BAYESIAN INFERENCE FOR CONTINUOUS-TIME ARMA MODELS DRIVEN BY NON-GAUSSIAN LÉVY PROCESSES

Simon J. Godsill and Gary (Ligong) Yang

Signal Processing and Communications Laboratory
Cambridge University Engineering Department
Cambridge, CB2 1PZ, UK
{sjg, ly221}@eng.cam.ac.uk

ABSTRACT

In this paper we present methods for estimating the parameters of a class of non-Gaussian continuous-time stochastic process, the continuous-time autoregressive moving average (CARMA) model driven by symmetric α-Stable ($S\alpha S$) Lévy processes. In this challenging framework we are not able to evaluate the likelihood function directly, and instead we use a disctretized approximation to the likelihood. The parameters are then estimated from this approximating model using a Bayesian Monte Carlo scheme, and employing a Kalman filter to marginalize and sample the trajectory of the state process.

An efficient exploration of the parameter space is achieved through a novel reparameterization in terms of an equivalent mechanical system. Simulations demonstrate the potential of the methods.

1. INTRODUCTION

Many physical phenomena are more accurately modeled in continuous time than by discrete time approximations, since the underlying physics are typically expressed using differential equations. Moreover, a direct continuous time treatment of a problem is often more flexible: it is conceptually straightforward to deal with irregularly sampled data, missing data, or changes in sampling frequency, for example. Here we consider the continuous-time ARMA (CARMA) model, which has received much attention in the literature, especially in the Brownian-driven (Gaussian) case [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Parameter estimation for such models is more challenging than in the corresponding discrete time classes of process, since the likelihood function is nonlinear in the parameters. In the Gaussian case one can find likelihood based approaches [5, 6, 1] or Bayesian approaches using Markov chain Monte Carlo (MCMC) [7, 8, 9, 10].

However, the fundamental assumption of Gaussianity is highly idealized for many real-world processes, which exhibit outliers and jumps. As a more realistic alternative we may introduce non-Gaussianity into the underlying driving processes by replacing the Brownian motion with a Lévy process (e.g. [11, 12, 13]). The existence of general Lévy-driven CARMA processes has been proven by Brockwell [11]. Here we investigate the possibilities for Bayesian parameter estimation problems using MCMC methods.

We acknowledge partial funding for this work from the EPSRC project 'Bayesian inference for partially observed continuous time diffusions'.

2. CARMA MODELS

The governing equation for a continuous-time AR (CAR) process in state-space form is ($X^{(i)}(t)$ denotes the i-th derivative of the signal $X(t)$)

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + \mathbf{h}dW(t),$$
$$\mathbf{X}(t) = \left[X^{(0)}(t), X^{(1)}(t), ..., X^{(P-1)}(t)\right]^T \quad (1)$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_P & -a_{P-1} & -a_{P-2} & \cdots & -a_1 \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \sigma_e \end{bmatrix},$$

and the driving noise process $W(t)$ is a non-Gaussian Lévy process. This stochastic differential equation has a simple solution:

$$\mathbf{X}(t) = e^{At}\mathbf{X}(0) + \int_0^t e^{A(t-u)}\mathbf{h}dW(u) \quad (2)$$

We are interested in discrete observations of the CAR model at discrete time instants $\{t_i\}$, $Y(t_i) = \mathbf{b}\mathbf{X}(t_i) + v_i$, $v_i \sim \mathcal{N}(0, \lambda\sigma_e^2)$, $i = 1, 2, ..., N$with $\mathbf{b} = [b_0, b_1, ..., b_{P-1}]$.

The resulting process $\{Y(t)\}$ is termed a CARMA process with parameters $\mathbf{a} = [a_1\, a_2\, ...\, a_P]$, $\mathbf{b}$ and σ_e. In the case where $W(t)$ is Brownian motion these involved convenient properties mean that the likelihood function $p(\mathbf{y}|\mathbf{a}, \mathbf{b}, \sigma_e)$ can be evaluated exactly using the Kalman filter [5, 14]. However, for non-Gaussian cases the likelihood cannot in general be computed exactly and we will consider approximation schemes. For the purposes of illustration we will consider $\mathbf{b}$ to be fixed in this paper, and the extension of the methods to estimate $\mathbf{b}$ is straightforward, following the same principles as described below.

2.1. Mechanical System Reparametrization

Here we introduce a convenient and physically interpretable reparameterization of the CARMA model in terms of mechanical second order linear systems. The AR filter polynomial, for example, can be factorized into its second order sections, such that (for P even),

$$s^p + \sum_{i=1}^{P} a_i s^{P-i} = \prod_{i=1}^{P/2} s^2 + 2sc_i\omega_{n,i} + \omega_{n,i}^2$$

where $c_i > 0$ is the damping factor and $\omega_{n,i} > 0$ is the natural frequency of the i-th second order section. With P odd we can extend the representation with an additional first order section. The MA part can similarly be factorized into second order sections in cases where $\mathbf{b}$ is also to be estimated. We note that there is a one-to-one mapping from $\{c_i, \omega_{n,i}\}$ to $\mathbf{a}$ and from here on we will refer only to $(c, \omega_n) = \{c_i, \omega_{n,i}\}$ and not to $\mathbf{a}$.

With this parametrization it is more straightforward to assign priors and to check for stability of the ARMA filter; moreover the proposed MCMC algorithms are found to converge faster owing to a more independent parameterization of the model.

For the Bayesian analysis we assign independent priors to the mechanical system parameters as follows (the index i is omitted for the sake of clarity):

$$p(\omega_n) = \begin{cases} 1/\pi & , \quad \omega_n \in \{0, \pi\} \\ 0 & , \quad \text{otherwise} \end{cases} \tag{3}$$

and the prior distribution for damping factor c is designed to show some preference for smaller c (i.e. more resonant systems):

$$p(c) = \begin{cases} 2.5^{(-0.5220)} & , \quad c \in \{0, 0.5220\} \\ 2.5^{(-c)} & , \quad c \in \{0.5220, \infty\} \\ 0 & , \quad \text{otherwise} \end{cases} \tag{4}$$

3. NON-GAUSSIAN DRIVING PROCESS

Let us now consider the non-Gaussian driving noise process $\{W(t)\}$. We would like to capture the observed heavy-tailed behaviour (shocks/impulses) in many engineering processes while retaining some degree of analytic simplicity. One possibility would be a finite activity pure jump process, as popular in finance applications see e.g. [12, 13], although such processes lead to some technical difficulties in estimation. A natural alternative is the α-stable Lévy process [15] in which the increments $dW(t)$ are modeled as independent α-stable random variables. Such a process allows for a wide range of heavy-tailed behaviour and has self-similarity properties. In discrete time settings α-stable processes have found application in a number of areas, see e.g. [16, 17, 18, 19]. The α-stable family of distributions $S_\alpha(\gamma, \beta, \delta)$ can be identified by means of the characteristic function (see [15, 20, 16] for reference), which depends on four parameters: $\alpha \in (0, 2]$, measuring the tail thickness (thicker tails for smaller values of the parameter), $\beta \in [-1, 1]$ determining the degree and sign of asymmetry, $\gamma > 0$ (scale) and $\delta \in \Re$ (location). Closed-form density functions do not exist in general, but simulation of random variables is relatively straightforward (see [21, 20]), which implies that Monte Carlo strategies are a feasible approach.

We will limit ourselves here to the symmetric α-stable ($S\alpha S$) family with $\beta = 0$. The $S\alpha S$ Lévy process $\{W(t)\}$ is defined as follows [15]:

1. $W(0) = 0$ (almost surely)
2. $\{W(t)\}$ has independent symmetric α-stable incrememts
3. $W(t) - W(s) \sim S_\alpha((t-s)^{1/\alpha}, 0, 0), (t > s)$

The $S\alpha S$ Lévy process has a number of attractive properties for analysis and estimation purposes. In particular, compared with other distributional assumptions for the Lévy process, the stable properties of this distribution imply that increments at all time scales are also α-stable, with known parameters. This fact will be employed in the design of our inference procedures. Nevertheless, the non-Gaussian and continuous-time nature of the process suggest substantial challenges in the optimal estimation of parameters for Lévy-driven CARMA models.

4. INFERENCE METHODOLOGY

One suitable objective for inference would be evaluation of the likelihood function $p(\mathbf{y}|\mathbf{a}, \mathbf{b}, \sigma_e)$, or in a Bayesian setting we can explore the posterior distribution $p(\mathbf{a}, \mathbf{b}, \sigma_e|\mathbf{y})$. These distributions are available analytically for the Gaussian (i.e. Brownian-driven) case [5], but not in general for the $S\alpha S$ case. We adopt a discretization scheme to approximate the process at very fine time-scale. This approximation scheme then lends itself well to a data augmentation Monte Carlo (MC) approach which allows MC estimation of the likelihood, or exploration of the posterior distribution using Markov chain Monte Carlo (MCMC) methods.

4.1. Approximation Scheme

The challenge here is to determine (approximately) the probability distribution of $\mathbf{X}(t)$ in (2). $\mathbf{X}(t)$ has a multivariate $S\alpha S$ distribution, following from the properties of stable stochastic integrals [15]. However, an exact characterization of this distribution which is amenable to inference procedures has remained elusive to date, except for the first order case, $P = 1$. Considering first $P = 1$, results for stochastic integration of stable processes (Property 1.2.1, Page 10 of [15]) lead to:

$$\begin{aligned} \int_0^t e^{a(t-u)}\mathbf{h} dW(u) &\sim S_\alpha((\int_{u=0}^t (e^{a(t-u)})^\alpha du)^{1/\alpha}, 0, 0) \\ &= S_\alpha((\frac{e^{at\alpha} - 1}{a\alpha})^{1/\alpha}, 0, 0) \end{aligned} \tag{5}$$

This exact result provides a possible benchmark for testing the convergence of the proposed approximation method outlined below.

For $P > 1$, instead of a direct consideration of the multivariate distribution, we propose to discretize and approximate the integral in (2) as follows

$$\begin{aligned} \int_0^t e^{A(t-u)}\mathbf{h} dW(u) &= \sum_{j=0}^{M-1} \int_{j\delta t}^{(j+1)\delta t} e^{A(t-u)}\mathbf{h} dW(u) \\ &\approx \sum_{j=0}^{M-1} e^{A(t-j\delta t)}\mathbf{h} \int_{j\delta t}^{(j+1)\delta t} dW(u) \\ &= \sum_{j=0}^{M-1} e^{A(t-j\delta t)}\mathbf{h}\, \delta W_j \end{aligned} \tag{6}$$

with $\delta W_j = W((j+1)\delta t) - W(j\delta t)$, where $\delta t = t/M$ is some 'sufficiently' small time step. Such an approximation will lead to convergence in distribution as δt tends to zero, see Page 376-390, Section 7.12 and 7.13 in [15] and [20, 22] for theoretical analysis of related distcretizations of diffusion processes.

In order to perform efficient inference in this approximating model the special properties of the $S\alpha S$ Lévy process are utilized. First, we know that the increments δW_j are independently distributed as $\delta W_j \sim S_\alpha((\delta_t)^{1/\alpha}, 0, 0)$. $\delta W_j \sim S_\alpha((\delta_t)^{1/\alpha}, 0, 0)$. Moreover there is a well known representation of δW_j as a scale mixture of normals (see Page 20 Property

1.3.1 with $\alpha' = 2$ of [15]) that enables us to write the following equivalent form for δW_j as $\delta W_j \sim N(0, 2\sigma_j^2 \delta_t^{2/\alpha})$, $\sigma_j^2 \sim S_{\alpha/2}((\cos(\pi\alpha/4))^{2/\alpha}, 1, 0)$, where $S_{\alpha/2}(., 1, 0)$ is the fully skewed positive stable distribution. Essentially this means that δW_j can be equivalently represented as a zero mean Gaussian whose variance is random, unknown, independent, and distributed as a positive stable random variate. Conditioning now on the sequence of variance parameters $\{\sigma_j^2,\ j = 0, 1, 2, ..., M-1\}$, we have, approximately:

$$\mathbf{X}(t) \sim \mathcal{N}\left(e^{At}\mathbf{X}(0), \sum_{j=0}^{M-1} 2\sigma_j^2 \delta_t^{2/\alpha}\, e^{A(t-j\delta t)}\mathbf{h}\mathbf{h}^T e^{A(t-j\delta_t)^T}\right) \quad (7)$$

Moreover the approximation allows us to write an augmented dynamical model at the δ_t scale:

$$p(\mathbf{X}((j+1)\delta_t)|\mathbf{X}(j\delta_t), \sigma_j^2) \approx \mathcal{N}\left(e^{A\delta t}\mathbf{X}(j\delta t), 2\sigma_j^2\delta_t^{2/\alpha}\, e^{A\delta t}\mathbf{h}\mathbf{h}^T e^{A^T\delta_t}\right) \quad (8)$$

This augmented dynamic model, combined with the likelihood model (2), defines a state-space model, from which the parameters may be inferred. Note that we are introducing an additional M parameters $\{\sigma_j^2\}$ into the model which will need to be sampled as part of the procedure. Discretization at a fine time-scale of the type proposed here can be expected in general to cause a dramatic slow-down in algorithmic convergence, with the sampler eventually becoming 'stuck' altogether ('reducible'). Such effects were observed in [23, 24]. Here, however, the entire sample path of the process is integrated out using the Kalman filter, and we only condition on the less critical scaling parameters $\{\sigma_j\}$. Thus we can expect much less of a harmful effect on convergence in our approach. This approach, while distinct, can be compared with the discretization method of [25] in which a reducible sampler is prevented by joint sampling of sample paths and parameters.

4.2. Computational Methods

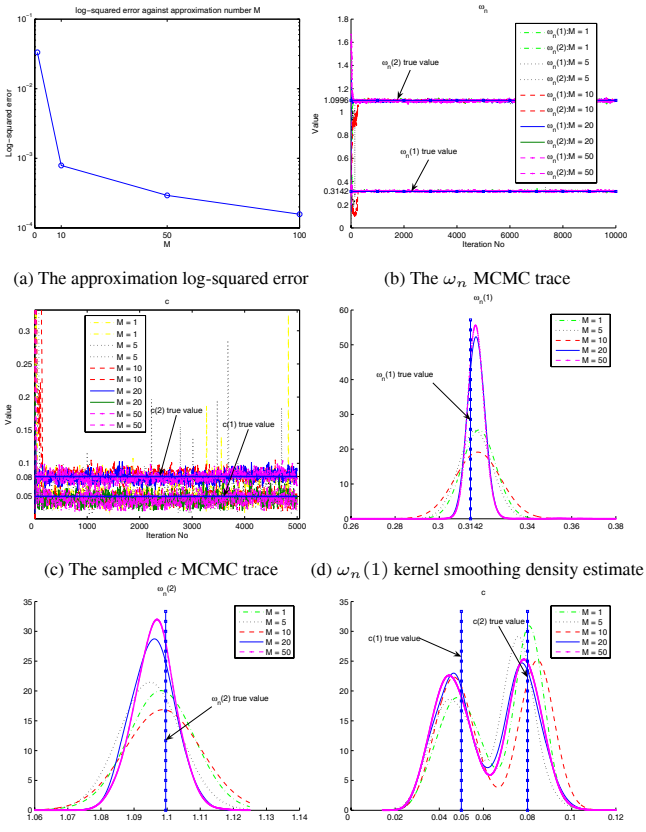
The approximating state space model of (2) combined with dynamic model (8) are suitable for inference by Monte Carlo methods. As the basic building block for the algorithms we can use a Kalman filter/simulation smoother to compute the conditional likelihood $p(\mathbf{y}|\omega_\mathbf{n}, \mathbf{c}, \{\sigma_j^2\})$ (see e.g. [14], prediction error decomposition) and to draw random samples from the distribution $p(\mathbf{x}, \sigma_e^2|\mathbf{y}, \omega_n, c)$ (see [26, 27], simulation smoother), where $\mathbf{y} = [Y(t_1), Y(t_2), ..., Y(t_N)]$ and $\mathbf{x} = [\mathbf{X}(\mathbf{0}), \mathbf{X}(\delta_t), \mathbf{X}(2\delta_t), ..., \mathbf{X}(M\delta_t)]$. In order to achieve this we have assumed additionally a prior distribution $p(\sigma_e^2) = IG(\alpha_e, \beta_e)$, where $IG(\cdot)$ denotes the standard inverted gamma prior for a variance component.

A schematic overview of the proposed MCMC algorithm is now given (all steps are straightforward applications of Metropolis-Hastings or Gibbs Sampling [28]):

- Randomly initialize parameters: $\omega_n^{(0)}, c^{(0)}, \{\sigma_j(i)^{(0)}\}$, sample $(\mathbf{x}^{(0)},\ \sigma_e^{2(0)}) \sim p(\mathbf{x}, \sigma_e^2|\mathbf{y}, \omega_n^{(0)}, c^{(0)}, \{\sigma_j(i)^{(0)}\})$
- For $k = 1$ to N_{its}
  - Sample $\omega_n^{(k)}$ and $\mathbf{c}^{(k)}$ from $p(\omega_n, \mathbf{c}|\mathbf{y}, \{\sigma_j\}^{(k-1)})$ using a mixture of Metropolis-Hastings updates (random walk plus an FFT-based independence sampler).
  - Sample $(\mathbf{x}^{(k)},\ \sigma_e^{2(k)}) \sim p(\mathbf{x}, \sigma_e^2|\mathbf{y}, \omega_n^{(k)}, c^{(k)}, \{\sigma_j(i)^{(k-1)}\})$ (simulation smoother)
  - Sample $\{\sigma_j^{2(k)}\}$ from $p(\{\sigma_j{}^2\}|\mathbf{x}^{(k)}, \sigma_e^{2(k)}, \mathbf{y}, \omega_n^{(k)}, c^{(k)})$ by a specialized hybrid rejection sampler (see [18, 19]).
- end

5. RESULTS

Extensive numerical experiments have been carried out using `MatLab` to validate our approach. In order to test convergence of the approximation as M increases for $P = 1, 200000$ random samples were generated with $a_1 = 0.5$, $\alpha = 1.1$ and various degrees of augmentation. In these simulations, M denotes the *factor* by which we augment the time axis, rather than the total number of time points (e.g. $M = 10$ corresponds to $10N$ uniformly spaced time points between t_1 and t_N. The corresponding squared errors are shown in Fig. 1 (a). Furthermore, we have examined a wide range of CARMA models with different parameters. The outcomes so far are very promising, as shown in a typical result for a 4-th order CAR model with $\alpha = 1.1$ (very heavy-tailed), in Fig. 1. Convergence is very rapid for all parameters (Fig. 1 (b) and (c)), and as M increases we see empirically a convergence of the estimated distributions, as shown in kernel density plots (Fig. 1 (d)-(f)).



(a) The approximation log-squared error (b) The ω_n MCMC trace

(c) The sampled c MCMC trace (d) $\omega_n(1)$ kernel smoothing density estimate

(e) $\omega_n(2)$ kernel smoothing density estimate (f) c kernel smoothing density estimate

Fig. 1. Experimental results

6. CONCLUSION

In this paper, we have proposed Bayesian inference approaches for non-Gaussian Lévy-driven models. A discretisation scheme was found very effective in parameter estimation for the model using MCMC. A mechanical system reparameterization further provides significant performance improvement and allows straightforward stability checking and prior assignment. Note that our basic procedure is readily extended to schemes that sample the MA parameters and to cases of unknown model order, as we have demonstrated in the Gaussian case [10].

7. ACKNOWLEDGEMENTS

The authors thank Gareth O. Roberts, Neil Shephard, Omiros Papaspiliopoulos and Chris Rogers for helpful discussions and suggestions.

8. REFERENCES

[1] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer-Verlag New York, Inc, 1996.

[2] H. Fan, T. Söderström, M. Mossberg, B. Carlsson and Y. Zou, "Estimation of continuous-time AR process parameters from discrete-time data," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1232–1244, 1999.

[3] T. Söderström, H. Fan, B. Carlsson and S. Bigi, "Least squares estimation of continuous-time ARX models from discrete-time data," *IEEE Transactions on Automatic Control*, vol. 42, no. 5, pp. 659–673, 1997.

[4] E. K. Larsson and E. G. Larsson, "Cramer-Rao bounds for continuous-time AR parameter estimation with irregular sampling," in *Proc. 26th ICASSP*, Salt Lake City, Utah, USA, May 2001.

[5] R. H. Jones, "Fitting a continuous-time autoregression to discrete data," in *Applied Time Series Analysis II*, D. F. Findley, Ed., pp. 651–682. Academic Press, London, 1981.

[6] R. H. Jones, "Fitting multivariate models to unequally spaced data," in *Time Series Analysis of Irregularly Observed Data, Lectures Notes in Statistics, 25*, E. Parzen, Ed., pp. 158–188. Springer-Verlag, 1984.

[7] P. Giannopoulos and S. J. Godsill, "Estimation of CAR processes observed in noise using Bayesian inference," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2001.

[8] P. Giannopoulos, *Modelling data using continuous-time autoregressions*, Ph.D. thesis, Department of Engineering, University of Cambridge, UK, 2002.

[9] F. Li and M. West, *Time Deformation Models: Theory and Practice*, Ph.D. thesis, Institute of Statistics and Decision Sciences, Duke University, 1997.

[10] Gary (Ligong) Yang and S. J. Godsill, "Bayesian inference for continuous-time ARMA models in Gaussian enviroment," Tech. Rep., Cambridge University Engineering Department (CUED), 2005.

[11] P. J. Brockwell, "Lévy-driven CARMA processes," *Ann. Inst. Stat. Math*, vol. 53, pp. 113–124, 2001.

[12] O. Papaspiliopoulos G. O. Roberts and P. Dellaportas, "Bayesian inference for non-gaussian ornstein-uhlenbeck stochastic volatility processes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, no. 2, pp. 369–393, 2004.

[13] O. E. Barndorff-Nielsen and N. Shephard, "Non-gaussian ou based models and some of their uses in financial economics," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 167–241, 2001.

[14] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1991.

[15] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, 1994.

[16] C. L. Nikias and M. Shao, *Signal Processing with α-Stable Distributions and Applications*, John Wiley and Sons, 1995.

[17] M. Lombardi and S. J. Godsill, "On-line Bayesian estimation of AR signals in symmetric alpha-stable noise," *IEEE Transactions on Signal Processing*, In Press.

[18] S. J. Godsill, "MCMC and EM-based methods for inference in heavy-tailed processes with alpha-stable innovations," in *Proc. IEEE Signal processing workshop on higher-order statistics*, Caesarea, Israel, June 1999.

[19] S. J. Godsill and E. E. Kuruoglu, "Bayesian inference for time series with heavy-tailed symmetric alpha -stable noise processes," in *In Proc. Applications of heavy tailed distributions in economics, engineering and statistics*, Washington DC, USA, June 1999.

[20] A. Janicki and A. Weron, *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*, Marcel Dekker, Inc., 1994.

[21] C. L. Mallows J. M. Chambers and B. W. Stuck, "A method for simulating stable random variables," *Journal of American Statistical Association*, vol. 71, pp. 340–344, 1976.

[22] P. Protter and D. Talay, "The Euler scheme for Lévy driven stochastic differential equations," *The Annals of Probability*, vol. 25, no. 1, pp. 393–423, 1997.

[23] S. Chib O. Elerian and N. Shephard, "Likelihood inference for discretely observed nonlinear diffusions," *Econometrica*, vol. 69, no. 4, pp. 959–993, 2001.

[24] G. O. Roberts and O. Stramer, "On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm," *Biometrika*, vol. 88, no. 3, pp. 603–621, 2001.

[25] A. Golightly and D. J. Wilkinson, "Bayesian inference for stochastic kinetic models using a diffusion approximation," *Biometrics*, vol. 61, no. 3, pp. 781–788, 2005.

[26] C. K. Carter and R. Kohn, "On Gibbs sampling for state space models," *Biometrika*, vol. 81, no. 3, pp. 541–553, 1994.

[27] P. DeJong and N. Shephard, "The simulation smoother for time series models," *Biometrika*, vol. 82, no. 2, pp. 339–350, 1995.

[28] S. Richardson W. R. Gilks and D. J. Spiegelhalter, Eds., *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, 1996.