_t. We may however write that

$$\text{Var}(dX_t) = S\text{diag}\{h(X_t, \theta)\}S' dt,$$

giving rise to the SDE

$$dX_t = Sh(X_t, \theta)\, dt + \sqrt{S\text{diag}\{h(X_t, \theta)\}S'}\, dW_t, \quad X_0 = x_0, \tag{C.4}$$

where W_t is now a d-vector of (uncorrelated) standard Brownian motion processes (with the same dimension as X_t). The SDE in (C.4) is the one most commonly referred to as the chemical Langevin equation, and characterises the diffusion process which most closely reflects the dynamics of the attached system (expressed through the reaction network). In keeping with the notation of this thesis (and to harmonise with the SDE presented in (3.1)) we note that

$$\alpha(X_t, \theta) = Sh(X_t, \theta) \quad \text{and} \quad \beta(X_t, \theta) = S\text{diag}\{h(X_t, \theta)\}S'.$$

We can then proceed with the methods discussed in this thesis using the standard techniques outlined previously, that is, a realisation of X_t may be obtained through the Euler-Maruyama approximation.

A more formal derivation of the chemical Langevin equation can be ascertained via the Fokker-Planck equation (2.21), and we refer the reader to (Gillespie, 1992, 2000) for details. Further discussion of the diffusion approximation can be found in Allen (2010), Wilkinson (2011) or Golightly and Gillespie (2013) amongst others.

C.4 Example: Lotka-Volterra model

The Lotka-Volterra model of predator-prey dynamics was previously considered in Sections 2.7.4 and 3.5.2. It was developed independently by Lotka (1925) and Volterra (1926)

and is characterised by the three reactions

$$\begin{aligned} R_1 : &\quad \mathcal{X}_1 &\xrightarrow{\theta_1}& \quad 2\mathcal{X}_1 \\ R_2 : &\quad \mathcal{X}_1 + \mathcal{X}_2 &\xrightarrow{\theta_2}& \quad 2\mathcal{X}_2 \\ R_3 : &\quad \mathcal{X}_2 &\xrightarrow{\theta_3}& \quad \emptyset. \end{aligned} \tag{C.5}$$

Clearly the three reactions involve only two species, with $\mathcal{X}_1$ representing the prey and $\mathcal{X}_2$ the predators. To that end, R_1 may be thought of as a prey reproduction, R_2 represents a predator reproduction via a prey death and R_3 is a predator death. For the reaction network in (C.5) we have the stoichiometry matrix

$$S = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix},$$

and through the implementation of (C.1) we obtain the hazard functions

$$\begin{aligned} h_1(X, \theta_1) &= \theta_1 X_1, \\ h_2(X, \theta_2) &= \theta_2 X_1 X_2, \\ h_3(X, \theta_3) &= \theta_3 X_2. \end{aligned}$$

Denoting the state of the system at time t by $X_t = (X_{1,t}, X_{2,t})'$, we arrive (using (C.4)) at the SDE

$$dX_t = \begin{pmatrix} \theta_1 X_{1,t} - \theta_2 X_{1,t} X_{2,t} \\ \theta_2 X_{1,t} X_{2,t} - \theta_3 X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \theta_1 X_{1,t} + \theta_2 X_{1,t} X_{2,t} & -\theta_2 X_{1,t} X_{2,t} \\ -\theta_2 X_{1,t} X_{2,t} & \theta_3 X_{2,t} + \theta_2 X_{1,t} X_{2,t} \end{pmatrix}^{\frac{1}{2}} dW_t, \tag{C.6}$$

with initial condition $X_0 = x_0$, which is the SDE presented in (3.42).

We obtain realisations of the true discrete stochastic Lotka-Volterra system by means of the Gillespie algorithm (Algorithm 8), and realisations of the diffusion approximation (C.6) using the Euler-Maruyama approximation. Figure C.1 depicts a single realisation of the Lotka-Volterra model from both the true discrete stochastic system and the diffusion approximation. As in Section 3.5.2, we again follow Boys et al. (2008) and set $\theta = (\theta_1, \theta_2, \theta_3)' = (0.5, 0.0025, 0.3)'$, with $x_0 = (71, 79)'$. Note that the diffusion approximation is found using the Euler-Maruyama approximation with time-step, $\Delta t = 0.01$. Figure C.2 compares 95% credible regions for the number of prey and predators in the Lotka-Volterra model obtained from 100K simulations of the Gillespie algorithm and the Euler-Maruyama approximation ($\Delta t = 0.01$), for the conditions given above. Whilst simulation occurs on a finer grid (or in the case of the Gillespie algorithm, any moment in time),

output is collated every 0.1 such that the populations of each species may be compared. We note the remarkable similarities between simulations of the true discrete stochastic system and the diffusion approximation.

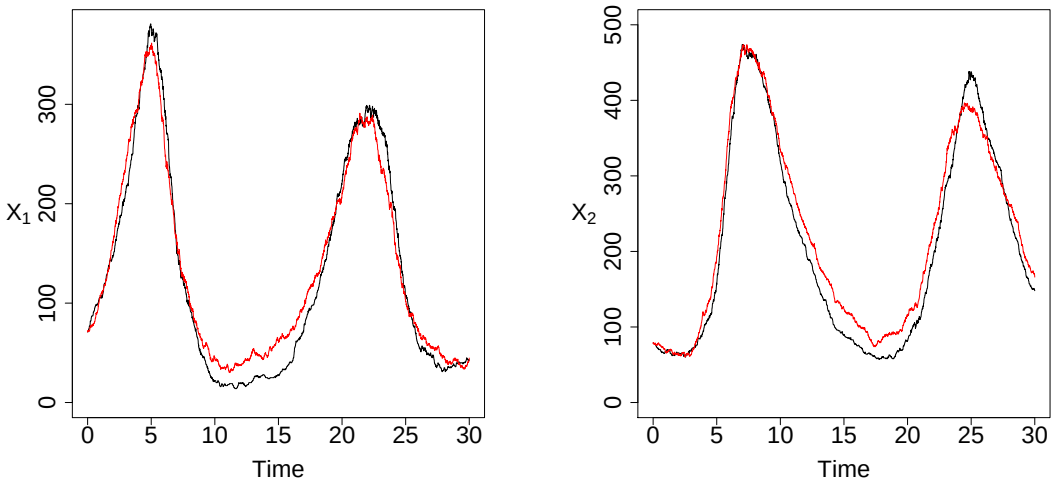

Figure C.1: A single realisation of prey (X_1) and predator (X_2) in the Lotka-Volterra model, $x_0 = (71, 79)'$ and $\theta = (0.5, 0.0025, 0.3)'$. Black: Gillespie algorithm. Red: Diffusion approximation (Euler-Maruyama, $\Delta t = 0.01$).

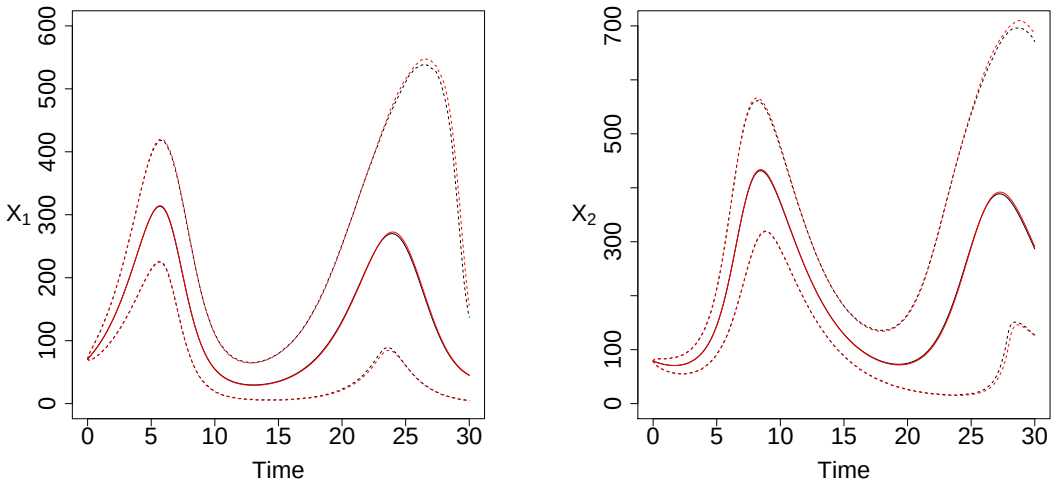

Figure C.2: Lotka-Volterra model. 95% credible region (dashed line) and mean (solid line) for numbers of prey (X_1) and predator (X_2) on a uniform grid of step-size 0.1, $x_0 = (71, 79)'$ and $\theta = (0.5, 0.0025, 0.3)'$. Black: Gillespie algorithm. Red: Diffusion approximation (Euler-Maruyama, $\Delta t = 0.01$).

Bibliography

Aït-Sahalia, Y. (2002), 'Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach', *Econometrica* **70**(1), 223–262.

Aït-Sahalia, Y. (2008), 'Closed-form likelihood expansions for multivariate diffusions', *The Annals of Statistics* **36**(2), 906–937.

Allen, L. J. S. (2008), An introduction to stochastic epidemic models, *in* F. Brauer, P. van den Driessche and J. Wu, eds, 'Mathematical Epidemiology', Vol. 1945 of *Lecture Notes in Mathematics*, Springer Berlin Heidelberg, pp. 81–130.

Allen, L. J. S. (2010), *An Introduction to Stochastic Processes with Applications to Biology*, 2nd edn, CRC Press.

Alonso, D., McKane, A. J. and Pascual, M. (2007), 'Stochastic amplification in epidemics', *Journal of the Royal Society Interface* **4**(14), 575–582.

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.

Arnold, L. (1974), *Stochastic Differential Equations: Theory and Applications*, John Wiley & Sons, New York.

Bachelier, L. (1964), *Theory of Speculation, the Random Character of Stock Prices, translation of Bacheliers 1900 doctoral thesis*, Cambridge: MIT.

Barbour, A. D. (1974), 'On a functional central limit theorem for Markov population processes', *Advances in Applied Probability* **6**(1), 21–39.

Berglund, M., Sunnåker, M., Adiels, M., Jirstrand, M. and Wennberg, B. (2011), 'Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations', *Mathematical Medicine and Biology* **29**(4), 361–384.

Beskos, A., Kalogeropoulos, K. and Pazos, E. (2013), 'Advanced MCMC methods for sampling on diffusion pathspace', *Stochastic Processes and their Applications* **123**(4), 1415–1453.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2009), 'Monte Carlo maximum likelihood estimation for discretely observed diffusion processes', *Annals of Statistics* **37**, 223–245.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O. 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**, 333–382.

Bibby, B. M. and Sørensen, M. (2001), 'Simplified estimating functions for diffusion models with a high-dimensional parameter', *Scandinavian Journal of Statistics* **28**(1), 99–112.

Black, F. and Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**(3), 637–654.

Boys, R. J., Wilkinson, D. J. and Kirkwood, T. B. L. (2008), 'Bayesian inference for a discretely observed stochastic kinetic model', *Statistics and Computing* **18**(2), 125–135.

Brown, R. (1828), 'A brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies', *Philosophical Magazine Series 2* **4**(21), 161–173.

Capasso, V. and Morale, D. (2009), 'Stochastic modelling of tumour-induced angiogenesis', *Journal of Mathematical Biology* **58**(1-2), 219–233.

Chen, W. Y. and Bokka, S. (2005), 'Stochastic modeling of nonlinear epidemiology', *Journal of Theoretical Biology* **234**(4), 455–470.

Chiarella, C., Hung, H. and Tô, T. D. (2009), 'The volatility structure of the fixed income market under the HJM framework: a nonlinear filtering approach', *Computational Statistics & Data Analysis* **53**(6), 2075–2088.

Chib, S., Pitt, M. K. and Shephard, N. (2004), Likelihood based inference for diffusion driven models, Economics Papers 2004-W20, Economics Group, Nuffield College, University of Oxford.

Clark, J. (1990), The simulation of pinned diffusions, *in* 'Decision and Control, 1990', Proceedings of the 29th IEEE Conference on, IEEE, pp. 1418–1420.

Cox, J. C., Ingersoll Jr., J. E. and Ross, S. A. (1985), 'An intertemporal general equilibrium model of asset prices', *Econometrica* **53**(2), 363–384.

Dargatz, C. (2007), A diffusion approximation for an epidemic model, Discussion paper, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München 517.

Del Moral, P. and Murray, L. M. (2015), 'Sequential Monte Carlo with highly informative observations', *SIAM/ASA Journal on Uncertainty Quantification* **3**(1), 969–997.

Delyon, B. and Hu, Y. (2006), 'Simulation of conditioned diffusion and application to parameter estimation', *Stochastic Processes and their Applications* **116**, 1660–1675.

Ditlevsen, S. and De Gaetano, A. (2005), 'Mixed effects in stochastic differential equation models', *REVSTAT-Statistical Journal* **3**(2), 137–153.

Donnet, S., Foulley, J. and Samson, A. (2010), 'Bayesian analysis of growth curves using mixed models defined by stochastic differential equations', *Biometrics* **66**(3), 733–741.

Donnet, S. and Samson, A. (2008), 'Parametric inference for mixed models defined by stochastic differential equations', *ESAIM: Probability and Statistics* **12**, 196–218.

Durham, G. B. and Gallant, A. R. (2002), 'Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes', *Journal of Business and Economic Statistics* **20**, 297–338.

Einstein, A. (1905), 'The theory of Brownian movement', *Annalen der Physik* **17**, 549.

Elerian, O., Chib, S. and Shephard, N. (2001), 'Likelihood inference for discretely observed nonlinear diffusions', *Econometrica* **69**, 959–993.

Elf, J. and Ehrenberg, M. (2003), 'Fast evaluation of fluctuations in biochemical networks with the linear noise approximation', *Genome Research* **13**(11), 2475–2484.

Eraker, B. (2001), 'MCMC analysis of diffusion models with application to finance', *Journal of Business and Economic Statistics* **19**(2), 177–191.

Fearnhead, P. (2006), 'The stationary distribution of allele frequencies when selection acts at unlinked loci', *Theoretical Population Biology* **70**(3), 376–386.

Fearnhead, P. (2008), 'Computational methods for complex stochastic systems: a review of some alternatives to MCMC', *Statistics and Computing* **18**, 151–171.

Fearnhead, P., Giagos, V. and Sherlock, C. (2014), 'Inference for reaction networks using the linear noise approximation', *Biometrics* **70**(2), 457–466.

Featherstone, K., Hey, K., Momiji, H., McNamara, A. V., Patist, A. L., Woodburn, J., Spiller, D. G., Christian, H. C., McNeilly, A. S., Mullins, J. J., Finkenstdt, B. F., Rand, D. A., White, M. R. H. and Davis, J. R. E. (2016), 'Spatially coordinated dynamic gene transcription in living pituitary tissue', *eLife* **5**, e08494.

Fehlberg, E. (1969), *Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems*, NASA technical report, National Aeronautics and Space Administration.

Feller, W. (1952), 'The parabolic differential equations and the associated semi-groups of transformations', *Annals of Mathematics* **55**, 468–519.

Ferm, L., Lötstedt, P. and Hellander, A. (2008), 'A hierarchy of approximations of the master equation scaled by a size parameter', *Journal of Scientific Computing* **34**(2), 127–151.

Finkenstädt, B., Heron, E. A., Komorowski, M., Edwards, K., Tang, S., Harper, C. V., Davis, J. R. E., White, M. R. H., Millar, A. J. and Rand, D. A. (2008), 'Reconstruction of transcriptional dynamics from gene reporter data using differential equations', *Bioinformatics* **24**(24), 2901–2907.

Finkenstädt, B., Woodcock, D. J., Komorowski, M., Harper, C. V., Davis, J. R. E., White, M. R. H. and Rand, D. A. (2013), 'Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: an application to single cell data', *Annals of Applied Statistics* **7**, 1960–1982.

Fogelson, A. (1984), 'A mathematical model and numerical method for studying platelet adhesion and aggregation during blood clotting', *Journal of Computational Physics* **56**(1), 111–134.

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

Gamerman, D. and Lopes, H. F. (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall/CRC Texts in Statistical Science, 2nd edn, Chapman & Hall.

Gelfand, A. E. and Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409.

Gelman, A. (1996), Inference and monitoring convergence, *in* W. Gilks, S. Richardson and D. Spiegelhalter, eds, 'Markov Chain Monte Carlo in Practice', Chapman & Hall/CRC Interdisciplinary Statistics, Taylor & Francis.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013), *Bayesian Data Analysis*, Chapman & Hall/CRC Texts in Statistical Science, 3rd edn, Taylor & Francis.

Gelman, A., Roberts, G. O. and Gilks, W. R. (1996), Efficient Metropolis jumping rules, *in* 'Bayesian Statistics, 5 (Alicante, 1994)', Oxford Sci. Publ., Oxford Univ. Press, pp. 599–607.

Geman, S. and Geman, D. (1984), 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **PAMI-6**(6), 721–741.

Geweke, J. (1989), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica* **57**(6), 1317–1339.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to calculating posterior moments, *in* J. M. Bernardo, J. Berger, A. P. Dawid and J. F. M. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, pp. 169–193.

Giagos, V. (2011), Inference for auto-regulatory genetic networks using diffusion process approximations, PhD thesis, Lancaster University.

Gibson, M. A. and Bruck, J. (2000), 'Efficient exact stochastic simulation of chemical systems with many species and many channels', *The Journal of Physical Chemistry A* **104**(9), 1876–1889.

Gilioli, G., Pasquali, S. and Ruggeri, F. (2008), 'Bayesian inference for functional response in a stochastic predator–prey system', *Bulletin of Mathematical Biology* **70**(2), 358–381.

Gillespie, C. S. and Golightly, A. (2010), 'Bayesian inference for generalized stochastic population growth models with application to aphids', *Journal of the Royal Statistical Society - Series C (Applied Statistics)* **59**(2), 341–357.

Gillespie, D. T. (1976), 'A general method for numerically simulating the stochastic time evolution of coupled chemical reactions', *Journal of Computational Physics* **22**(4), 403–434.

Gillespie, D. T. (1977), 'Exact stochastic simulation of coupled chemical reactions', *Journal of Physical Chemisty* **81**(25), 2340–2361.

Gillespie, D. T. (1992), 'A rigorous derivation of the chemical master equation', *Physica A: Statistical Mechanics and its Applications* **188**(13), 404–425.

Gillespie, D. T. (2000), 'The chemical langevin equation', *The Journal of Chemical Physics* **113**(1), 297–306.

Gillespie, D. T. (2001), 'Approximate accelerated stochastic simulation of chemically reacting systems', *The Journal of Chemical Physics* **115**(4), 1716–1733.

Golightly, A. and Gillespie, C. S. (2013), Simulation of stochastic kinetic models, *in* 'In Silico Systems Biology', Humana Press, pp. 169–187.

Golightly, A., Henderson, D. A. and Sherlock, C. (2015), 'Delayed acceptance particle MCMC for exact inference in stochastic kinetic models', *Statistics and Computing* **25**(5), 1039–1055.

Golightly, A. and Wilkinson, D. J. (2005), 'Bayesian inference for stochastic kinetic models using a diffusion approximation', *Biometrics* **61**(3), 781–788.

Golightly, A. and Wilkinson, D. J. (2006), 'Bayesian sequential inference for nonlinear multivariate diffusions', *Statistics and Computing* **16**, 323–338.

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

Golightly, A. and Wilkinson, D. J. (2010), Markov chain Monte Carlo algorithms for SDE parameter estimation, *in* 'Learning and Inference in Computational Systems Biology', Computational Molecular Biology, MIT Press.

Golightly, A. and Wilkinson, D. J. (2011), 'Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo', *Interface Focus* **1**(6), 807–820.

Gray, A., Greenhalgh, D., Hu, L., Mao, X. and Pan, J. (2011), 'A stochastic differential equation SIS epidemic model', *SIAM Journal on Applied Mathematics* **71**(3), 876–902.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**(1), 97–109.

Heidelberger, P. and Welch, P. D. (1983), 'Simulation run length control in the presence of an initial transient', *Operations Research* **31**(6), 1109–1144.

Hey, K. L., Momiji, H., Featherstone, K., Davis, J. R. E., White, M. R. H., Rand, D. A. and Finkenstdt, B. (2015), 'A stochastic transcriptional switch model for single cell imaging data', *Biostatistics* **16**(4), 655–669.

Heydari, J. J., Lawless, C., Lydall, D. A. and Wilkinson, D. J. (2014), 'Fast Bayesian parameter estimation for stochastic logistic growth models', *BioSystems* **122**, 55–72.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of basic Engineering* **82**(1), 35–45.

Kalogeropoulos, K., Roberts, G. O. and Dellaportas, P. (2010), 'Inference for stochastic volatility models using time change transformations', *The Annals of Statistics* **38**(2), 784–807.

Kermack, W. O. and McKendrick, A. G. (1927), 'A contribution to the mathematical theory of epidemics', *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **115**(772), 700–721.

Klim, S., Mortensen, S. B., Kristensen, N. R., Overgaard, R. V. and Madsen, H. (2009), 'Population stochastic modelling (PSM)-an R package for mixed-effects models based on stochastic differential equations', *Computer Methods and Programs in Biomedicine* **94**(3), 279–289.

Kloeden, P. E. and Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York.

Komorowski, M., Finkenstadt, B., Harper, C. and Rand, D. (2009), 'Bayesian inference of biochemical kinetic parameters using the linear noise approximation', *BMC Bioinformatics* **10**(1), 343.

Komorowski, M., Finkenstädt, B. and Rand, D. A. (2010), 'Using a single fluorescent reporter gene to infer half-life of extrinsic noise and other parameters of gene expression', *Biophysical Journal* **98**(12), 2759–2769.

Koopman, S. J., Shephard, N. and Creal, D. (2009), 'Testing the assumptions behind importance sampling', *Journal of Econometrics* **49**(1), 2–11.

Kou, S. C., Olding, B. P., Lysy, M. and Liu, J. S. (2012), 'A multiresolution method for parameter estimation of diffusion processes', *Journal of the American Statistical Association* **107**, 1558–1574.

Kurtz, T. G. (1970), 'Solutions of ordinary differential equations as limits of pure jump Markov processes', *Journal of Applied Probability* **7**(1), 49–58.

Kurtz, T. G. (1971), 'Limit theorems for sequences of jump Markov processes approximating ordinary differential processes', *Journal of Applied Probability* **8**(2), 344–356.

Leung, H. K. (1985), 'Expansion of the master equation for a biomolecular selection model', *Bulletin of Mathematical Biology* **47**(2), 231–238.

Lin, M., Chen, R. and Mykland, P. (2010), 'On generating Monte Carlo samples of continuous diffusion bridges', *Journal of the American Statistical Association* **105**, 820–838.

Lindström, E. (2012), 'A regularized bridge sampler for sparsely sampled diffusions', *Statistics and Computing* **22**(1), 615–623.

Lotka, A. J. (1925), *Elements of Physical Biology*, Williams & Wilkins Company.

Lu, Z. H., Chow, S. M., Sherwood, A. and Zhu, H. (2015), 'Bayesian analysis of ambulatory blood pressure dynamics with application to irregularly spaced sparse data', *Annals of Applied Statistics* . In press.

Marchand, J. L. (2011), Conditioning diffusions with respect to partial observations. Available from `http://arxiv.org/abs/1105.1608`.

Matis, J. H., Kiffe, T. R., Matis, T. I. and Stevenson, D. E. (2006), 'Application of population growth models based on cumulative size to Pecan aphids', *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 425–449.

Matis, T. I., Parajulee, M. N., Matis, J. H. and Shrestha, R. B. (2008), 'A mechanistic model based analysis of cotton aphid population dynamics data', *Agricultural and Forest Entomology* **10**(4), 355–362.

McCulloch, C. E. and Searle, S. R. (2004), *Generalized, Linear, and Mixed Models*, Wiley series in probability and statistics. Applied probability and statistics section, Wiley.

McNeil, D. R. (1973), 'Diffusion limits for congestion models', *Journal of Applied Probability* **10**(2), 368–376.

Merton, R. C. (1973), 'Theory of rational option pricing', *The Bell Journal of Economics and Management Science* **4**(1), 141–183.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**(6), 1087–1092.

Øksendal, B. K. (2003), *Stochastic Differential Equations: An Introduction with Applications*, Universitext (1979), Springer.

Overgaard, R. V., Jonsson, N., Tornøe, C. W. and Madsen, H. (2005), 'Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm', *Journal of Pharmacokinetics and Pharmacodynamics* **32**(1), 85–107.

Papaspiliopoulos, O. and Roberts, G. O. (2012), Importance sampling techniques for estimation of diffusion models, *in* 'Statistical Methods for Stochastic Differential Equations, Monographs on Statistics and Applied Probability', Chapman and Hall, pp. 311–337.

Papaspiliopoulos, O., Roberts, G. O. and Skold, M. (2003), Non-centered parameterisations for hierarchical models and data augmentation, *in* J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith and M. West, eds, 'Bayesian Statistics 7', pp. 307–327.

Papaspiliopoulos, O., Roberts, G. O. and Stramer, O. (2013), 'Data augmentation for diffusions', *Journal of Computational and Graphical Statistics* **22**, 665–688.

Pedersen, A. R. (1995), 'A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations', *Scandinavian Journal of Statistics* **22**(1), 55–71.

Petzold, L. (1983), 'Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations', *SIAM Journal on Scientific and Statistical Computing* **4**(1), 136–148.

Picchini, U. (2014), 'Inference for SDE models via Approximate Bayesian Computation', *Journal of Computational and Graphical Statistics* **23**(4), 1080–1100.

Picchini, U. and Ditlevsen, S. (2011), 'Practical estimation of high dimensional stochastic differential mixed-effects models', *Computational Statistics and Data Analysis* **55**(3), 1426–1444.

Picchini, U., Gaetano, S. and Ditlevsen, S. (2010), 'Stochastic differential mixed-effects models', *Scandinavian Journal of Statistics* **37**, 67–90.

Pinheiro, J. C. and Bates, D. (2009), *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing, Springer.

Pitt, M., Silva, R., Giordani, P. and Kohn, R. (2010), Auxiliary particle filtering within adaptive Metropolis-Hastings sampling. Available from `http://arxiv.org/abs/1006.1914`.

Raftery, A. E. and Lewis, S. (1992), 'How many iterations in the Gibbs sampler?', *Bayesian Statistics* **4**(2), 763–773.

Raftery, A. E. and Lewis, S. (1996), Implementing MCMC, *in* 'Markov Chain Monte Carlo in Practice', W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, Chapman & Hall, London, pp. 115–130.

Ramshaw, J. (1985), 'Augmented langevin approach to fluctuations in nonlinear irreversible processes', *Journal of Statistical Physics* **38**(3-4), 669–680.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997), 'Weak convergence and optimal scaling of random walk Metropolis algorithms', *Annals of Applied Probability* **7**, 110–120.

Roberts, G. O. and Rosenthal, J. S. (2001), 'Optimal scaling for various Metropolis-Hastings algorithms', *Statistical Science* **16**(4), 351–367.

Roberts, G. O. and Stramer, O. (2001), 'On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm', *Biometrika* **88**(3), 603–621.

Rogers, L. C, G. and Williams, D. (2000), *Diffusions, Markov processes and martingales*, 2nd edn, Cambridge University Press, Cambridge.

Schauer, M., van der Meulen, F. and van Zanten, H. (2016), 'Guided proposals for simulating multi-dimensional diffusion bridges', *Bernoulli* . In press.

Sermaidis, G., Papaspiliopoulos, O., Roberts, G. O., Beskos, A. and Fearnhead, P. (2013), 'Markov chain Monte Carlo for exact inference for diffusions', *Scandinavian Journal of Statistics* **40**, 294–321.

Shephard, N. and Pitt, M. K. (1997), 'Likelihood analysis of non-Gaussian measurement time series', *Biometrika* **84**(3), 653–667.

Sherlock, C. (2013), 'Optimal scaling of the random walk Metropolis: general criteria for the 0.234 acceptance rule', *Journal of Applied Probability* **50**(1), 1–15.

Sherlock, C. and Roberts, G. (2009), 'Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets', *Bernoulli* **15**(3), 774–798.

Shoji, I. (2011), 'A note on convergence rate of a linearization method for the discretization of stochastic differential equations', *Communications in Nonlinear Science and Numerical Simulation* **16**(7), 2667–2671.

Shoji, I. and Ozaki, T. (1998), 'A statistical method of estimation and simulation for systems of stochastic differential equations', *Biometrika* **85**(1), 240–243.

Stramer, O. and Bognar, M. (2011), 'Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach', *Bayesian Analysis* **6**(2), 231–258.

Stramer, O., Bognar, M. and Scheider, P. (2010), 'Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions', *Journal of Financial Econometrics* **8**, 450–480.

Stramer, O. and Yan, J. (2006), On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation, Technical report, University of Iowa.

Tanner, M. A. and Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association* **82**(398), 528–540.

Tian, T., Burrage, K., Burrage, P. M. and Carletti, M. (2007), 'Stochastic delay differential equations for genetic regulatory networks', *Journal of Computational and Applied Mathematics* **205**(2), 696–707.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions (with discussion)', *The Annals of Statistics* **22**(4), 1701–1762.

Tornøe, C. W., Overgaard, R. V., Agerso, H., Nielsen, H. A., Madsen, H. and Jonsson, E. (2005), 'Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations', *Pharmaceutical Research* **22**(8), 1247–1258.

Tuckwell, H. C. (1987), 'Diffusion approximations to channel noise', *Journal of Theoretical Biology* **127**(4), 427–438.

van der Meulen, F. and Schauer, M. (2015), Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. Available from `http://arxiv.org/abs/1406.4704`.

van Kampen, N. G. (1981), 'The validity of nonlinear langevin equations', *Journal of Statistical Physics* **25**(3), 431–442.

van Kampen, N. G. (2007), *Stochastic Processes in Physics and Chemistry*, 3rd edn, North Holland.

Volterra, V. (1926), 'Fluctuations in the abundance of a species considered mathematically', *Nature* **118**, 558–560.

Wallace, E. W. J. (2010), 'A simplified derivation of the linear noise approximation'. **URL:** *http://arxiv.org/abs/1004.4280*

Wallace, E. W. J., Gillespie, D. T., Sanft, K. R. and Petzold, L. R. (2012), 'Linear noise approximation is valid over limited times for any chemical system that is sufficiently large', *IET Systems Biology* **6**(4), 102–115.

Walsh, J. B. (1981), 'A stochastic model of neural response', *Advances in Applied Probability* **13**(2), 231–281.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, 2nd edn, Springer-Verlag New York, Inc., New York, NY, USA.

Whitaker, G. A., Golightly, A., Boys, R. J. and Sherlock, C. (2016*a*), 'Bayesian inference for diffusion driven mixed-effects models', *Bayesian Analysis* . To appear.

Whitaker, G. A., Golightly, A., Boys, R. J. and Sherlock, C. (2016*b*), 'Improved bridge constructs for stochastic differential equations', *Statistics and Computing* . To appear.

Wiener, N. (1923), 'Differential-space', *Journal of Mathematics and Physics* **2**(1-4), 131–174.

Wilkinson, D. J. (2011), *Stochastic Modelling for Systems Biology*, Chapman & Hall/CRC Mathematical and Computational Biology, 2nd edn, CRC Press.