

Improved bridge constructs for stochastic differential equations

Gavin A. Whitaker¹ · Andrew Golightly¹ · Richard J. Boys¹ · Chris Sherlock²

Received: 30 September 2015 / Accepted: 21 April 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract We consider the task of generating discrete-time realisations of a nonlinear multivariate diffusion process satisfying an Itô stochastic differential equation conditional on an observation taken at a fixed future time-point. Such realisations are typically termed diffusion bridges. Since, in general, no closed form expression exists for the transition densities of the process of interest, a widely adopted solution works with the Euler–Maruyama approximation, by replacing the intractable transition densities with Gaussian approximations. However, the density of the conditioned discrete-time process remains intractable, necessitating the use of computationally intensive methods such as Markov chain Monte Carlo. Designing an efficient proposal mechanism which can be applied to a noisy and partially observed system that exhibits nonlinear dynamics is a challenging problem, and is the focus of this paper. By partitioning the process 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.

Keywords Stochastic differential equation · Multivariate diffusion bridge · Guided proposal · Markov chain Monte Carlo · Linear noise approximation

✉ Andrew Golightly
andrew.golightly@ncl.ac.uk

¹ School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

² Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK

1 Introduction

Diffusion processes satisfying stochastic differential equations (SDEs) provide a flexible class of models for describing many continuous-time physical processes. Some application areas and indicative references include finance, e.g. Kalogeropoulos et al. (2010), Stramer et al. (2010), reaction networks, e.g. Fuchs (2013), Golightly et al. (2015) and population dynamics, e.g. Heydari et al. (2014). Fitting such models to data observed at discrete-times can be problematic since the transition densities of the diffusion process are likely to be intractable. A review of inferential methods for diffusions can be found in Fuchs (2013). A widely adopted solution is to approximate the unavailable transition densities either analytically (Aït-Sahalia 2002, 2008) or numerically (Pedersen 1995; Elerian et al. 2001; Eraker 2001; Roberts and Stramer 2001). Within the Bayesian paradigm, the numerical approach can be seen as a data augmentation problem. The simplest implementation augments 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 data. The key challenges of designing such an MCMC scheme include overcoming dependence between the parameters and missing data (first highlighted as a problem by Roberts and Stramer (2001)) and overcoming dependence between successive values of the missing data. Dealing with the latter requires repeatedly generating realisations known as *diffusion bridges* from an approximation of the conditioned process. Methods built upon exact simulation, that avoid use of the Euler–Maruyama approximation and the associated discretisation error, have been proposed

by Beskos et al. (2006) (see also Beskos et al. 2009). However, these exact methods are limited to diffusions which can be transformed to have unit diffusion coefficient, known as reducible diffusions.

Designing bridge constructs for irreducible, multivariate diffusions is a challenging problem and has received much attention in recent literature. The simplest approach (see e.g. Pedersen 1995) is based on the forward dynamics of the diffusion process and generates a bridge by sampling iteratively from the Euler–Maruyama approximation of the unconditioned SDE. This myopic approach induces a discontinuity at the observation time (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) (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 is less effective 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 Pedersen and MDB approaches, with a tuning parameter governing the precise dynamics of the resulting sampler. Del Moral and Murray (2014) (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.

1.1 Contributions and organisation of the paper

Our contribution is the development of a novel class of bridge constructs that are 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 MDB sampler of Durham and Gallant (2002) to the end-point conditioned residual process. We consider two implementations of this approach. Firstly, we use the bridge introduced by Whitaker et al. (2015) that constructs the residual process by subtracting the solution of an ordinary differential equation (ODE) 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 MDB sampler. In addition, we adapt the guided proposal proposed by Schauer et al. (2016) to a partial and noisy observation regime.

We evaluate the performance of each bridge construct (as well as the constructs proposed by Durham and Gallant (2002) and Lindström (2012)) using three examples: a simple birth–death model, a Lotka–Volterra system and a model of aphid growth.

The remainder of this article is organised as follows. Section 2 provides a brief introduction to the problem of sampling conditioned SDEs and examines two previously proposed approaches. In Sect. 3 we describe a novel class of bridge constructs and adapt an existing approach to a more general observation regime. Applications are considered in Sect. 4 and a discussion is provided in Sect. 5.

2 Sampling conditioned SDEs

Consider a continuous-time d -dimensional Itô process $\{X_t, t \geq 0\}$ governed by the SDE parameterised by $\theta = (\theta_1, \dots, \theta_p)'$ of the form

$$dX_t = \alpha(X_t, \theta) dt + \sqrt{\beta(X_t, \theta)} dW_t, \quad X_0 = x_0. \quad (1)$$

Here, α is a d -vector of drift functions, the diffusion matrix 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 of (uncorrelated) standard Brownian motion processes. We assume that α and β are sufficiently regular so that the SDE has a weak non-explosive solution (Øksendal 2003).

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

$$Y_T = F'X_T + \epsilon_T, \quad \epsilon_T | \Sigma \sim N(0, \Sigma). \quad (2)$$

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 for some $d_o \leq d$. This flexible setup allows for only observing a subset of components. 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. The complication of multiple partial and/or noisy observations is discussed in Sect. 5.

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

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

giving m intervals of equal length $\Delta\tau = T/m$. Since, in general, the form of the SDE in (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. Specifically, we have that

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

where $\Delta W_{\tau_k} \sim N(0, \Delta\tau I_d)$ and I_d is the $d \times d$ identity matrix. The continuous-time conditioned process is then approximated by the discrete-time skeleton bridge, with the latent values $x_{(0,T)} = (x_{\tau_1}, \dots, x_{\tau_{m-1}} = x_T)'$ having the (posterior) density

$$\pi(x_{(0,T)}|x_0, y_T, \theta, \Sigma) \propto \pi(y_T|x_T, \Sigma) \prod_{k=0}^{m-1} \pi(x_{\tau_{k+1}}|x_{\tau_k}, \theta) \tag{3}$$

where

$$\pi(x_{\tau_{k+1}}|x_{\tau_k}, \theta) = N(x_{\tau_{k+1}}; x_{\tau_k} + \alpha(x_{\tau_k}, \theta)\Delta\tau, \beta(x_{\tau_k}, \theta)\Delta\tau)$$

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 $x_{(0,T)} = (x_{\tau_1}, \dots, x_{\tau_{m-1}})'$ have the density

$$\pi(x_{(0,T)}|x_0, x_T, \theta) \propto \prod_{k=0}^{m-1} \pi(x_{\tau_{k+1}}|x_{\tau_k}, \theta). \tag{4}$$

For nonlinear forms of the drift and diffusion coefficients, the products in (3) and (4) will be intractable and samples can be generated via computationally intensive algorithms such as Markov chain Monte Carlo 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) as

$$\pi(x_{(0,T)}|x_0, y_T, \theta, \Sigma) \propto \prod_{k=0}^{m-1} \pi(x_{\tau_{k+1}}|x_{\tau_k}, y_T, \theta, \Sigma). \tag{5}$$

The density in (4) can be factorised in a similar manner. This suggests seeking proposal 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 (5). 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

$$q(x_{\tau_{k+1}}|x_{\tau_k}, y_T, \theta, \Sigma) = N(x_{\tau_{k+1}}; x_{\tau_k} + \mu(x_{\tau_k})\Delta\tau, \Psi(x_{\tau_k})\Delta\tau) \tag{6}$$

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

2.1 Myopic simulation

Ignoring the information in the observation y_T and simply applying the Euler–Maruyama approximation over each interval of length $\Delta\tau$ leads to a proposal density of the form given by (6) with $\mu_{EM}(x_{\tau_k}) = \alpha(x_{\tau_k})$ and $\Psi_{EM}(x_{\tau_k}) = \beta(x_{\tau_k})$. Sampling iteratively according to (6) for $k = 0, 1, \dots, m-1$ gives a proposed bridge which we denote by $x_{(0,T)}^*$. The Metropolis–Hastings (MH) acceptance probability for a move from $x_{(0,T)}$ to $x_{(0,T)}^*$ is

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

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 Σ is reduced, the MH 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

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

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

2.2 Modified diffusion bridge

For known x_T , Durham and Gallant (2002) 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). In brief, the joint distribution of $X_{\tau_{k+1}}$ and Y_T (conditional on x_{τ_k}) is approximated by

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

where $\alpha_k = \alpha(x_{\tau_k})$, $\beta_k = \beta(x_{\tau_k})$ and $\Delta_k = T - \tau_k$. Conditioning on $Y_T = y_T$ gives

$$\mu_{\text{MDB}}(x_{\tau_k}) = \alpha_k + \beta_k F (F' \beta_k F \Delta_k + \Sigma)^{-1} \times \{y_T - F'(x_{\tau_k} + \alpha_k \Delta_k)\} \quad (7)$$

and

$$\Psi_{\text{MDB}}(x_{\tau_k}) = \beta_k - \beta_k F (F' \beta_k F \Delta_k + \Sigma)^{-1} F' \beta_k \Delta \tau. \quad (8)$$

In the case of no measurement error and observation of all components (so that x_T is known), (7) and (8) become

$$\mu_{\text{MDB}}^*(x_{\tau_k}) = \frac{x_T - x_{\tau_k}}{T - \tau_k} \quad \text{and} \quad \Psi_{\text{MDB}}^*(x_{\tau_k}) = \frac{T - \tau_{k+1}}{T - \tau_k} \beta(x_{\tau_k}).$$

2.2.1 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

$$dX_t = \tilde{\alpha}(X_t) dt + \sqrt{\beta(X_t)} dW_t, \quad X_0 = x_0 \quad (9)$$

where the drift is

$$\tilde{\alpha}(X_t) = \alpha(X_t) + \beta(X_t) \nabla_{x_t} \log p(x_T | x_t). \quad (10)$$

See for example chap. 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 (1). Approximating $\alpha(X_t)$ and $\beta(X_t)$ in (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

$$dX_t = \frac{X_T - X_t}{T - t} dt + \sqrt{\beta(X_t)} dW_t. \quad (11)$$

Use of (11) 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 (11). As discussed by [Papaspiliopoulos et al. \(2013\)](#), it is impossible to simulate exact (discrete-time) realisations of (11) unless $\beta(\cdot)$ is constant. They also note that performing a local linearisation of (11) according to

[Shoji and Ozaki \(1998\)](#) (see also [Shoji 2011](#)) gives a tractable process with transition density

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

that is, the transition density of the modified diffusion bridge discussed in the previous section. Plainly, taking the Euler–Maruyama approximation of (11) yields the MDB construct, albeit without the time dependent multiplier of $\beta(x_{\tau_k})$ 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.

Unfortunately, the MDB is only efficient when the drift of (1) is approximately constant. When this is not the case, so that realisations of the SDE started from the same point exhibit strong and similar non-linearity over the inter-observation time, the modified diffusion bridge is likely to be unsatisfactory.

2.3 Lindström bridge

A bridge construct that combines the myopic sampler with the MDB is proposed in [Lindström \(2012\)](#), for the special case of known x_T . Extending the sampler to the observation scenario in (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

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

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

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

with $\gamma > 0$. This particular choice of C_{Heur} ensures that $\text{Var}(Y_T | x_{\tau_k})$ 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

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

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

$$\mu_{LB}(x_{\tau_k}) = \alpha_k + \beta_k F (F' \beta_k F \Delta_k^\gamma + \Sigma)^{-1} \times \{y_T - F'(x_{\tau_k} + \alpha_k \Delta_k)\} \tag{12}$$

and

$$\Psi_{LB}(x_{\tau_k}) = \beta_k - \beta_k F (F' \beta_k F \Delta_k^\gamma + \Sigma)^{-1} F' \beta_k \Delta\tau. \tag{13}$$

In the case of no measurement error and observation of all components, (12) and (13) become

$$\mu_{LB}^*(x_{\tau_k}) = w_k^\gamma \mu_{MDB}^*(x_{\tau_k}) + (1 - w_k^\gamma) \alpha(x_{\tau_k})$$

and

$$\Psi_{LB}^*(x_{\tau_k}) = w_k^\gamma \Psi_{MDB}^*(x_{\tau_k}) + (1 - w_k^\gamma) \beta(x_{\tau_k})$$

where

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

The Lindström bridge 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 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.

3.1 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

$$d\zeta_t = f(\zeta_t)dt, \quad \zeta_0 = x_0, \\ dR_t = \{\alpha(X_t) - f(\zeta_t)\}dt + \sqrt{\beta(X_t)} dW_t, \quad R_0 = 0. \tag{14}$$

We then aim to choose ζ_t (and therefore $f(\cdot)$) to adequately account for nonlinear dynamics (so that the drift in (14) is approximately constant), and construct the MDB of Sect. 2.2 for the residual stochastic process rather than the target process itself. Suitable choices of ζ_t and $f(\cdot)$ can be found in Sects. 3.1.1 and 3.1.2. It should be clear from the discussion in Sect. 2.2 that for known x_T , the MDB approximates the density of $R_{\tau_{k+1}} | r_{\tau_k}, r_T$ by

$$q(r_{\tau_{k+1}} | r_{\tau_k}, r_T) = N \left(r_{\tau_{k+1}} ; r_{\tau_k} + \frac{r_T - r_{\tau_k}}{T - \tau_k} \Delta\tau, \frac{T - \tau_{k+1}}{T - \tau_k} \beta(x_{\tau_k}) \Delta\tau \right). \tag{15}$$

In this case, the connection between (15) and the intractable continuous-time conditioned residual process can be established by following the arguments of Sect. 2.2.1. By approximating the drift and diffusion matrix in (14) 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

$$dR_t = \frac{R_T - R_t}{T - t} dt + \sqrt{\beta(X_t)} dW_t. \tag{16}$$

The density in (15) is then obtained by a local linearisation of (16).

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.

3.1.1 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

$$d\eta_t = \alpha(\eta_t)dt, \quad \eta_0 = x_0, \tag{17}$$

so that

$$dR_t = \{\alpha(X_t) - \alpha(\eta_t)\}dt + \sqrt{\beta(X_t)} dW_t, \quad R_0 = 0. \tag{18}$$

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

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

As in Sect. 2.2, we obtain the (approximate) joint distribution

$$\begin{aligned} & \left(\begin{array}{c} R_{\tau_{k+1}} \\ Y_T - F'\eta_T \end{array} \right) \Big|_{r_{\tau_k}} \\ & \sim N \left\{ \left(\begin{array}{cc} r_{\tau_k} + (\alpha_k - \alpha_k^\eta) \Delta \tau & \beta_k \Delta \tau \\ F'(r_{\tau_k} + (\alpha_k - \alpha_k^\eta) \Delta \tau) & F' \beta_k F \Delta \tau + \Sigma \end{array} \right), \left(\begin{array}{cc} \beta_k \Delta \tau & \beta_k F \Delta \tau \\ F' \beta_k \Delta \tau & F' \beta_k F \Delta \tau + \Sigma \end{array} \right) \right\} \end{aligned} \quad (19)$$

where $\alpha_k^\eta = \alpha(\eta_{\tau_k})$ and α_k , β_k and Δ_k are as defined in Sect. 2.2. Note that the mean in (19) 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 (17) 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 (19) with δ_k^η , conditioning on $y_T - F'\eta_T$ and using the partition $X_t = \eta_t + R_t$ gives $\Psi_{\text{RB}}(x_{\tau_k}) = \Psi_{\text{MDB}}(x_{\tau_k})$ and

$$\begin{aligned} \mu_{\text{RB}}(x_{\tau_k}) &= \alpha_k + \beta_k F (F' \beta_k F \Delta_k + \Sigma)^{-1} \\ & \times \{ y_T - F'(\eta_T + r_{\tau_k} + (\alpha_k - \delta_k^\eta) \Delta_k) \}. \end{aligned} \quad (20)$$

Note that in the case of known x_T , $\Psi_{\text{RB}}^*(x_{\tau_k}) = \Psi_{\text{MDB}}^*(x_{\tau_k})$ and (20) becomes

$$\mu_{\text{RB}}^*(x_{\tau_k}) = \delta_k^\eta + \frac{(x_T - x_{\tau_k}) - (\eta_T - \eta_{\tau_k})}{T - \tau_k}.$$

3.1.2 Further subtraction using the linear noise approximation

Whilst the solution of the SDE governing the residual stochastic process in (18) 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 $\rho_t = E(R_t | r_0, y_T)$, which we denote by $\hat{\rho}_t = E(\hat{R}_t | r_0, y_T)$. Here, $\{\hat{R}_t, t \in [0, T]\}$ is obtained through the linear noise approximation (LNA) of (18). The LNA can be derived in a number of more or less formal ways (see e.g. Kurtz (1970), van Kampen (2001) and Fearnhead et al. (2014)). Here, we give a brief exposition of the LNA and refer the reader to Fearnhead et al. (2014) and the references therein for a complete derivation.

By Taylor expanding $\alpha(X_t)$ and $\beta(X_t)$ about η_t (the solution of (17)), truncating the expansion of α at the first two terms and taking only the first term of the expansion of β , we obtain

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

where $H(\eta_t)$ is the Jacobian matrix with (i, j) th element $(H(\eta_t))_{i,j} = \partial \alpha_i(\eta_t) / \partial \eta_{j,t}$. It should be clear from the truncations used in the Taylor expansions of the drift and diffusion coefficients that the key assumption underpinning the LNA is that the stochastic term $\beta(X_t)$ is “small”. Now, for a fixed initial condition $\hat{R}_0 = \hat{r}_0$, it is straightforward to show that

$$\hat{R}_t | \hat{R}_0 = \hat{r}_0 \sim N(P_t \hat{r}_0, P_t \psi_t P_t') \quad (21)$$

where P_t and ψ_t satisfy the ODE system

$$\frac{dP_t}{dt} = H(\eta_t) P_t, \quad P_0 = I_d, \quad (22)$$

$$\frac{d\psi_t}{dt} = P_t^{-1} \beta(\eta_t) (P_t^{-1})', \quad \psi_0 = 0. \quad (23)$$

The joint distribution of \hat{R}_t and $Y_T - F'\eta_T$ (conditional on \hat{r}_0) is

$$\begin{aligned} & \left(\begin{array}{c} \hat{R}_t \\ Y_T - F'\eta_T \end{array} \right) \Big|_{\hat{r}_0} \\ & \sim N \left\{ \left(\begin{array}{cc} P_t \hat{r}_0 & P_t \psi_t P_t' \\ F' P_T \hat{r}_0 & F' P_T \psi_T P_T' F + \Sigma \end{array} \right), \left(\begin{array}{cc} P_t \psi_t P_t' & P_t \psi_t P_t' F \\ F' P_T \psi_T P_T' & F' P_T \psi_T P_T' F + \Sigma \end{array} \right) \right\}. \end{aligned} \quad (24)$$

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

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

Having obtained an explicit, closed-form (subject to the solution of (17), (22) and (23)) 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

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

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

$$\mu_{\text{RB}^-}(x_{\tau_k}) = \alpha_k + \beta_k F (F' \beta_k F \Delta_k + \Sigma)^{-1} \times \{ y_T - F'(\eta_T + \hat{\rho}_T + r_{\tau_k}^- + (\alpha_k - \delta_k^\eta - \delta_k^\rho) \Delta_k) \}. \tag{25}$$

Note that in the case of known x_T , $\Psi_{\text{RB}^*}^*(x_{\tau_k}) = \Psi_{\text{MDB}}^*(x_{\tau_k})$ and (25) becomes

$$\mu_{\text{RB}^*}^*(x_{\tau_k}) = \delta_k^\eta + \delta_k^\rho + \frac{(x_T - x_{\tau_k}) - (\eta_T - \eta_{\tau_k}) - (\hat{\rho}_T - \hat{\rho}_{\tau_k})}{T - \tau_k}.$$

3.2 Guided proposals

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 (9) and (10) 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

$$d\hat{X}_t = B(t)\hat{X}_t dt + b(t) dt + \sqrt{\sigma(t)} dW_t, \quad \hat{X}_0 = x. \tag{26}$$

Here, $B(t)$ and $\sigma(t)$ are $d \times d$ matrices and $b(t)$ is a d -vector. Note that the LNA (see Sect. 3.1.2) satisfies (26) with $B(t) = H(\eta_t)$, $b(t) = \alpha(\eta_t) - H(\eta_t)\eta_t$ and $\sigma(t) = \beta(\eta_t)$.

The guided proposal can be extended to the Gaussian additive noise regime in (2) by noting that in this case, the drift in (10) becomes

$$\tilde{\alpha}(X_t) = \alpha(X_t) + \beta(X_t) \nabla_{x_t} \log p(y_T|x_t). \tag{27}$$

Given a tractable approximation of $p(y_T|x_t)$, the Euler-Maruyama approximation of (9) can be applied over the discretisation of $[0, T]$ to give a proposal density of the form (6) with $\mu_{\text{GP}}(x_{\tau_k}) = \tilde{\alpha}(x_{\tau_k})$ and $\Psi_{\text{GP}}(x_{\tau_k}) = \beta(x_{\tau_k})$.

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 (21) with the observation regime defined in (2) gives

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

where $P_{T|t}$ and $\psi_{T|t}$ are found by integrating the ODE system in (22) and (23) from t to T with $P_{t|t} = I_d$ and $\psi_{t|t} = 0$. Hence the drift (27) becomes

$$\tilde{\alpha}(X_t) = \alpha(X_t) + \beta(X_t) P'_{T|t} F (F' P_{T|t} \psi_{T|t} P'_{T|t} F + \Sigma)^{-1} \times \{ y_T - F'(\eta_T + P_{T|t}(x_t - \eta_t)) \}. \tag{28}$$

Note that a computationally efficient implementation of this approach is obtained by using the identities $P_{T|t} = P_T P_t^{-1}$ and $\psi_{T|t} = P_t(\psi_T - \psi_t)P'_t$. Hence, the LNA ODEs in (17), (22) and (23) 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 $\text{tr}(\Sigma)$ 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 (28) 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 (28) at time T . This naive approach (henceforth referred to as GP-N) can be alleviated by integrating the ODE system given by (17), (22) and (23) for each interval $[\tau_k, T]$, $k = 0, 1, \dots, m - 1$, with $\eta_{\tau_k} = x_{\tau_k}$. In this case, the drift (27) is given by

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

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

$$\mu_{\text{GP-N}}^*(x_{\tau_k}) = \alpha(x_{\tau_k}) + \beta(x_{\tau_k}) P'_{T|\tau_k} (P_{T|\tau_k} \psi_{T|\tau_k} P'_{T|\tau_k})^{-1} \times \{ x_T - [\eta_T + P_{T|\tau_k}(x_{\tau_k} - \eta_{\tau_k})] \}$$

and

$$\mu_{\text{GP}}^*(x_{\tau_k}) = \alpha(x_{\tau_k}) + \beta(x_{\tau_k}) P'_{T|\tau_k} (P_{T|\tau_k} \psi_{T|\tau_k} P'_{T|\tau_k})^{-1} \times (x_T - \eta_T).$$

The limiting 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 $\sigma(t) \rightarrow \beta(x_T)$ as $t \rightarrow T$ by integrating (17), (22) and (23) for each τ_k is likely to be time consuming, unless the LNA ODE system is tractable. In the case of exact observations, a computationally less demanding approach is obtained in van der Meulen and Schauer (2015) by taking the transition density of (26) 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 (6) with $\Psi_{\text{GP-S}}^*(x_{\tau_k}) = \beta(x_{\tau_k})$ and

$$\mu_{\text{GP-S}}^*(x_{\tau_k}) = \alpha(x_{\tau_k}) + \beta(x_{\tau_k})\beta(x_T)^{-1} \times \left\{ \frac{x_T - x_{\tau_k} - (\eta_T - \eta_{\tau_k})}{T - \tau_k} \right\}.$$

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).

3.2.1 Use of the MDB variance

Using the Euler–Maruyama approximation of (9) gives the variance of $X_{\tau_{k+1}}|x_{\tau_k}, y_T$ in the guided proposal process as $\Psi_{\text{GP}}(x_{\tau_k})\Delta\tau = \beta(x_{\tau_k})\Delta\tau$. In Sect. 4 we investigate the effect of using the variance (8) 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. In the case of known x_T , use of the MDB variance in place of $\beta(x_{\tau_k})\Delta\tau$ comes at almost no additional computational cost. We denote this construct GP-MDB.

3.3 Computational considerations

For the observation regime in (2), 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 , $k = 1, 2, \dots, m-1$ 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 and guided proposal require the inversion of a $d \times d$ matrix at each intermediate time.

The Lindström bridge and modified diffusion bridge 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 $[\tau_k, T]$, $k = 0, 1, \dots, m-1$ for each simulated skeleton bridge, in order to maintain reasonable statistical efficiency (as measured by, for example, estimated acceptance rate of a Metropolis–Hastings independence sampler).

4 Applications

We now compare the accuracy and efficiency of the bridging methods discussed in the previous sections, by using them to make proposals inside a Metropolis–Hastings independence sampler. 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 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) or (4), depending on the observation regime. We measure the statistical efficiency of each bridge via their empirical acceptance probability. R code for the implementation of the M-H scheme can be found at <https://github.com/gawhitaker/bridges-apps>. The bridge constructs used in each example, together with their relative computational cost can be found in Table 1. 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.

4.1 Birth–death

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

$$dX_t = (\theta_1 - \theta_2)X_t dt + \sqrt{(\theta_1 + \theta_2)X_t} dW_t, \quad X_0 = x_0 \quad (29)$$

which can be seen as a degenerate case of a Feller square-root diffusion (Feller 1952). The ODE system ((17), (22) and (23)) governing the linear noise approximation of (29) is tractable, and we obtain $\eta_t = x_0 e^{(\theta_1 - \theta_2)t}$, $P_t = e^{(\theta_1 - \theta_2)t}$ and

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

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 $X_T|X_0 = x_0$,

Table 1 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

	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	–	–

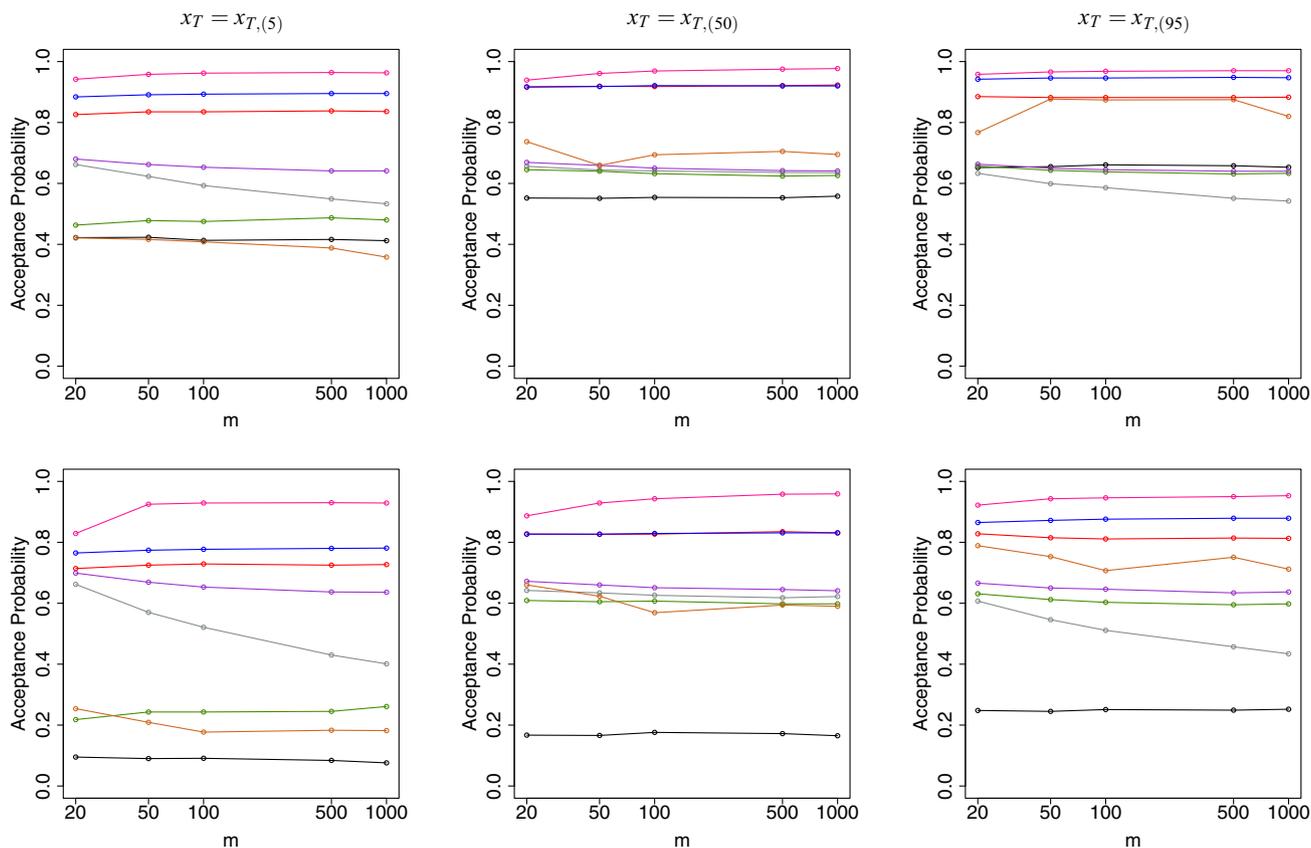


Fig. 1 Birth–death model. Empirical acceptance probability against m with $T = 1$ (1st row) and $T = 2$ (2nd row). The results are based on 100K iterations of a Metropolis–Hastings independence sampler. *Black* MDB, *brown* LB, *red* RB, *blue* RB[−], *grey* GP-N, *green* GP-S, *purple* GP, *pink* GP-MDB

found by repeatedly applying the Euler–Maruyama approximation to (29) with a small time-step. 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 $(x_{2,(5)}, x_{2,(50)}, x_{2,(95)}) = (6.97, 12.00, 18.35)$.

Since the ODE system governing the LNA is tractable for this example, there is little difference in CPU cost between the

bridges (see Table 1). Therefore, we use statistical efficiency (as measured by empirical Metropolis–Hastings acceptance probability) as a proxy for overall efficiency of each bridge, with higher probabilities preferred.

Figure 1 shows empirical acceptance probabilities against the number of sub-intervals m for each bridge and each x_T . Figures 2 and 3 compare 95% credible regions of the proposal under various bridging strategies with the true con-

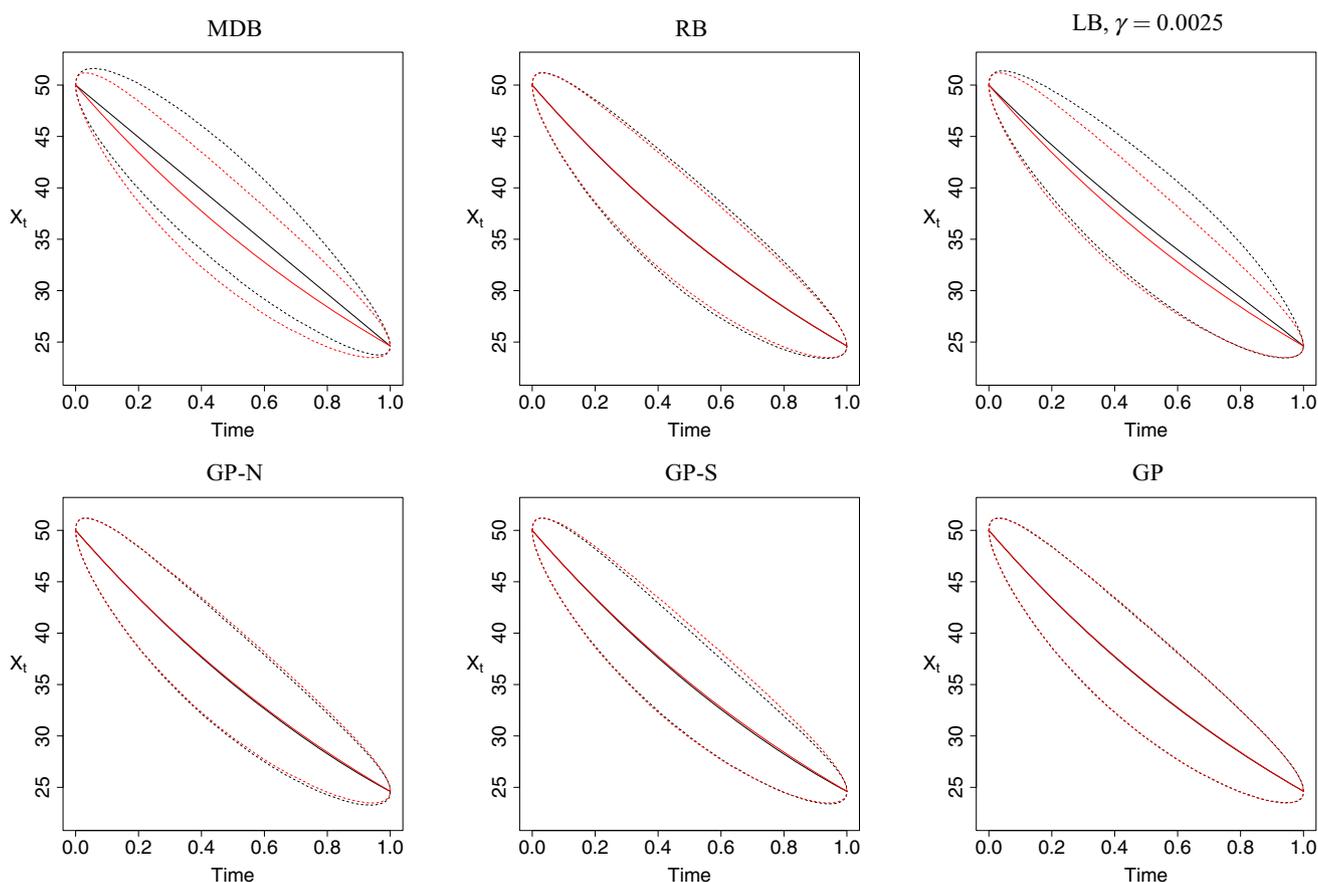


Fig. 2 Birth–death model. 95% credible region (*dashed line*) and mean (*solid line*) of the true conditioned process (*red*) and various bridge constructs (*black*) using $x_T = x_{1,(50)}$

ditioned 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 Fig. 3) 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 Fig. 1) 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.

4.2 Lotka–Volterra

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 takes the form

$$dX_t = \begin{pmatrix} \theta_1 X_{1,t} - \theta_2 X_{1,t} X_{2,t} \\ \theta_2 X_{1,t} X_{2,t} - \theta_3 X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \theta_1 X_{1,t} + \theta_2 X_{1,t} X_{2,t} & -\theta_2 X_{1,t} X_{2,t} \\ -\theta_2 X_{1,t} X_{2,t} & \theta_3 X_{2,t} + \theta_2 X_{1,t} X_{2,t} \end{pmatrix}^{\frac{1}{2}} dW_t. \quad (30)$$

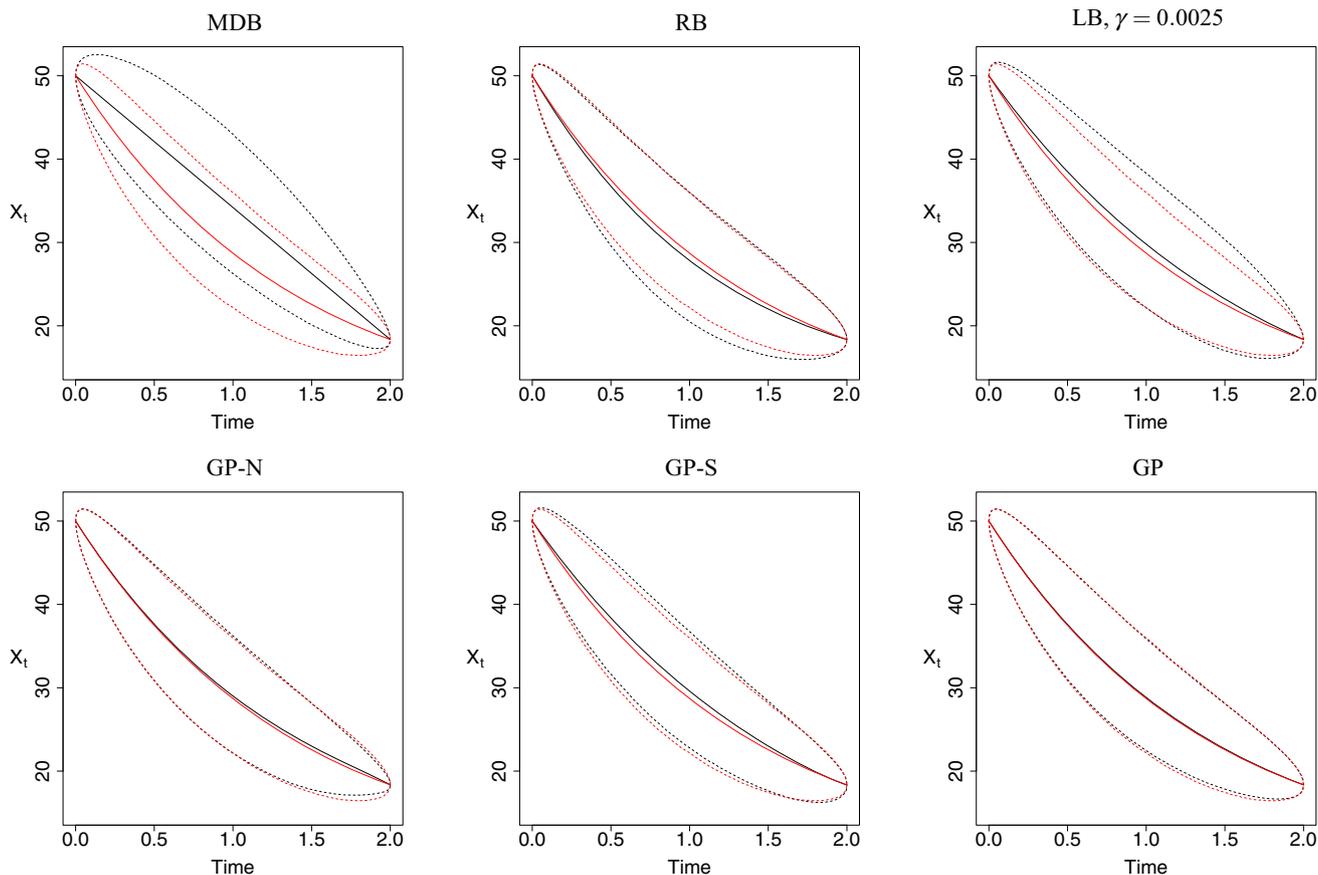


Fig. 3 Birth–death model. 95% credible region (*dashed line*) and mean (*solid line*) of the true conditioned process (*red*) and various bridge constructs (*black*) using $x_T = x_{2,(95)}$

Table 2 Lotka–Volterra model. Quantiles of $X_T|X_0 = (71, 79)'$ found by repeatedly simulating from the Euler–Maruyama approximation of (30) with $\theta = (0.5, 0.0025, 0.3)'$

	$T = 1$	$T = 2$	$T = 3$	$T = 4$
$x_{T,(5)}$	(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)

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. Note that the ODE system ((17), (22) and (23)) governing the linear noise approximation of (30) is intractable and we therefore use the R package `lsoda` to numerically solve the system when necessary.

Following Boys et al. (2008) we impose 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 2. Note that for this parameter choice, the expectation of $X_t|X_0 = (71, 79)'$ is approximately periodic with a period around 17.

We fixed the discretisation by taking $m = 50$, but note no appreciable difference in results for finer discretisations (e.g. $m = 1000$). As in the previous example, GP-N and GP-S perform relatively poorly, therefore in what follows we omit these bridges from the results. Note that we include MDB for reference. Figure 4 shows empirical acceptance probabilities against T for each bridge and each x_T . Figure 5 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

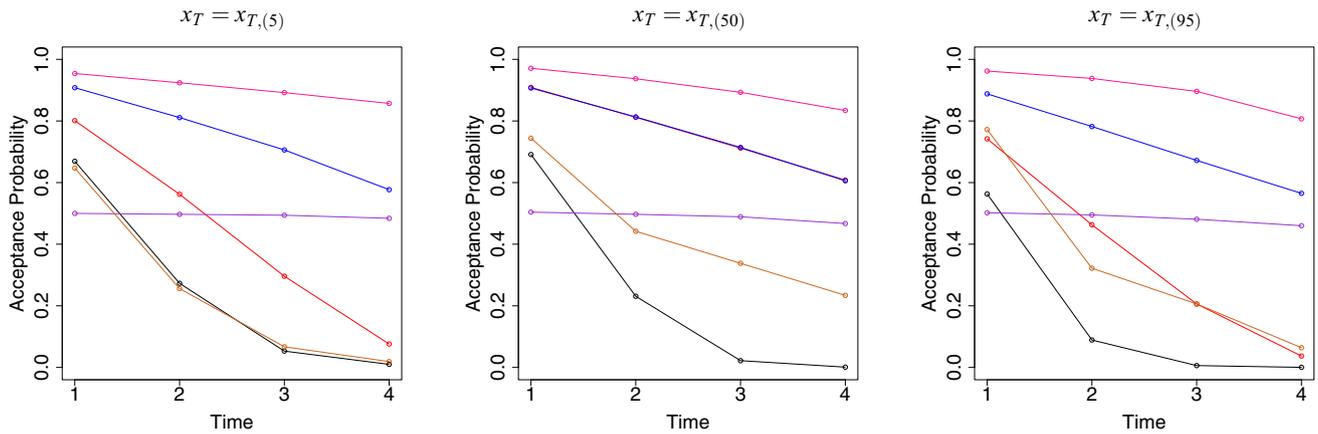


Fig. 4 Lotka–Volterra model. Empirical acceptance probabilities against T . The results are based on 100 K iterations of a Metropolis–Hastings independence sampler. *Black* MDB, *brown* LB, *red* RB, *blue* RB^- , *purple* GP, *pink* GP-MDB

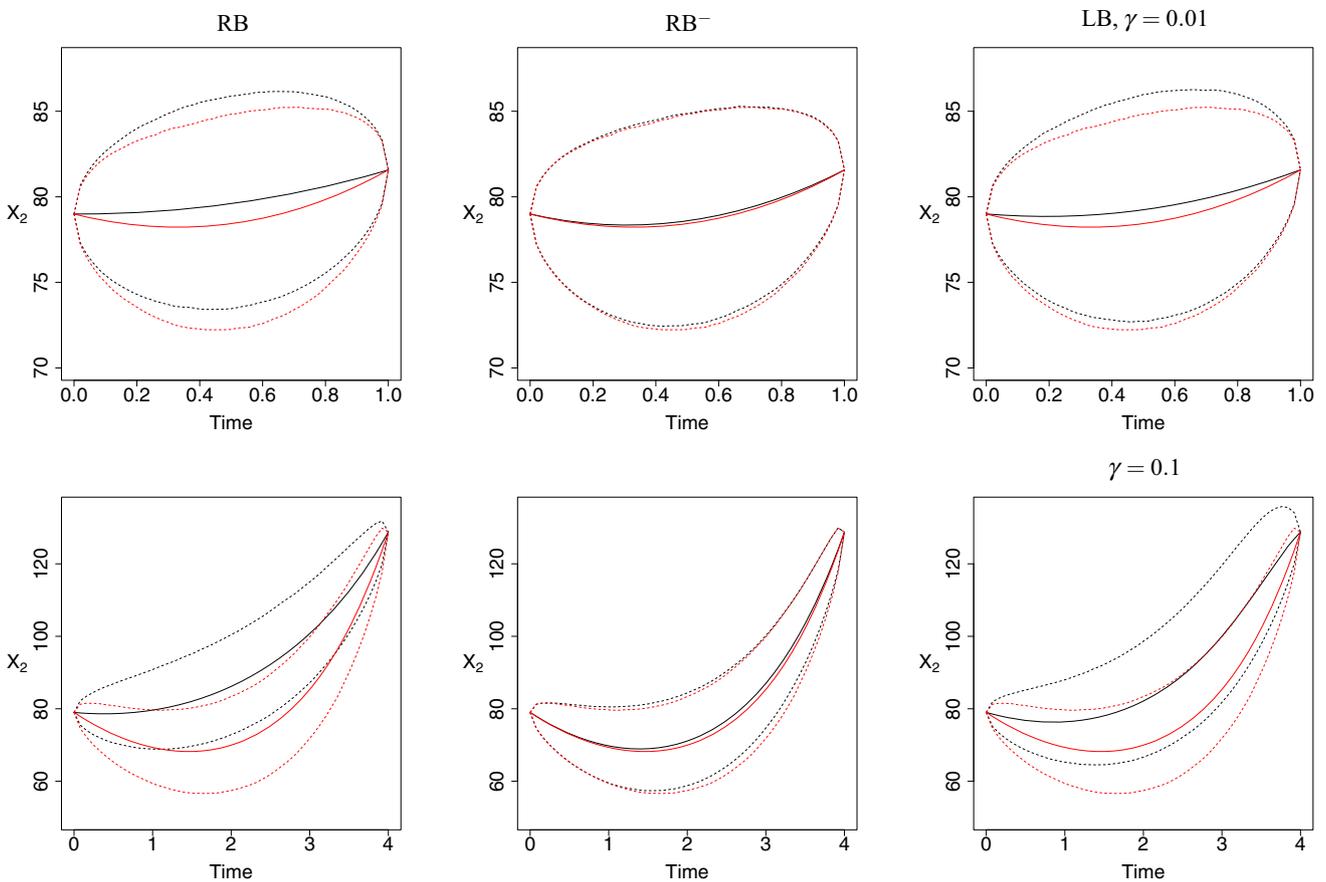


Fig. 5 Lotka–Volterra model. 95% credible region (*dashed line*) and mean (*solid line*) of the true conditioned predator component $X_{2,t}|x_0, x_T$ (*red*) and various bridge constructs (*black*) using $x_T = x_{T,(95)}$ with $T = 1$ (1st row) and $T = 4$ (2nd row)

required larger values of γ , reflecting the need for more weight to be placed on the myopic component of the construct. As for the previous example, unless x_T is the median of $X_T|x_0$, RB is comprehensively outperformed by RB^- (see Fig. 5 for the effect of increasing T on RB and RB^-). How-

ever, we see that the acceptance probabilities are decreasing in T for both constructs. As noted by Fearnhead et al. (2014), the LNA 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 article. However, the difference in estimated acceptance probabilities between GP-MDB and RB^- is fairly small, even when $T = 4$ (e.g. 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 (ESS) 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 ESS/s. When $x_T = x_{T,(5)}$ and $T = 1$, ESS/s scales roughly as 1 : 3 : 56 : 83 for GP : GP-MDB : RB : RB^- . When $T = 4$, ESS/s 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 Fig. 4.

4.3 Aphid growth

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

$$\begin{pmatrix} dN_t \\ dC_t \end{pmatrix} = \begin{pmatrix} \theta_1 N_t - \theta_2 N_t C_t \\ \theta_1 N_t \end{pmatrix} dt + \begin{pmatrix} \theta_1 N_t + \theta_2 N_t C_t & \theta_1 N_t \\ \theta_1 N_t & \theta_1 N_t \end{pmatrix}^{1/2} dW_t \quad (31)$$

where the components of $\theta = (\theta_1, \theta_2)'$ characterise the birth and death rate respectively. 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 (31) is the focus of Whitaker et al. (2015).

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 exactly. Rather, we observe

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

Table 3 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 (31) with $\theta = (1.45, 0.0009)'$, and corrupting $N_{3.57}$ with additive $N(0, \sigma^2)$ noise

	$\sigma = 5$	$\sigma = 10$	$\sigma = 50$
$y_{3.57,(5)}$	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

where $\Sigma = \sigma^2$ and $F = (1, 0)'$ so that only noisy observation of N_T is possible, and 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. (2015), 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 $Y_{3.57}|X_{2.29} = (347.55, 398.94)'$, σ . To investigate the effect of measurement error, we further take $\sigma \in \{5, 10, 50\}$. The resulting quantiles are shown in Table 3. As with the previous example, the ODE system governing the linear noise approximation of (31) is intractable and we again use the `lsoda` package to numerically solve the system when necessary.

Figure 6 shows empirical acceptance probabilities against σ for EM, RB, RB^- , GP and GP-MDB. Figure 7 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 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, $\Psi_{MDB}(x_{\tau_k})$, approaches the Euler–Maruyama variance, as used in GP.

The relative computational cost of each scheme can be found in Table 1. EM is particularly cheap to implement, given the simple form of the construct and the M-H 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 propos-

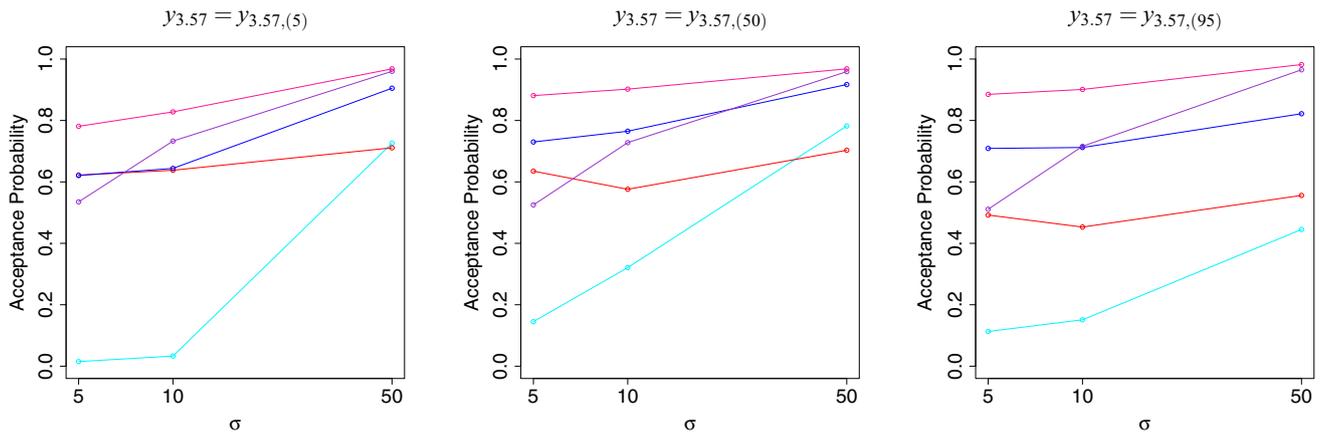


Fig. 6 Aphid growth model. Empirical acceptance probabilities against σ . The results are based on 100K iterations of a Metropolis–Hastings independence sampler. *Turquoise* EM, *red* RB, *blue* RB^- , *purple* GP, *pink* GP-MDB

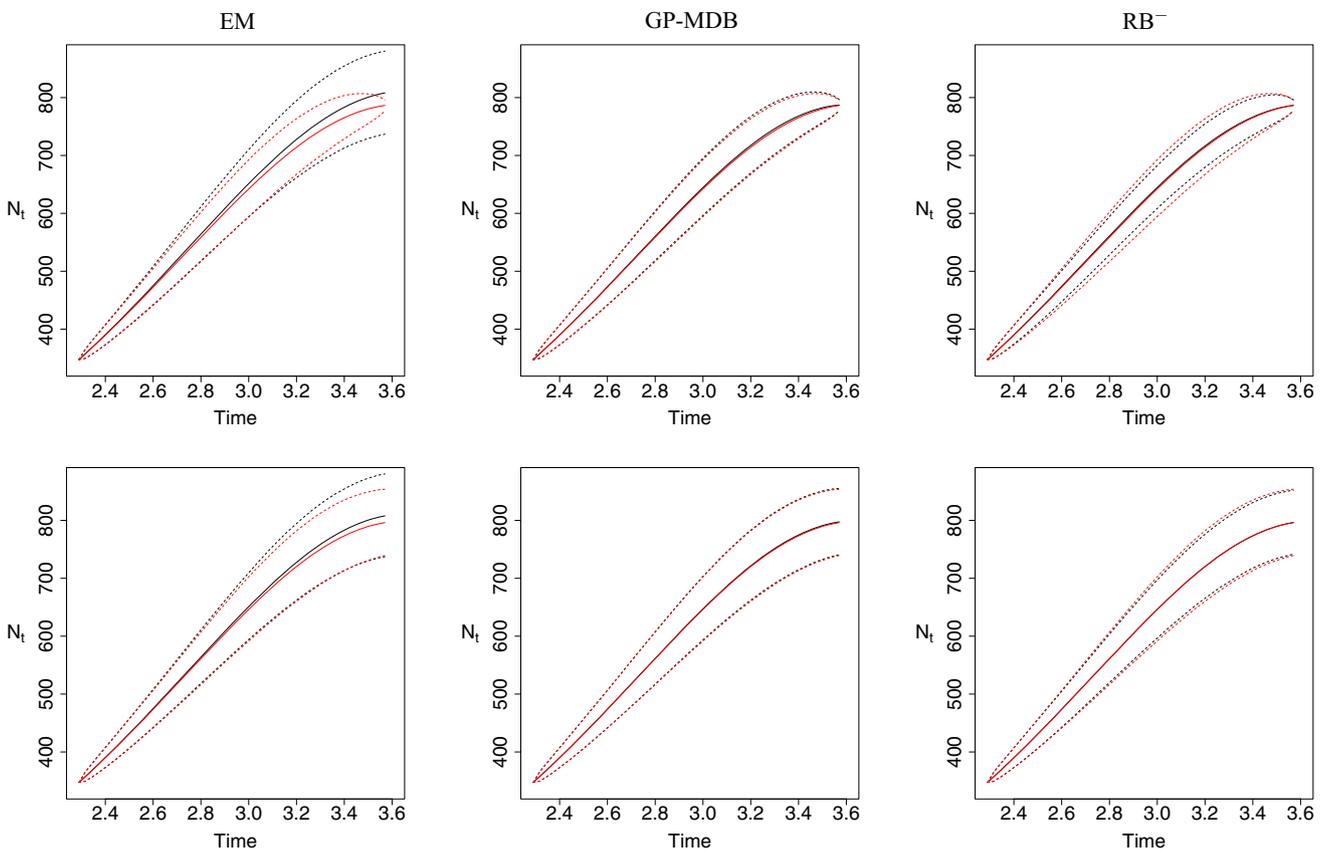


Fig. 7 Aphid growth model. 95% credible region (*dashed line*) and mean (*solid line*) of the true conditioned aphid population component $N_t | x_{2,29}, y_{3,57}$ (*red*) and various bridge constructs (*black*) using $y_{3,57} = y_{3,57,(50)}$ with $\sigma = 5$ (1st row) and $\sigma = 50$ (2nd row)

als, we found that 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.

5 Discussion

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 approximated using the modified diffusion bridge of [Durham and Gallant \(2002\)](#). Using three examples, we have 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 (denoted RB). The second variant further subtracts the linear noise approximation (LNA) of the expected conditioned residual process (denoted 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.

5.1 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 [Sect. 4.2](#)) as an example, overall efficiency (as measured by minimum effective sample size per second, ESS/s, 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 RB : RB⁻. Therefore, the relative difference in ESS/s 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 $d < 5$.

5.2 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](#)), Lind-

ström 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.

5.3 Extensions

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 (as $\Delta\tau \rightarrow 0$) 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. Interest also lies in the comparison of the bridge constructs for SDEs that exhibit multimodal behaviour, although we anticipate that further modification of the constructs will be required to efficiently deal with such a scenario.

Acknowledgments The authors would like to thank the associate editor and three anonymous referees for their suggestions for improving this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ait-Sahalia, Y.: Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* **70**(1), 223–262 (2002)

- Ait-Sahalia, Y.: Closed-form likelihood expansions for multivariate diffusions. *Ann. Stat.* **36**(2), 906–937 (2008)
- Beskos, A., Papaspiliopoulos, O., Roberts, G.O., Fearnhead, P.: Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. R. Stat. Soc.* **68**, 333–382 (2006)
- Beskos, A., Papaspiliopoulos, O., Roberts, G.O.: Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *Ann. Stat.* **37**, 223–245 (2009)
- Boys, R.J., Wilkinson, D.J., Kirkwood, T.B.L.: Bayesian inference for a discretely observed stochastic kinetic model. *Stat. Comput.* **18**(2), 125–135 (2008)
- Clark, J.: The simulation of pinned diffusions. In: *Decision and control, 1990, IEEE, Proceedings of the 29th IEEE Conference on*, pp. 1418–1420 (1990)
- Del Moral, P., Murray, L.M.: Sequential Monte Carlo with highly informative observations. *SIAM/ASA J. Uncertain. Quant.* **3**(1), 969–997 (2015)
- Delyon, B., Hu, Y.: Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.* **116**, 1660–1675 (2006)
- Durham, G.B., Gallant, A.R.: Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econ. Stat.* **20**, 297–338 (2002)
- Elerian, O., Chib, S., Shephard, N.: Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69**, 959–993 (2001)
- Eraker, B.: MCMC analysis of diffusion models with application to finance. *J. Bus. Econ. Stat.* **19**(2), 177–191 (2001)
- Fearnhead, P.: Computational methods for complex stochastic systems: a review of some alternatives to MCMC. *Statist. Comput.* **18**, 151–171 (2008)
- Fearnhead, P., Giagos, V., Sherlock, C.: Inference for reaction networks using the linear noise approximation. *Biometrics* **70**(2), 457–466 (2014)
- Feller, W.: The parabolic differential equations and the associated semi-groups of transformations. *Ann. Math.* **55**, 468–519 (1952)
- Fuchs, C.: *Inference for Diffusion Processes with Applications in Life Sciences*. Springer, Heidelberg (2013)
- Golightly, A., Wilkinson, D.J.: Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Stat. Data Anal.* **52**(3), 1674–1693 (2008)
- Golightly, A., Wilkinson, D.J.: Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* **1**(6), 807–820 (2011)
- Golightly, A., Henderson, D.A., Sherlock, C.: Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Stat. Comput.* **25**(5), 1039–1055 (2015)
- Heydari, J.J., Lawless, C., Lydall, D.A., Wilkinson, D.J.: Fast bayesian parameter estimation for stochastic logistic growth models. *Biosystems* **122**, 55–72 (2014)
- Kalogeropoulos, K., Roberts, G.O., Dellaportas, P.: Inference for stochastic volatility models using time change transformations. *Ann. Stat.* **38**(2), 784–807 (2010)
- Kurtz, T.G.: Solutions of ordinary differential equations as limits of pure jump markov processes. *J. Appl. Probab.* **7**(1), 49–58 (1970)
- Lin, M., Chen, R., Mykland, P.: On generating Monte Carlo samples of continuous diffusion bridges. *J. Am. Stat. Assoc.* **105**, 820–838 (2010)
- Lindström, E.: A regularized bridge sampler for sparsely sampled diffusions. *Stat. Comput.* **22**(1), 615–623 (2012)
- Lu, Z.H., Chow, S.M., Sherwood, A., Zhu, H.: Bayesian analysis of ambulatory blood pressure dynamics with application to irregularly spaced sparse data. *Ann. Appl. Stat.* **9**(3), 1601–1620 (2015)
- Marchand, J.L.: Conditioning diffusions with respect to partial observations. *arXiv preprint arXiv* **1105**, 1608 (2011)
- Matis, T.I., Parajulee, M.N., Matis, J.H., Shrestha, R.B.: A mechanistic model based analysis of cotton aphid population dynamics data. *Agric. For. Entomol.* **10**(4), 355–362 (2008)
- Øksendal, B.K.: *Stochastic Differential Equations: An Introduction with Applications*. Springer, New York (2003)
- Papaspiliopoulos, O., Roberts, G.O.: Importance sampling techniques for estimation of diffusion models. *Statistical Methods for Stochastic Differential Equations, Monographs on Statistics and Applied Probability*, pp. 311–337. Chapman and Hall, (2012)
- Papaspiliopoulos, O., Roberts, G.O., Stramer, O.: Data augmentation for diffusions. *J. Comput. Graph. Stat.* **22**, 665–688 (2013)
- Pedersen, A.R.: A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Stat.* **22**(1), 55–71 (1995)
- Picchini, U.: Inference for SDE models via approximate bayesian computation. *J. Comput. Graph. Stat.* **23**(4), 1080–1100 (2014)
- Roberts, G.O., Stramer, O.: On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88**(3), 603–621 (2001)
- Rogers, G.L.C., Williams, D.: *Diffusions, Markov Processes and Martingales*, vol. 1, 2nd edn. Cambridge University Press, Cambridge (2000)
- Schauer, M., van der Meulen, F., van Zanten, H.: *Guided Proposals for Simulating Multi-dimensional Diffusion Bridges*. Bernoulli In Press, Orlando (2016)
- Shoji, I.: A note on convergence rate of a linearization method for the discretization of stochastic differential equations. *Commun. Nonlinear Sci. Numer. Simul.* **16**(7), 2667–2671 (2011)
- Shoji, I., Ozaki, T.: A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika* **85**(1), 240–243 (1998)
- Stramer, O., Yan, J.: On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. University of Iowa, Technical Report (2006)
- Stramer, O., Bognar, M., Scheider, P.: Bayesian inference for discretely sampled markov processes with closed-form likelihood expansions. *J. Financ. Econ.* **8**, 450–480 (2010)
- van der Meulen, F., Schauer, M.: Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals, available from <http://arxiv.org/abs/1406.4704> (2015)
- van Kampen, N.G.: *Stochastic Processes in Physics and Chemistry*. North Holland Publishing Co., Redwood (2001)
- Whitaker, G.A., Golightly, A., Boys, R.J., Sherlock, C.: (2015) Bayesian inference for diffusion driven mixed-effects models, available from <http://arxiv.org/abs/1507.06807>