Bayesian Inference for Diffusion-Driven Mixed-Effects Models

Gavin A. Whitaker[*], Andrew Golightly[†], Richard J. Boys[‡], and Chris Sherlock[§]

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 paper. We extend a recently proposed MCMC scheme to include the SDE driven mixed-effects framework. Fundamental to our approach is the development of a novel construct that allows for efficient sampling of conditioned SDEs that may exhibit nonlinear dynamics between observation times. We apply the resulting scheme to synthetic data generated from a simple SDE model of orange tree growth, and real data on aphid numbers recorded under a variety of different treatment regimes. In addition, 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.

Keywords: stochastic differential equation, mixed-effects, Markov chain Monte Carlo, modified innovation scheme, linear noise approximation.

1 Introduction

Diffusion processes satisfying Itô stochastic differential equations (SDEs) are a class of continuous-time, continuous-valued Markov stochastic processes that can be used to model a wide range of phenomena. Examples include (but are not limited to) epidemics, financial time series, population dynamics (including predator-prey systems) and intracellular processes. 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).

Unfortunately, analytic intractability of SDEs governing most nonlinear multivariate diffusions can make likelihood-based inference methods problematic. Methods to overcome this difficulty include closed-form expansion of the transition density (Aït-Sahalia, 2002, 2008; Stramer et al., 2010), exact simulation approaches (Beskos et al.,

[*]School of Mathematics & Statistics, Newcastle University, Newcastle Upon Tyne, NE1 7RU
[†]School of Mathematics & Statistics, Newcastle University, Newcastle Upon Tyne, NE1 7RU, andrew.golightly@ncl.ac.uk
[‡]School of Mathematics & Statistics, Newcastle University, Newcastle Upon Tyne, NE1 7RU
[§]Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF

2006; Sermaidis et al., 2013) and use of the Euler–Maruyama approximation coupled with data augmentation (Pedersen, 1995; Elerian et al., 2001; Eraker, 2001; Durham and Gallant, 2002; Golightly and Wilkinson, 2008; Stramer and Bognar, 2011; Kou et al., 2012). 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 computationally inefficient when 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, that is, one which can be transformed to give a unit diffusion coefficient. 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 suggest that a solution might be found using the Euler–Maruyama discretisation.

1.1 Contributions and organisation of the paper

In this article we provide a method 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. The method makes use of a novel scheme that allows for observations made sparsely in time, as the process of interest may exhibit nonlinear dynamics between measurement times.

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 density for parameters and the components of the latent process. The intractability of the posterior density necessitates simulation techniques such as Markov chain Monte Carlo. As is well documented in Roberts and Stramer (2001), care must be taken in the design of the MCMC sampler due to dependence between the parameters entering the diffusion coefficient and the latent process. We therefore adapt the reparameterisation

technique (known as the modified innovation scheme) of Golightly and Wilkinson (2008) and Golightly and Wilkinson (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 approaches 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 bridge construct that is simple to implement 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 parameters of interest, allowing a marginal Metropolis–Hastings scheme targeting their posterior distribution. It should be noted, however, that evaluation of the Gaussian transition densities under the LNA require the solution of an ordinary differential equation (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.

We apply the methods to two 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. Second, 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.

The remainder of the article is organised as follows. The SDMEM framework is introduced in Section 2. Section 3 provides MCMC methods for Bayesian inference, with a novel bridge construct outlined in Section 3.2. The linear noise approximation and its application as an inferential model is discussed in Section 4. The methods are applied in Section 5 before conclusions are drawn in Section 6.

2 Stochastic differential mixed-effects models

Consider the case where we have N experimental units randomly chosen from a theoretical population, and associated with each unit i is a continuous-time d-dimensional

Itô process $\{X_t^i, t \geq 0\}$ governed by the SDE

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

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 $\sqrt{\beta}\sqrt{\beta}^T = \beta$ and W_t^i is a d-vector of (uncorrelated) standard Brownian motion processes. The p-vector parameter $\theta = (\theta_1, \ldots, \theta_p)^T$ is common to all units whereas the q-vectors $b^i = (b_1^i, \ldots, b_q^i)^T$, $i = 1, \ldots, N$, 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 $\psi = (\psi_1, \ldots, \psi_r)^T$. The model defined by (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 respectively.

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

$$Y_t^i = F^T X_t^i + \epsilon_t, \qquad \epsilon_t|\Sigma \overset{indep}{\sim} N(0, \Sigma), \qquad (2)$$

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 ($d_o < d$) and this aspect is explored further in Section 5.2.

Together (1) and (2) completely specify the stochastic differential mixed-effects model. However, 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

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

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

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

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

$$\Delta\tau \equiv \tau_{j,k+1} - \tau_{j,k} = \frac{t_{j+1} - t_j}{m}. \qquad (3)$$

The Euler–Maruyama approximation 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

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

for the latent process over the time interval $[t_j, t_{j+1}]$ for unit i. Hence, the complete latent trajectory associated with unit i is given by

$$(x^i)^T = ((x^i_{[0,1]})^T, (x^i_{(1,2]})^T \ldots, (x^i_{(n-1,n]})^T)$$

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, $b = (b^1, \ldots, b^N)^T$, ψ and Σ.

3 Bayesian inference

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

$$\pi(\theta, \psi, \Sigma, b, x | y) \propto \pi(\theta)\pi(\psi)\pi(\Sigma)\pi(b|\psi)\pi(x|\theta, b)\pi(y|x, \Sigma) \tag{4}$$

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

$$\pi(x|\theta, b) = \prod_{i=1}^{N} \prod_{j=0}^{n-1} \prod_{k=1}^{m} \pi(x^i_{\tau_{j,k}} | x^i_{\tau_{j,k-1}}, \theta, b^i) \tag{5}$$

where

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

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

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

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) we aim to construct a Markov chain Monte Carlo (MCMC) scheme which generates realisations from this posterior. 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. $\pi(\psi|\theta, \Sigma, b, x, y) = \pi(\psi|b)$.

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. 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.

3.1   Path updates

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

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

which suggests a scheme where unit-specific paths are updated separately. We now 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.

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

$$\pi(x_{(j,j+2)}|x_{t_j}, y_{t_{j+1}}, x_{t_{j+2}}) \propto \pi(y_{t_{j+1}}|x_{t_{j+1}}) \prod_{l=j}^{j+1} \prod_{k=1}^{m} \pi(x_{\tau_{l,k}}|x_{\tau_{l,k-1}}). \tag{6}$$

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 (6). We use an independence sampler with proposal density of the form

$$q(x_{(j,j+2)}|x_{t_j}, y_{t_{j+1}}, x_{t_{j+2}}) = q_1(x_{(j,j+1]}|x_{t_j}, y_{t_{j+1}})\, q_2(x_{(j+1,j+2)}|x_{t_{j+1}}, x_{t_{j+2}}). \tag{7}$$

Figure 1 gives an illustration of the updating procedure which can be applied over intervals (t_j, t_{j+2}), $j = 0, 1, \ldots, n-2$, with two additional Metropolis–Hastings steps (such as those described in Golightly and Wilkinson (2006)) that allow for updating x at times t_0 and t_n. Deriving 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 resulting trajectory is typically referred to as a diffusion bridge.

Several strategies for constructing diffusion bridges have been proposed in the literature. For example, Pedersen (1995) used the Euler–Maruyama scheme to generate
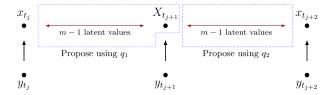
Figure 1: Path update illustration over a block of size $2m + 1$.

bridges myopically of the end point. Durham and Gallant (2002) use a linear Gaussian approximation of the distribution of the process conditional on the value at a previous and future time point, giving a construct known as the modified diffusion bridge. Extensions of this construct to the case of partial observation with additive Gaussian noise can be found in Golightly and Wilkinson (2008). 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. Consequently, use of the modified diffusion bridge 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. Several attempts to overcome this issue have been proposed in the recent literature.

A time-dependent combination of the Pedersen and modified diffusion bridge approaches was proposed by Lindström (2012). However, the resulting construct requires a model specific tuning parameter governing the relative weight of each contribution (either Pedersen or modified diffusion bridge). Moreover, the optimal value (in terms of maximising acceptance rate) may vary between observation intervals. 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 tuning parameters (namely the number of steps (and their size) in the leapfrog integrator) to maximise efficiency. Schauer et al. (2014) (see also Papaspiliopoulos and Roberts (2012)) 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 target SDE. The guiding term requires a tractable approximation of the unavailable transition densities governing the target process over the length of the inter-observation interval. Schauer et al. (2014) suggest using the transition densities associated with a class of linear processes, although we note that finding an approximation that is both accurate and computationally efficient may be difficult in practice. Moreover, such an approximation can suffer from computational efficiency due to the fact that it must be obtained at each intermediate time point.

In the next section we describe a novel bridge construct that requires no tuning parameters, is simple to implement (even when only a subset of components are observed

with Gaussian noise), computationally efficient and explicitly allows for the effect of the drift governing the target SDE.

3.2 An improved bridge construct

Consider a typical interval $[t_j, t_{j+1}]$, partitioned into m sub-intervals as in (3), over which we wish to generate a realisation of $\{X_t, t \in [t_j, t_{j+1}]\}$ conditional on $x_{t_j} \equiv x_{\tau_{j,0}}$ and the noisy measurement $y_{t_{j+1}} \equiv y_{\tau_{j,m}}$. Our approach builds on the modified diffusion bridge of Durham and Gallant (2002), which we briefly review before describing our extension.

Modified diffusion bridge

Key to constructing the modified diffusion bridge is an approximation of the joint distribution of $X_{\tau_{j,k}}$ and $Y_{t_{j+1}}$ (conditional on $x_{\tau_{j,k-1}}$). Under the Euler–Maruyama approximation $X_{\tau_{j,k}}|X_{\tau_{j,k-1}}$ and $Y_{t_{j+1}}|X_{\tau_{j,k}}$ are Gaussian, with the expressions for the mean and variance of the latter evaluated at $X_{\tau_{j,k-1}}$ to give a linear Gaussian structure. This leads to the approximation

$$\begin{pmatrix} X_{\tau_{j,k}} \\ Y_{t_{j+1}} \end{pmatrix} \bigg| x_{\tau_{j,k-1}} \sim N \left\{ \begin{pmatrix} x_{\tau_{j,k-1}} + \alpha_{j,k-1}\Delta\tau \\ F^T[x_{\tau_{j,k-1}} + \alpha_{j,k-1}\Delta^-] \end{pmatrix}, \right. $$
$$\left. \begin{pmatrix} \beta_{j,k-1}\Delta\tau & \beta_{j,k-1}F\Delta\tau \\ F^T\beta_{j,k-1}\Delta\tau & F^T\beta_{j,k-1}F\Delta^- + \Sigma \end{pmatrix} \right\} \quad (8)$$

where $\Delta^- = t_{j+1} - \tau_{j,k-1}$ and we have used the shorthand notation $\alpha(x_{\tau_{j,k-1}}) = \alpha_{j,k-1}$ and $\beta(x_{\tau_{j,k-1}}) = \beta_{j,k-1}$. Conditioning further on $y_{t_{j+1}}$ gives a Gaussian approximation of $\pi(x_{\tau_{j,k}}|x_{\tau_{j,k-1}}, y_{t_{j+1}})$, denoted $\widehat{\pi}(x_{\tau_{j,k}}|x_{\tau_{j,k-1}}, y_{t_{j+1}})$, which can be sampled recursively to give a bridge $x_{\tau_{j,0}}, \ldots, x_{\tau_{j,m}}$. In the case of no measurement error and observation of all components (so that $y_{t_{j+1}} = x_{t_{j+1}}$ and $F = I_d$, the $d \times d$ identity matrix), we obtain

$$\widehat{\pi}(x_{\tau_{j,k}}|x_{\tau_{j,k-1}}, x_{t_{j+1}}) = N \left(x_{\tau_{j,k}} \; ; \; x_{\tau_{j,k-1}} + \frac{x_{t_{j+1}} - x_{\tau_{j,k-1}}}{t_{j+1} - \tau_{j,k-1}}\Delta\tau \, , \, \frac{t_{j+1} - \tau_{j,k}}{t_{j+1} - \tau_{j,k-1}}\beta_{j,k-1}\Delta\tau \right)$$

which is the form of the modified diffusion bridge first described by Durham and Gallant (2002). In this case, $\widehat{\pi}(x_{\tau_{j,k}}|x_{\tau_{j,k-1}}, y_{t_{j+1}})$ can be seen as a linear approximation of the Brownian bridge SDE

$$dX_t = \frac{X_{t_{j+1}} - X_t}{t_{j+1} - t} \, dt + \sqrt{\beta(X_t)} \, dW_t. \quad (9)$$

Use of (9) has been justified by Delyon and Hu (2006), who show that the distribution of the target process (conditional on $x_{t_{j+1}}$) is absolutely continuous with respect to the distribution of the solution to (9). We may therefore expect that a Metropolis–Hastings scheme that uses a proposal based on a discretisation of (9) will yield a non-zero acceptance rate as $\Delta\tau \to 0$ (for a rigorous treatment of the limiting forms, we refer the reader to Delyon and Hu (2006), Stramer and Yan (2007) and to Papaspiliopoulos and Roberts (2012) for a recent discussion). However, it should also be noted that the linear

drift function governing (9) is independent of the drift function $\alpha(\cdot)$ governing the target process. Consequently, in situations where realisations of the target SDE (with the same initial condition) exhibit strong and similar nonlinearity over the inter-observation time, the modified diffusion bridge is likely to be unsatisfactory.

Residual bridge

To allow explicitly for dynamics based on the drift, we partition X_t into two parts, one that accounts for the drift in a deterministic way, and another as a residual stochastic process. The modified diffusion bridge is then applied to the *residual stochastic process* rather than the target process itself. The partition we require is

$$X_t = \eta_t + R_t, \tag{10}$$

where $\{\eta_t, t \in [t_j, t_{j+1}]\}$ is a deterministic process satisfying the ODE

$$\frac{d\eta_t}{dt} = \alpha(\eta_t), \quad \eta_{t_j} = x_{t_j}, \tag{11}$$

and $\{R_t, t \in [t_j, t_{j+1}]\}$ is a residual stochastic process satisfying

$$dR_t \equiv dX_t - d\eta_t = \{\alpha(X_t) - \alpha(\eta_t)\}dt + \sqrt{\beta(X_t)} \, dW_t. \tag{12}$$

We note that the partition in (10) is used by Fearnhead et al. (2014) (see also Section 4) to derive a tractable approximation to the intractable transition densities governing X_t, whereas our primary motivation for (10) is the application of the modified diffusion bridge construct 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. Therefore, we aim to derive an approximation $\widehat{\pi}(r_{\tau_{j,k}}|r_{\tau_{j,k-1}}, y_{t_{j+1}})$, that can be sampled recursively for $k = 1, \ldots, m$ and combined with the deterministic process (through numerical solution of (11)) via (10) to give a bridge $x_{\tau_{j,0}}, \ldots, x_{\tau_{j,m}}$. The scheme is illustrated in Figure 2.

The initial condition $\eta_{t_j} = x_{t_j}$ together with the Gaussian measurement error process imply that $r_{t_j} = 0$ and
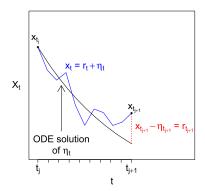
$$Y_{t_{j+1}} - F^T\eta_{t_{j+1}} = F^T R_{t_{j+1}} + \epsilon_{t_{j+1}}, \qquad \epsilon_{t_{j+1}}|\Sigma \overset{indep}{\sim} N(0, \Sigma).$$

Hence, it should be clear that the joint distribution of $R_{\tau_{j,k}}$ and $Y_{t_{j+1}} - F^T\eta_{t_{j+1}}$ conditional on $R_{\tau_{j,k-1}} = r_{\tau_{j,k-1}}$ can be approximated as

$$\begin{pmatrix} R_{\tau_{j,k}} \\ Y_{t_{j+1}} - F^T\eta_{t_{j+1}} \end{pmatrix} \bigg| r_{\tau_{j,k-1}} \sim N \left\{ \begin{pmatrix} r_{\tau_{j,k-1}} + (\alpha_{j,k-1} - \alpha^\eta_{j,k-1})\Delta\tau \\ F^T r_{\tau_{j,k-1}} + F^T(\alpha_{j,k-1} - \alpha^\eta_{j,k-1})\Delta^- \end{pmatrix}, \right.$$
$$\left. \begin{pmatrix} \beta_{j,k-1}\Delta\tau & \beta_{j,k-1}F\Delta\tau \\ F^T\beta_{j,k-1}\Delta\tau & F^T\beta_{j,k-1}F\Delta^- + \Sigma \end{pmatrix} \right) \right\}$$

where $\alpha(\eta_{\tau_{j,k-1}}) = \alpha^\eta_{j,k-1}$. Conditioning further on $y_{t_{j+1}} - F^T\eta_{t_{j+1}}$ gives

$$\widehat{\pi}(r_{\tau_{j,k}}|r_{\tau_{j,k-1}}, y_{t_{j+1}}) = N(r_{\tau_{j,k}}; \mu_{j,k}, \Psi_{j,k}), \tag{13}$$
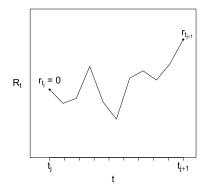
Figure 2: An illustration of the improved bridge construct. Left: The full bridge. Right: A sample path of R_t.

where

$$\mu_{j,k} = r_{\tau_{j,k-1}} + (\alpha^x_{j,k-1} - \alpha^\eta_{j,k-1})\Delta\tau + \beta_{j,k-1}F\Delta\tau(F^T\beta_{j,k-1}F\Delta^- + \Sigma)^{-1} \\ \times (y_{t_{j+1}} - F^T\eta_{t_{j+1}} - \{F^T r_{\tau_{j,k-1}} + F^T(\alpha^x_{j,k-1} - \alpha^\eta_{j,k-1})\Delta^-\}) \tag{14}$$

and

$$\Psi_{j,k} = \beta_{j,k-1}\Delta\tau - \beta_{j,k-1}F\Delta\tau(F^T\beta_{j,k-1}F\Delta^- + \Sigma)^{-1}F^T\beta_{j,k-1}\Delta\tau. \tag{15}$$

Together (13)–(15) define our bridge construct. These can be used to define the proposal mechanism in (7) for generating $\{X_t, t \in [t_j, t_{j+2}]\}$ by taking

$$q_1(x_{(j,j+1]}|x_{t_j}, y_{t_{j+1}}) = \prod_{k=1}^m \widehat{\pi}(x_{\tau_{j,k}} - \eta_{\tau_{j,k}}|r_{\tau_{j,k-1}}, y_{t_{j+1}})$$

and

$$q_2(x_{(j+1,j+2]}|x_{t_{j+1}}, x_{t_{j+2}}) = \prod_{k=1}^{m-1} \widehat{\pi}(x_{\tau_{j+1,k}} - \eta_{\tau_{j+1,k}}|r_{\tau_{j+1,k-1}}, x_{t_{j+2}}),$$

where $\widehat{\pi}(x_{\tau_{j+1,k}} - \eta_{\tau_{j+1,k}}|r_{\tau_{j+1,k-1}}, x_{t_{j+2}})$ can be sampled using (10) and (13)–(15) with j replaced by $j+1$, $\Sigma = 0$ and $F = I_d$.

In the special case of no measurement error and observation of all components we have that

$$\widehat{\pi}(r_{\tau_{j,k}}|r_{\tau_{j,k-1}}, x_{t_{j+1}}) = N\left(r_{\tau_{j,k}}; r_{\tau_{j,k-1}} + \frac{r_{t_{j+1}} - r_{\tau_{j,k-1}}}{t_{j+1} - \tau_{j,k-1}}\Delta\tau, \frac{t_{j+1} - \tau_{j,k}}{t_{j+1} - \tau_{j,k-1}}\beta_{j,k-1}\Delta\tau\right),$$

which can be seen as a linear approximation of the Brownian bridge SDE

$$dR_t = \frac{R_{t_{j+1}} - R_t}{t_{j+1} - t} + \sqrt{\beta(X_t)}\,dW_t. \tag{16}$$

We also note that (16) has the same diffusion coefficient as the target process and appeal again to Delyon and Hu (2006), to deduce that the distribution of the residual process governed by (12) (conditional on $r_{t_{j+1}}$) is absolutely continuous with respect to the distribution of the solution to (9).

3.3 Parameter updates

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. For the remaining parameters θ and $b = (b^1, \ldots, b^N)^T$ we have

$$\pi(\theta|b,x) \propto \pi(\theta)\pi(x|\theta,b) \qquad \text{and} \qquad \pi(b|\theta,\psi,x) \propto \pi(b|\psi)\pi(x|\theta,b) = \prod_{i=1}^{N} \pi(b^i|\psi)\pi(x^i|\theta,b^i)$$

where the last expression suggests unit-specific updates of the components of b.

As discussed earlier, since θ and the components of b enter into the diffusion coefficient of (1), sampling the full conditionals of $\theta|b,x$ and $b|\theta,\psi,x$ as part of a Gibbs sampler will result in a reducible Markov chain as $m \to \infty$ (or $\Delta\tau \to 0$). To overcome this problem we use a reparameterisation which is outlined in the next section.

Modified innovation scheme

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 there is a one-to-one relationship between the increments of the process (ΔX_t) and the increments of the driving Brownian motion (ΔW_t). 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*. Conditioning on the Brownian increment innovations in a Gibbs update should therefore 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)) 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 their innovation process 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

$$dZ_t^i = \beta(X_t^i, \theta, b^i)^{-1/2} \left(dX_t^i - \frac{x_{t_{j+1}}^i - X_t^i}{t_{j+1} - t}\, dt \right), \tag{17}$$

$$= \beta(X_t^i, \theta, b^i)^{-1/2} \left\{ \alpha(X_t^i, \theta, b^i) - \frac{x_{t_{j+1}}^i - X_t^i}{t_{j+1} - t} \right\} dt + dW_t^i$$

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 (17) between observation times. We therefore define $x_o^i = (x_{t_0}^i, \ldots, x_{t_n}^i)^T$ 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. We have for $k = 1, \ldots, m$

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

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

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

Note that our discretisation of (17) follows Golightly and Wilkinson (2008) by using the modified diffusion bridge to construct the innovation process. Now define a function f so that $X_{\tau_{j,k}}^i = f(Z_{\tau_{j,k}}^i, \theta, b^i)$ and $Z_{\tau_{j,k}}^i = f^{-1}(X_{\tau_{j,k}}^i, \theta, b^i)$. Let z_{imp}^i denote the (unit-specific) innovation values over $[t_0, t_n]$ and stack all z_{imp}^i values into the matrix z_{imp}. Define x_{imp}^i and x_{imp} similarly. The modified innovation scheme samples $\theta|b, z_{imp}, x_o$ and $b^i|\theta, \psi, z_{imp}^i, x_o^i$, $i = 1, \ldots, N$. Note that for an updated value of b^i, say b^{i*}, a new x_{imp}^{i*} is updated deterministically through $x_{imp}^{i*} = f(z_{imp}^{i*}, \theta, b^{i*})$. Likewise, for a new θ^*, a new x_{imp}^* is updated deterministically through $x_{imp}^{i*} = f(z_{imp}^{i*}, \theta^*, b^i)$, $i = 1, \ldots, N$. The full conditional density of θ is

$$\pi(\theta|b, z_{imp}, x_o) \propto \pi(\theta) \prod_{i=1}^{N} \prod_{j=1}^{n-1} \left[\prod_{k=1}^{m} \pi(x_{\tau_{j,k}}^i | x_{\tau_{j,k-1}}^i, \theta, b^i) \prod_{k=1}^{m-1} J\{f(z_{\tau_{j,k}}^i, \theta, b^i)\} \right], \quad (18)$$

where

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

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

$$\pi(b^i|\theta, \psi, z_{imp}^i, x_o^i) \propto \pi(b^i|\psi) \prod_{j=1}^{n-1} \left[\prod_{k=1}^{m} \pi(x_{\tau_{j,k}}^i | x_{\tau_{j,k-1}}^i, \theta, b^i) \prod_{k=1}^{m-1} J\{f(z_{\tau_{j,k}}^i, \theta, b^i)\} \right]. \quad (19)$$

Naturally, the full conditionals in (18) and (19) will typically be intractable, requiring the use of Metropolis-within-Gibbs updates.

4   Linear noise approximation

In this section we outline a competing solution which uses an inference scheme based on a linear noise approximation (LNA) to the SDMEM. The LNA typically refers to an approximation to the solution of the forward Kolmogorov equation governing the transition probability of a Markov jump process (Kurtz, 1970; Ferm et al., 2008; Komorowski et al., 2009; Finkenstädt et al., 2013). Specifically, the forward Kolmogorov equation is approximated by a Fokker–Planck equation with linear coefficients. Equivalently, a general Fokker–Planck equation can be deduced and then linearised. In this context, therefore, the LNA aims to replace intractable transition densities with Gaussian approximations. In what follows, we give a brief informal derivation of the LNA and refer the reader to Fearnhead et al. (2014) and the references therein for further details.

4.1   Setup

For notational simplicity and clarity of exposition, we suppress parameter dependence and the unit-specific i for the remainder of this sub-section.

Without loss of generality, consider a time $t \in [t_j, t_{j+1}]$ at which we wish to approximate the intractable transition density associated with $X_t | X_{t_j} = x_{t_j}$. The LNA uses the same partition of X_t given in (10), that is $X_t = \eta_t + R_t$ where the deterministic process η_t satisfies (11) and the residual stochastic process satisfies (12). The key assumption underpinning the LNA is that the residual stochastic perturbation is "small" relative to the deterministic process, allowing suitable truncation of a Taylor series expansion of $\alpha(X_t)$ and $\beta(X_t)$ about η_t. Taking the first two terms in the expansion of $\alpha(X_t)$, and the first term in the expansion of $\beta(X_t)$ gives an SDE satisfied by an approximate residual process $\{\tilde{R}_t, t \in [t_j, t_{j+1}]\}$ of the form

$$d\tilde{R}_t = H_t \tilde{R}_t \, dt + \sqrt{\beta(\eta_t)} \, dW_t, \tag{20}$$

where H_t is the Jacobian matrix with (i, j)th element $(H_t)_{i,j} = \partial \alpha_i(\eta_t)/\partial \eta_{j,t}$.

Assuming fixed or Gaussian initial conditions $\tilde{R}_{t_j} \sim N(m_{t_j}, V_{t_j})$ gives $\tilde{R}_t \sim N(m_t, V_t)$, where m_t and V_t satisfy the ODE system

$$\frac{dm_t}{dt} = H_t m_t, \tag{21}$$

$$\frac{dV_t}{dt} = H_t V_t + \beta(\eta_t, \theta, b) + V_t H_t^T. \tag{22}$$

In the absence of an analytic solution, the system of coupled ODEs (11) and (21)–(22) which characterise the LNA, must be solved numerically. For initial conditions $\eta_{t_j} = x_{t_j}$, we have $m_{t_j} = 0$ and $V_{t_j} = 0$ so that (21) does not need to be solved, and the approximating transition distribution is $X_t | X_{t_j} = x_{t_j} \sim N(\eta_t, V_t)$.

It is worth noting here that the linear form of the SDE (20) satisfied by the approximate residual process coupled with the additive Gaussian observation regime admits a closed form expression for densities of the form $\widehat{\pi}(\tilde{r}_{\tau_{j,k}} | \tilde{r}_{\tau_{j,k-1}}, y_{t_{j+1}})$, suggesting use of

the LNA as a proposal mechanism inside the Bayesian imputation approach of Section 3. Whilst the LNA could in principle be used to directly approximate the conditioned residual process governed by the SDE in (12) we note that the SDEs in (12) and (20) have different diffusion coefficients. Consequently, the probability law governing $\tilde{R}_t$ is not absolutely continuous with respect to the law of R_t. We therefore do not advocate use of the LNA in this way.

In the next section we outline an inference scheme for SDMEMs of the form (1) based on the LNA. It exploits the computational efficiency of a filtering algorithm proposed by Fearnhead et al. (2014) that allows closed-form calculation of the marginal likelihood $\pi(y|\theta, b, \Sigma)$ under our Gaussian observation regime (2); see the supplementary material for further details (Whitaker et al., 2016).

4.2   Application to SDMEMs

Under the linear noise approximation of (1) the marginal posterior for all parameters is given by

$$\pi(\theta, \psi, \Sigma, b|y) \propto \pi(\theta)\pi(\psi)\pi(\Sigma)\pi(b|\psi)\pi(y|\theta, \Sigma, b)$$

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

This factorisation suggests a Gibbs sampler with blocking that sequentially takes draws from the full conditionals $\pi(\Sigma|\theta, \psi, b, y) = \pi(\Sigma|y)$, $\pi(\theta|\psi, \Sigma, b, y) = \pi(\theta|b, y)$, $\pi(b|\theta, \psi, \Sigma, y) = \pi(b|\theta, \psi, y)$ and $\pi(\psi|\theta, \Sigma, b, y) = \pi(\psi|b)$. A Metropolis–Hastings step can be used when a full conditional density is intractable. An algorithm for computing the marginal likelihood $\pi(y^i|\theta, \Sigma, b^i)$ for each experimental unit is given in the supplementary material. Interest may also lie in the joint posterior $\pi(\theta, \psi, \Sigma, b, x|y)$ where, since no imputation is required for the LNA, $x^i = (x_{t_0}, \ldots, x_{t_n})^T$ 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 $i = 1, \ldots, N$. An efficient mechanism for making such draws can also be found in the supplementary material. The method uses a forward filter, backward sampling (FFBS) algorithm.

5   Applications

We now compare the accuracy and efficiency of our Bayesian imputation approach (coupled with the modified innovation scheme) with an LNA-based solution. We consider two scenarios: one in which the ODEs governing the LNA are tractable and one in which numerical solvers are required. In the first we use synthetic data generated from a simple univariate SDE description of orange tree growth. 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. 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). We also include (in

the supplementary material) a simulation study based on synthetic data generated from the model of aphid dynamics, to explore further any differences between the Bayesian imputation and LNA-based approaches.

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) of individual trees (mm) by

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

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 $b^i = (\phi_1^i, \phi_2^i)^T$, $i = 1, \ldots, N$ and the parameter vector governing the random effects distributions is $\psi = (\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2})^T$. Note that the ϕ_1^i can be interpreted as asymptotic circumferences and the ϕ_2^i as the time-distance 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 $\phi_2^i \sim N(350, 52.5^2)$. 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

$$\pi(\theta, \psi, b|x) \propto \pi(\theta)\pi(\psi)\pi(b|\psi)\pi(x|\theta, b).$$

In the Bayesian imputation approach, $\pi(x|\theta, b)$ is as in (5) whereas for the LNA–based solution

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

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

$$\frac{d\eta_t^i}{dt} = \frac{1}{\phi_1^i \phi_2^i} \eta_t^i (\phi_1^i - \eta_t^i), \qquad \eta_{t_j}^i = x_{t_j}^i,$$

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

Fortunately this ODE system can be solved analytically giving $\eta_t^i = A\phi_1^i e^{t/\phi_2^i}/(1 + Ae^{t/\phi_2^i})$ and

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

where $A = x_{t_0}^i/(\phi_1^i - x_{t_0}^i)$ and $B = \sigma^2 A\phi_1^i e^{2t/\phi_2^i}/(1 + Ae^{t/\phi_2^i})^4$.

The MCMC scheme can make use of simple semi-conjugate updates for ϕ_1, ϕ_2, σ_{ϕ_1} and σ_{ϕ_2}. 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 to update the latent trajectories between observation times: the improved bridge construct of Section 3.2 is not needed.

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 sample output was also used to estimate the marginal posterior variances of σ and the b^i, to provide sensible innovation variances in the random walk Metropolis updates. 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 3 shows the marginal posterior densities and autocorrelations for the common parameter σ and the parameters governing the random effects distributions. The marginal posterior means and standard deviations of $(\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2}, \sigma)$ are given in Table 1. The figures and table show that for these parameters 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. Similar results are obtained for the random effects parameters (see the supplementary material).

Both schemes were coded in C and run on an Intel Xeon 3.0 GHz processor; the modified innovation scheme took 43504 seconds to run whilst the LNA inference scheme took 2483 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 and the LNA–based approach gave 7821. 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 numerical schemes will degrade its performance.

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

	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 1: Marginal posterior means (standard deviations) of the random effects hyperparameters (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 $\sigma = 0.08$.
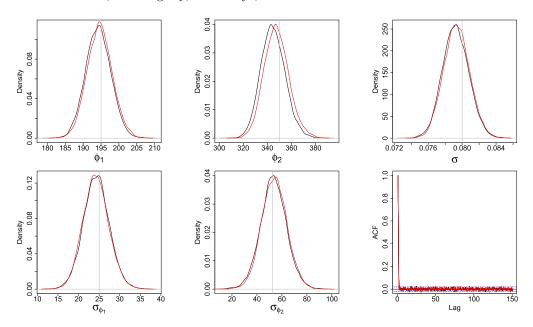
Figure 3: Marginal posterior densities for the random effects hyper-parameters (ϕ_1, ϕ_2, σ_{ϕ_1}, σ_{ϕ_2}) and common parameter σ in the orange tree growth SDMEM, together with their (overlayed) autocorrelation functions. Black: Bayesian imputation. Red: LNA. The vertical grey lines indicate the ground truth.

5.2 Cotton aphid dynamics

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

$$\Pr(N_{t+dt} = n_t + 1, C_{t+dt} = c_t + 1 \,|\, n_t, c_t) = \lambda n_t \, dt + o(dt),$$
$$\Pr(N_{t+dt} = n_t - 1, C_{t+dt} = c_t \,|\, n_t, c_t) = \mu n_t c_t \, dt + o(dt). \tag{23}$$

The diffusion approximation of the Markov jump process defined by (23) is

$$\begin{pmatrix} dN_t \\ dC_t \end{pmatrix} = \begin{pmatrix} \lambda N_t - \mu N_t C_t \\ \lambda N_t \end{pmatrix} dt + \begin{pmatrix} \lambda N_t + \mu N_t C_t & \lambda N_t \\ \lambda N_t & \lambda N_t \end{pmatrix}^{1/2} dW_t. \tag{24}$$

Matis et al. (2008) also provide a dataset of cotton aphid counts collected from three blocks (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).

We now formulate an appropriate SDMEM model driven by (24) 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} the corresponding cumulative population size. We write $X_t^{ijk} = (N_t^{ijk}, C_t^{ijk})^T$ and consider the SDMEM

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

where

$$\alpha(X_t^{ijk}, b^{ijk}) = \begin{pmatrix} \lambda^{ijk} N_t^{ijk} - \mu^{ijk} N_t^{ijk} C_t^{ijk} \\ \lambda^{ijk} N_t^{ijk} \end{pmatrix},$$

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

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

$$\begin{aligned} \lambda^{ijk} &= \lambda + \lambda_{W_i} + \lambda_{N_j} + \lambda_{B_k} + \lambda_{WN_{ij}} + \lambda_{WB_{ik}} + \lambda_{NB_{jk}} \\ \mu^{ijk} &= \mu + \mu_{W_i} + \mu_{N_j} + \mu_{B_k} + \mu_{WN_{ij}} + \mu_{WB_{ik}} + \mu_{NB_{jk}}. \end{aligned} \quad (25)$$

Also for identifiability we use the corner constraints $\lambda_{W_1} = \lambda_{N_1} = \lambda_{B_1} = 0$, $\lambda_{WN_{ij}} = \lambda_{WN_{ij}}(1 - \kappa_{ij})$, $\lambda_{WB_{ik}} = \lambda_{WB_{ik}}(1 - \kappa_{ik})$ and $\lambda_{NB_{jk}} = \lambda_{NB_{jk}}(1 - \kappa_{jk})$, where $\kappa_{rs} = \max(\delta_{1r}, \delta_{1s})$ and $\delta_{1\cdot}$ is the Kronecker delta, with equivalent constraints on the death rates. The interpretation of (25) 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 further differs 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

$$Y_t^{ijk} | N_t^{ijk}, \sigma \overset{indep}{\sim} N(N_t^{ijk}, \sigma^2 N_t^{ijk}), \quad t = 0, 1.14, 2.29, 3.57, 4.57. \quad (26)$$

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 (25) 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. Note that the form of the prior for σ gives a semi-conjugate update. The remaining parameters in (25) are updated using random walk Metropolis on the pairwise λ, μ component blocks $(\lambda, \mu), (\lambda_{W_2}, \mu_{W_2}), (\lambda_{W_3}, \mu_{W_3}), \ldots, (\lambda_{NB_{33}}, \mu_{NB_{33}})$.

The nonlinear form of the observation model (26) 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 in Section 3.2, we replace Σ in (14) and (15) with $\sigma^2 \eta_{N,t_{j+1}}$, where $\eta_{t_{j+1}} = (\eta_{N,t_{j+1}}, \eta_{C,t_{j+1}})^T$ is the solution of (11). 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) and our improved bridge construct of Section 3.2 in a scenario typical of the real dataset. Using the simulation study of Gillespie and Golightly (2010), we take $(\lambda, \mu)^T = (1.75, 0.00095)^T$, $x_0 = (28, 28)^T$ and recursively apply the Euler–Maruyama approximation to give $x_{3.57} = (829.08, 1406.07)^T$. 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 (26) with $\sigma = 1$. Figure 4 shows 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 clear 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 acceptance rate of around 58% for the improved 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 an ODE system given by (11) and (22) where the Jacobian matrix is

$$H_t = \begin{pmatrix} \lambda - \mu \eta_{C,t} & -\mu \eta_{N,t} \\ \lambda & 0 \end{pmatrix}.$$
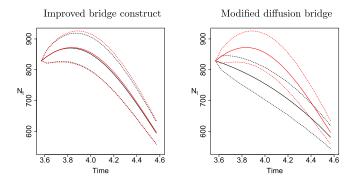
Figure 4: 95% credible region (dashed line) and mean (solid line) of the true conditioned aphid population component $N_t | x_{3.57}, y_{4.57}$ (red) and two competing bridge constructs (black).

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 the FFBS algorithm in the supplementary material, we make an approximation to the marginal likelihood calculation by replacing Σ with $\sigma^2 \eta_{N, t_{j+1}}$.

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 (25), 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 shows the marginal posterior densities of the baseline parameters, the parameter σ controlling the observation error variance and a selection of the remaining parameters. 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. For example, 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.
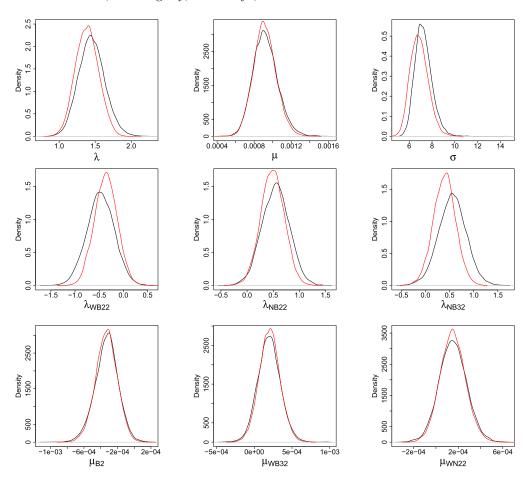
Figure 5: Marginal posterior densities for a selection of the aphid model parameters. Black: Bayesian imputation. Red: LNA.

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 (26). 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

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

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-

ment combination. This is estimated by averaging realisations of N_t (obtained by applying the Euler–Maruyama approximation to (24)) over draws from the marginal posterior $\pi(n_0^{ijk}, b^{ijk} | y)$ obtained using either Bayesian imputation or the LNA. Figures 6 and 7 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 7) suggesting that in some situations, using the inferences made under the 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 (25). Figure 8 shows marginal posterior densities of the overall birth rates (λ^{ijk}) associated with the six treatment-block combinations for which predictives are presented in Figure 7. 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 1039 under the modified innovation scheme. The LNA, however, clearly benefits from analytically integrating out the latent process and gave a minESS of 8908. 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 3.0 GHz 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 demonstrated by the simulation study in the supplementary material.
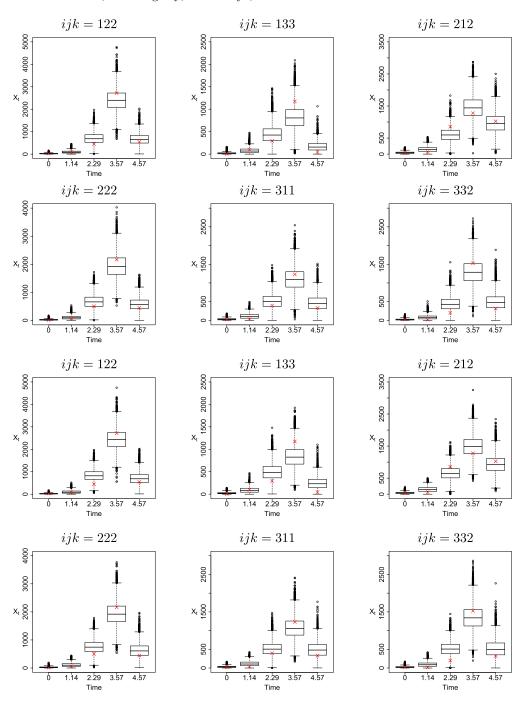
Figure 6: Within sample predictive distributions for Bayesian imputation (top 2 rows) and LNA (bottom 2 rows). The red crosses indicate the observed values.
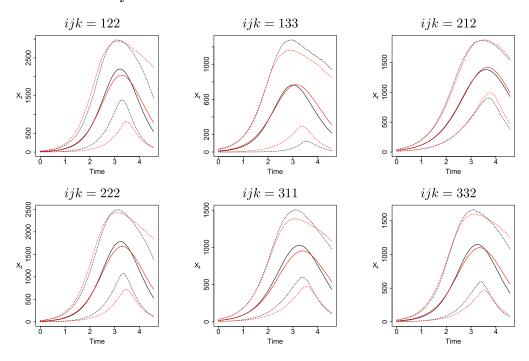
Figure 7: Out-of-sample predictive intervals for the aphid population size (N_t^{ijk}) against time for a random selection of treatment combinations. The mean is depicted by the solid line with the dashed representing a 95% credible region. Black: Bayesian imputation. Red: LNA.

6   Discussion

We have provided a framework that permits (simulation-based) Bayesian inference for a large class of multivariate SDMEMs using discrete-time observations that may be incomplete and subject to measurement error. 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. Fundamental to our approach is the development of a novel bridge construct that can be used to sample a discretisation of a conditioned diffusion process, and does not break down when the process exhibits strong nonlinearity over inter-observation times of interest. 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 here 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 using two applications. The computational efficiency of the LNA depends on the dimension of the SDE driving the
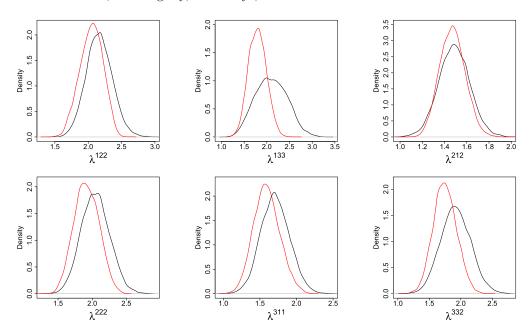
Figure 8: Marginal posterior densities for a random selection of the birth rates associated with specific treatment combinations in the aphid model. Black: Bayesian imputation. Red: LNA.

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 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 is 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 is 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 (given in the supplementary material) highlighted further differences between the LNA and Bayesian imputation approaches. Care must therefore be taken in trying to fit the SDMEM by using an LNA–based inference approach.

Supplementary Material

Supplementary material for Bayesian inference for diffusion-driven mixed-effects models
(DOI: 10.1214/16-BA1009SUPP; .pdf).

References

Aït-Sahalia, Y. (2002). "Maximum likelihood estimation of discretely sampled diffusions:
a closed-form approximation approach." *Econometrica*, 70(1): 223–262. MR1926260.
doi: http://dx.doi.org/10.1111/1468-0262.00274. 435, 436

Aït-Sahalia, Y. (2008). "Closed-form likelihood expansions for multivariate diffusions."
The Annals of Statistics, 36(2): 906–937. MR2396819. doi: http://dx.doi.org/
10.1214/009053607000000622. 435, 436

Berglund, M., Sunnåker, M., Adiels, M., Jirstrand, M., and Wennberg, B. (2011). "In-
vestigations of a compartmental model for leucine kinetics using non-linear mixed
effects models with ordinary and stochastic differential equations." *Mathemati-
cal Medicine and Biology*, 29(4): 361–384. MR3000271. doi: http://dx.doi.org/
10.1093/imammb/dqr021. 436

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. MR3016228. doi: http://dx.doi.org/10.1016/j.spa.2012.12.001.
441

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 (Sta-
tistical Methodology)*, 68: 333–382. MR2278331. doi: http://dx.doi.org/10.1111/
j.1467-9868.2006.00552.x. 435

Chib, S., Pitt, M. K., and Shephard, N. (2004). "Likelihood based inference for diffusion
driven models." Technical Report 2004-W20, Nuffield College, Oxford. 445

Clark, J. (1990). "The simulation of pinned diffusions." In *Proceedings of the 29th IEEE
Conference on Decision and Control, 1990*, 1418–1420. IEEE. 441

Delyon, B. and Hu, Y. (2006). "Simulation of conditioned diffusion and application to
parameter estimation." *Stochastic Processes and Their Applications*, 116: 1660–1675.
MR2269221. doi: http://dx.doi.org/10.1016/j.spa.2006.04.004. 441, 442, 445

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. MR2758209. doi: http://dx.doi.org/10.1111/j.1541-0420.2009.01342.x.
436

Durham, G. B. and Gallant, A. R. (2002). "Numerical techniques for maximum like-
lihood estimation of continuous-time diffusion processes." *Journal of Business and
Economic Statistics*, 20: 297–338. MR1939904. doi: http://dx.doi.org/10.1198/
073500102288618397. 436, 437, 441, 442, 445, 453

Elerian, O., Chib, S., and Shephard, N. (2001). "Likelihood inference for discretely observed nonlinear diffusions." *Econometrica*, 69: 959–993. MR1839375. doi: http://dx.doi.org/10.1111/1468-0262.00226. 436

Eraker, B. (2001). "MCMC analysis of diffusion models with application to finance." *Journal of Business and Economic Statistics*, 19(2): 177–191. MR1939708. doi: http://dx.doi.org/10.1198/073500101316970403. 436

Fearnhead, P., Giagos, V., and Sherlock, C. (2014). "Inference for reaction networks using the linear noise approximation." *Biometrics*, 70(2): 457–466. MR3258050. doi: http://dx.doi.org/10.1111/biom.12152. 443, 447, 448

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. MR2373035. doi: http://dx.doi.org/10.1007/s10915-007-9179-z. 447

Finkenstädt, B., Woodcock, D., Komorowski, M., Harper, C., Davis, J., White, M., 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. MR3161709. doi: http://dx.doi.org/10.1214/13-AOAS669. 447

Fuchs, C. (2013). *Inference for diffusion processes with applications in Life Sciences*. Heidelberg: Springer. MR3015023. doi: http://dx.doi.org/10.1007/978-3-642-25969-2. 437, 445, 446

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. MR2744477. doi: http://dx.doi.org/10.1111/j.1467-9876.2009.00696.x. 452, 453, 454

Golightly, A. and Wilkinson, D. J. (2006). "Bayesian sequential inference for nonlinear multivariate diffusions." *Statistics and Computing*, 16: 323–338. MR2297534. doi: http://dx.doi.org/10.1007/s11222-006-9392-x. 440

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. MR2422763. doi: http://dx.doi.org/10.1016/j.csda.2007.05.019. 436, 437, 440, 441, 445, 446, 458

Golightly, A. and Wilkinson, D. J. (2010). "Markov chain Monte Carlo algorithms for SDE parameter estimation." In Lawrence, N. D., Girolami, M., Rattray, M., and Sanguinetti, G. (eds.), *Learning and inference in Computational Systems Biology*, Computational Molecular Biology. MIT Press. 437, 445

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. 440

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.   436

Komorowski, M., Finkenstädt, B., Harper, C., and Rand, D. (2009). "Bayesian inference of biochemical kinetic parameters using the linear noise approximation." *BMC Bioinformatics*, 10(1): 343.   447

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.   MR3036416. doi: http://dx.doi.org/10.1080/01621459.2012.720899.   436

Kurtz, T. G. (1970). "Solutions of ordinary differential equations as limits of pure jump Markov processes." *Journal of Applied Probability*, 7(1): 49–58. MR0254917.   447

Lindström, E. (2012). "A regularized bridge sampler for sparsely sampled diffusions." *Statistics and Computing*, 22(1): 615–623.   MR2865039. doi: http://dx.doi.org/10.1007/s11222-011-9255-y.   441

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.   451

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.   437, 448, 451, 459

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.   436

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*, 311–337. Chapman and Hall.   MR2976985. doi: http://dx.doi.org/10.1201/b12126-5.   441, 442

Papaspiliopoulos, O., Roberts, G. O., and Stramer, O. (2013). "Data augmentation for diffusions." *Journal of Computational and Graphical Statistics*, 22: 665–688. MR3173736. doi: http://dx.doi.org/10.1080/10618600.2013.783484.   437

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. MR1334067.   436, 440

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.   MR2741425. doi: http://dx.doi.org/10.1016/j.csda.2010.10.003.   436, 437, 449

Picchini, U., Gaetano, S., and Ditlevsen, S. (2010). "Stochastic differential mixed-effects models." *Scandinavian Journal of Statistics*, 37: 67–90.   MR2675940. doi: http://dx.doi.org/10.1111/j.1467-9469.2009.00665.x.   436, 437, 449

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. MR1859397. doi: http://dx.doi.org/10.1093/biomet/88.3.603. 436, 440

Schauer, M., van der Meulen, F., and van Zanten, H. (2014). "Guided proposals for simulating multi-dimensional diffusion bridges." Available from http://arxiv.org/abs/1311.3606. 441

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. MR3066416. doi: http://dx.doi.org/10.1111/j.1467-9469.2012.00812.x. 435

Stramer, O. and Bognar, M. (2011). "Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach." *Bayesian Analysis*, 6(2): 231–258. MR2806243. doi: http://dx.doi.org/10.1214/11-BA608. 436, 437, 440

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. 435

Stramer, O. and Yan, J. (2007). "Asymptotics of an efficient Monte Carlo estimation for the transition density of diffusion processes." *Methodology and Computing in Applied Probability*, 9(4): 483–496. MR2404739. doi: http://dx.doi.org/10.1007/s11009-006-9006-2. 442

Tornøe, C. W., Overgaard, R. V., Agerso, H., Nielsen, H. A., Madsen, H., and Jonsson, E. N. (2005). "Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations." *Pharmaceutical Research*, 22(8): 1247–1258. 436

Whitaker G. A., Golightly A., Boys R. J., and Sherlock C. (2016). Supplementary material for "Bayesian inference for diffusion-driven mixed-effects models". *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/16-BA1009SUPP. 448

Acknowledgments