

# Efficient Statistical Inference for Stochastic Reaction Processes

Andreas Ruttor and Manfred Opper

*Artificial Intelligence Group, TU Berlin, Berlin, Germany*

(Dated: June 29, 2009)

We address the problem of estimating unknown model parameters and state variables in stochastic reaction processes when only sparse and noisy measurements are available. Using an asymptotic system size expansion for the backward equation, we derive an efficient approximation for this problem. We demonstrate the validity of our approach on model systems and generalize our method to the case when some state variables are not observed.

PACS numbers: 02.50.Tt, 87.18.Vf, 02.50.Ga, 05.10.Gg

Stochastic reaction processes are models for the dynamics of a population of interacting species, e.g., chemical substances which can participate in several reactions. They play an important role in physics, chemistry, and biology [1, 2]. Given the reaction rates of the process, the probability of having a specific number of individuals (molecules) in the system at a certain time is computed by solving the chemical Master equation for which efficient exact and approximate approaches are known.

In recent years, this modelling framework has found increasing interest in the field of systems biology, where it has been applied to gene regulation and protein networks [3]. The fact that here reactions are known only qualitatively leaving a variety of numerical rate parameters unknown leads to a different type of computational problem. It becomes necessary to estimate these parameters using discrete noisy observations of the stochastic process. The information contained in the observations also changes the time evolution of the system's probability distribution. This problem of state and parameter estimation is known as data assimilation [4].

When observations are frequent and the process is in equilibrium, one may estimate parameters directly by fitting the stationary distribution or correlation functions of the process to their counterparts measured from the data. For an application of this method to models of financial markets, see [5]. But the problem of statistical inference becomes especially nontrivial when observations are sparse. In this case one would like to apply statistically efficient methods such as maximum likelihood (ML) estimators or Bayes estimators. Unfortunately, exact computations of such estimators or their simulation by stochastic techniques such as particle methods or Monte Carlo approaches may become time consuming or intractable [6–8]. To cope with this problem, we present in this Letter an efficient approximate solution based on an asymptotic system size expansion [1].

We begin by formulating a stochastic reaction process for  $d$  interacting species, which participate in several chemical reactions. The statistical dynamics for the state  $\mathbf{x} = (x_1, \dots, x_d)$  where  $x_j$  denotes the numbers of individuals (molecules) of species  $j$ , is governed by a continuous time Markov jump process. The rate of each re-

action  $i$  is given by its rate function  $h_i(\mathbf{x})$ . Waiting times between two consecutive reaction events are exponentially distributed with rate parameter  $\lambda(\mathbf{x}) = \sum_i h_i(\mathbf{x})$ . At each time such an event occurs, a specific reaction  $i$  is chosen with probability  $h_i(\mathbf{x})/\lambda(\mathbf{x})$  and the state  $\mathbf{x}$  changes deterministically into some other state  $\mathbf{x}'$  which depends on the reaction.

The probability  $p_t(\mathbf{x})$  of finding the system in state  $\mathbf{x}$  at any time  $t$  evolves through the Master equation

$$\frac{d}{dt}p_t(\mathbf{x}) = \sum_{\mathbf{x}' \neq \mathbf{x}} [p_t(\mathbf{x}')f_\theta(\mathbf{x}|\mathbf{x}') - p_t(\mathbf{x})f_\theta(\mathbf{x}'|\mathbf{x})], \quad (1)$$

where the total rate  $f_\theta(\mathbf{x}'|\mathbf{x})$  is the sum of the rates  $h_i(\mathbf{x})$  for all processes which lead from  $\mathbf{x}$  to  $\mathbf{x}'$ .  $\theta$  denotes a vector of parameters which determine the explicit form of the  $h_i$ . This might be, e.g., the rate constants appearing in the mass action kinetics of chemical reactions [9].

We will now address the problem of estimating the parameters  $\theta$  from a set of measurements  $D \equiv \{\mathbf{y}_i\}_{i=1}^N$ . The  $\mathbf{y}_i$  are noisy observations of the true process  $\mathbf{x}(t_i)$  at discrete times  $t_i$ . A related problem is the computation of  $p_t(\mathbf{x}|D)$ , the probability of the state at time  $t$  based on past and future observations.

Statistically efficient methods for parameter estimation can be based on the likelihood of the observations  $p(D|\theta)$ . A maximum likelihood (ML) approach would maximize  $p(D|\theta)$  with respect to  $\theta$ , whereas a Bayesian procedure would compute a posterior density of parameters  $p(\theta|D) \propto p(D|\theta)p(\theta)$  when prior knowledge in the form of a density  $p(\theta)$  is available.

It is straightforward to derive a recursion for  $p(D|\theta)$ . For any time  $t$ , we define  $r_t(\mathbf{x}) \equiv p(D_{\geq t}|\theta, \mathbf{x}_t = \mathbf{x})$ , the likelihood of future observations  $D_{\geq t} = \{\mathbf{y}_i\}_{t_i \geq t}$  conditioned on the present state  $\mathbf{x}(t) = \mathbf{x}$ . The likelihood of all data is then  $p(D|\theta) = \sum_{\mathbf{x}} p_0(\mathbf{x})r_0(\mathbf{x})$ , where  $p_0(\mathbf{x})$  is the distribution of the initial state. For times between two observations  $r_t$  obeys the Kolmogorov backward equation

$$\frac{d}{dt}r_t(\mathbf{x}) = \sum_{\mathbf{x}' \neq \mathbf{x}} f(\mathbf{x}'|\mathbf{x}) [r_t(\mathbf{x}) - r_t(\mathbf{x}')]. \quad (2)$$

Here and in the following we omit the parameters  $\theta$  in the rate  $f$  for notational clarity. The observations enter

through their conditional distributions (assuming independent noise)  $p(\mathbf{y}|\mathbf{x})$ , in the end condition  $r_{t_N}(\mathbf{x}) = p(\mathbf{y}_N|\mathbf{x}(t_N))$  and in the jump conditions

$$\lim_{t \rightarrow t_l^-} r(\mathbf{x}, t) = p(\mathbf{y}_l|\mathbf{x}(t_l)) \lim_{t \rightarrow t_l^+} r(\mathbf{x}, t), \quad (3)$$

where  $t_l^-$  and  $t_l^+$  denote the left and right side limits. A numerical inference method would process a sequence of parameter values  $\theta$  together with their likelihoods  $p(D|\theta)$  either for optimization or for drawing samples from the posterior. For each of these values a full solution of Eq. (2) backwards in time is required.

Given the exact parameters or good estimates, the conditional probability  $p_t(\mathbf{x}|D)$  is obtained using the Markov nature of  $\mathbf{x}(t)$  and Bayes' rule as  $p_t(\mathbf{x}|D) \propto p_t(\mathbf{x}|D_{<t}) r_t(\mathbf{x})$ , where  $p_t(\mathbf{x}|D_{<t})$  is the conditional distribution of the state based only on the observations  $D_{<t} \equiv \{\mathbf{y}_i\}_{t_i < t}$  before time  $t$ . For times between observations this probability fulfils the forward equation (1) with jump conditions at the observations. This result can be further simplified by noting that the conditional process is also Markovian. Hence,  $p_t(\mathbf{x}|D)$  itself fulfils a Master equation, where the original rates  $f(\mathbf{x}'|\mathbf{x})$  are replaced by time dependent rates  $g_t(\mathbf{x}'|\mathbf{x})$  which take the observations into account. By differentiating  $p_t(\mathbf{x}|D)$  with respect to time  $t$ , using the backward (2) and forward equations (1), the rates of the conditional process are found to satisfy

$$\frac{g_t(\mathbf{x}'|\mathbf{x})}{f(\mathbf{x}'|\mathbf{x})} = \frac{r_t(\mathbf{x}')}{r_t(\mathbf{x})}. \quad (4)$$

Note, that for times  $t > t_N$  after the last observation  $g_t(\mathbf{x}'|\mathbf{x}) = f(\mathbf{x}'|\mathbf{x})$ . Hence, the statistical inference problem requires the repeated solution of a matrix linear differential equation with dimensionality of the order  $M^d$ , where  $M$  is the typical number of states accessible to each species in the system.

We next present an approximate solution to this inference problem which follows van Kampen's idea of an asymptotic system size expansion [1]. This was developed to approximate the solution of the Master equation (1) in the limit when the typical number of individuals (molecules) is a macroscopic quantity and fluctuations are expected to be small. To lowest order, a classical deterministic rate equation for the process is obtained. The next order includes fluctuations which obey a Gaussian diffusion process of the Ornstein-Uhlenbeck type. The method is asymptotically valid under the conditions of a stable classical equilibrium. Formally, the two lowest order terms in the expansion can be derived [2] by approximating the jump process first with a diffusion process using a Kramers Moyal expansion followed by a subsequent weak noise limit, where the diffusion term is treated as a small quantity. In this Letter, we generalize this idea for solving the backward equation (2) instead. The first

step is a backward Kramers Moyal expansion [10] in (2) to obtain a backward Fokker-Planck equation (BFPE). The corresponding diffusion process has a drift vector  $\mathbf{f}$  and diffusion matrix  $\mathbf{D}$  defined by

$$\mathbf{f}(\mathbf{x}) = \sum_{\mathbf{x}' \neq \mathbf{x}} f(\mathbf{x}'|\mathbf{x})(\mathbf{x}' - \mathbf{x}), \quad (5)$$

$$\mathbf{D}(\mathbf{x}) = \sum_{\mathbf{x}' \neq \mathbf{x}} (\mathbf{x}' - \mathbf{x}) f(\mathbf{x}'|\mathbf{x})(\mathbf{x}' - \mathbf{x})^\top. \quad (6)$$

We then apply a weak noise expansion by formally rescaling  $\mathbf{D} \rightarrow \epsilon^2 \mathbf{D}$  and assuming that typical state vectors are close to a nonrandom time dependent state  $\mathbf{b}(t)$ . Setting  $\mathbf{x} = \mathbf{b}(t) + \epsilon \mathbf{u}$  and  $r_t(\mathbf{x}) = \psi_t(\mathbf{u})$  we expand the BFPE up to order  $\epsilon^2$ . The macroscopic state equation

$$\dot{\mathbf{b}} = \mathbf{f}(\mathbf{b}) \quad (7)$$

is obtained from terms of order  $\epsilon$  whereas from the second order, we obtain the partial differential equation

$$\partial_t \psi_t + (\mathbf{A}(t)\mathbf{u})^\top \nabla \psi_t + \frac{1}{2} \text{Tr}(\mathbf{D}(\mathbf{b}) \nabla \nabla^\top) \psi_t = 0 \quad (8)$$

where  $\nabla$  is the vector of derivatives with respect to  $\mathbf{u}$  and  $A_{ij}(t) = \partial_{x_j} f_i(\mathbf{x})|_{\mathbf{x}=\mathbf{b}(t)}$ . If we assume that the measurement noise can be modelled by a Gaussian, i.e.,  $p(\mathbf{y}|\mathbf{x}) \propto \exp(-\|\mathbf{y} - \mathbf{x}\|^2 / (2\sigma^2))$ , a simple solution to (8) which is compatible with the jump conditions (3), yields

$$r_t(\mathbf{x}) \approx \frac{z(t)}{\sqrt{|\mathbf{S}|}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{b})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{b}) \right], \quad (9)$$

where the matrix  $\mathbf{S}$  satisfies the differential equation

$$\dot{\mathbf{S}} = \mathbf{A}\mathbf{S} + \mathbf{S}\mathbf{A}^\top - \mathbf{D}(\mathbf{b}) \quad (10)$$

and the normalization ( $r$  is not a probability) obeys  $\dot{z}(t) = z(t) \text{Tr}(\mathbf{A}(t))$ . Note, that jump conditions lead to discontinuities for both  $\mathbf{b}$  and  $\mathbf{S}$  at the times of the observations. Hence, to compute our approximate solution  $r_t$  of the backward equation, we only have to solve the ordinary differential equations for  $\mathbf{b}$  and  $\mathbf{S}$  backward in time starting at the last observation  $\mathbf{b}(t_N) = \mathbf{y}_N$  and  $\mathbf{S}(t_N) = \sigma^2 \mathbf{I}$ , where  $\mathbf{I}$  is the unit matrix. In contrast to the exact backward equation, the dimensionality of these equations is only of the order  $d^2$ . In this approximation the likelihood  $p(D|\theta)$  is given by  $p(D|\theta) \approx (2\pi)^{d/2} z(0)$  where we have assumed an uninformative flat distribution for  $p_0(\mathbf{x})$ .

Strictly speaking, the weak noise expansion for the backward equation is only valid when the observations  $\mathbf{y}_i$  are sufficiently close to the classical path  $\mathbf{b}$ . Otherwise the assumption of Gaussian fluctuations implicit in our approximation would not be correct, and large deviation techniques would be required. However, our likelihood based approach of parameter estimation aims at making observations highly probable with respect to

TABLE I: Maximum likelihood estimates of reaction constants based on 11 observations with  $\sigma = 1$ . Mean values and standard deviations have been obtained by averaging over 1000 samples from the Lotka-Volterra process.

| parameter | inference result               | uncertainty                    | true value         |
|-----------|--------------------------------|--------------------------------|--------------------|
| $\alpha$  | $(4.0 \pm 1.2) \times 10^{-3}$ | $(0.9 \pm 0.3) \times 10^{-3}$ | $4 \times 10^{-3}$ |
| $\beta$   | $(2.1 \pm 0.9) \times 10^{-4}$ | $(0.5 \pm 0.4) \times 10^{-4}$ | $2 \times 10^{-4}$ |
| $\gamma$  | $(4.5 \pm 4.7) \times 10^{-3}$ | $(1.2 \pm 0.7) \times 10^{-3}$ | $4 \times 10^{-3}$ |
| $\delta$  | $(1.9 \pm 0.6) \times 10^{-4}$ | $(0.5 \pm 0.2) \times 10^{-4}$ | $2 \times 10^{-4}$ |

model parameters. This will drive typical paths sufficiently close to the observed data to justify the approximation for likely parameter values.

Although (8) can be viewed as the backward equation for a linear stochastic differential equation, our linearization differs from that used in the well known extended Kalman smoother [11] approach of data assimilation which has a similar computational complexity. This is based on a linearization applied to the forward filtering process  $p_t(\mathbf{x}|D_{<t})$  which requires the knowledge of initial conditions. Our backward approach can deal more robustly with vague initial conditions  $p_0(\mathbf{x})$ .

The Kalman approach would simply approximate the conditional probability  $p_t(\mathbf{x}|D)$  by that of the linearized model estimated in the filtering step. In contrast, we apply the combination of Kramers-Moyal and weak noise expansions a second time to a (forward) process with jump rate  $g_t(\mathbf{x}'|\mathbf{x})$  given in (4) leading to a new linearization at a different state. The drift of the resulting diffusion process is obtained by expanding  $r_t(\mathbf{x}')$  to first order around  $\mathbf{x}$  in Eq. (4). Using (9) and the definition of the diffusion matrix (6), we obtain

$$\mathbf{g}(\mathbf{x}, t) \approx \mathbf{f}(\mathbf{x}) - \mathbf{D}(\mathbf{b}(t))\mathbf{S}^{-1}(t)[\mathbf{x} - \mathbf{b}(t)]. \quad (11)$$

The subsequent weak noise expansion leads to a Gaussian approximation for  $p_t(\mathbf{x}|D)$  where the mean state vector  $\mathbf{m}$  and the covariance matrix  $\mathbf{C}$  evolve according to

$$\dot{\mathbf{m}} = \mathbf{g}(\mathbf{m}) \quad \dot{\mathbf{C}} = \mathbf{H}\mathbf{C} + \mathbf{C}\mathbf{H}^\top + \mathbf{D}(\mathbf{m}) \quad (12)$$

with  $H_{ij}(t) = \partial_{x_j} g_i(\mathbf{x})|_{\mathbf{x}=\mathbf{m}(t)}$ . The discontinuities in the drift  $\mathbf{g}(\mathbf{x}, t)$  at the observation times lead to discontinuities in the first derivative of  $\mathbf{m}$  and  $\mathbf{C}$ . When no explicit prior knowledge of the initial state is known, we start the forward equation with the most likely value  $\mathbf{m}(0) = \mathbf{b}(0)$  and with the uncertainty given by  $\mathbf{C}(0) = \mathbf{S}(0)$ .

We have successfully applied our method to observations generated by simulating two rather simple, but non-trivial reaction systems. The first one, the well-known Lotka-Volterra model [7, 12], consists of two interacting species, traditionally named preys and predators. In this system four reactions are possible, as both preys  $X_1$  and predators  $X_2$  can be created or destroyed at any point in

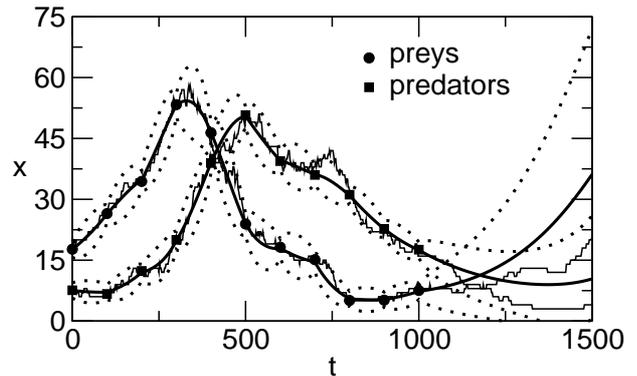
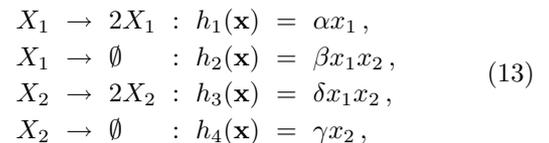
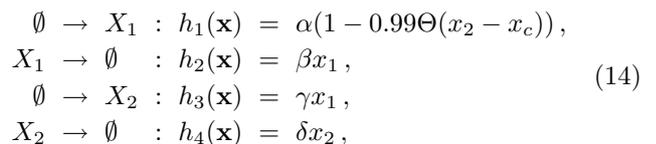


FIG. 1: Inference results for a Lotka-Volterra process with parameters given in Table I. Symbols denote noisy observations, thick lines show the posterior mean, and the 95% confidence region for parameters fixed at their maximum-likelihood values is surrounded by dotted lines. For comparison the original process is plotted as thin line. Parameter estimates for this data set:  $\alpha = (6.1 \pm 1.3) \times 10^{-3}$ ,  $\beta = (2.6 \pm 0.5) \times 10^{-4}$ ,  $\gamma = (4.1 \pm 0.9) \times 10^{-3}$ ,  $\delta = (1.8 \pm 0.4) \times 10^{-4}$ .

time. The reactions and rate laws are given by



where  $x_1$  denotes the number of preys and  $x_2$  the number of predators. The second reaction system is a simple genetic autoregulatory network [13], which can be found as parts of the transcriptional regulatory network in biological cells. It again consists of two species, mRNA  $X_1$  and protein  $X_2$ , and four reactions. Their associated rate laws are given by



where  $\Theta(x)$  is the Heaviside step function. Both mRNA and proteins decay exponentially, and proteins are produced by translation of mRNA with a rate proportional to the mRNA number  $x_1$ . In order to regulate the concentration of the protein, the production of mRNA is down-regulated by a factor 0.01 as soon as  $x_2$  increases beyond a critical threshold  $x_c$ .

Both reaction systems have been simulated using Gillespie's algorithm [14]. Observations have been obtained at regular intervals and corrupted by Gaussian noise with standard deviation  $\sigma$ . The exact initial conditions,  $x_1(0)$  and  $x_2(0)$ , have not been used for inference purposes.

As shown in Table I and Table II, the inferred reaction constants are reasonable close to their true values.

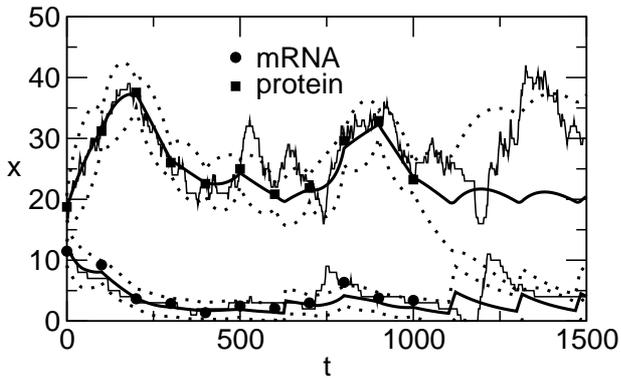


FIG. 2: Inference results for the genetic autoregulatory network with parameters given in Table II. Symbols and lines as in Fig. 1. Results of parameter estimation for this data set:  $\alpha = 1.9 \times 10^{-1}$ ,  $\beta = 6.8 \times 10^{-3}$ ,  $\gamma = 3.0 \times 10^{-2}$ ,  $\delta = 4.2 \times 10^{-3}$ ,  $x_c = 19.7$ .

Additionally, our method is quite fast, because only one backward integration is needed in each step of the maximization algorithm. In fact, its computational complexity is nearly independent of the number of observations.

In the case of the Lotka-Volterra model we used a standard gradient-based method for optimization. From the curvature of the log-likelihood, we also get predictions for the uncertainty of the parameter estimates (shown in the third column in table I) which compare fairly well to the average estimation errors obtained from the simulations.

The success of our method for the case of the genetic autoregulatory network is somewhat more surprising. The discontinuity of the reactions as a function of the threshold  $x_c$  makes this a highly nonlinear model. For this case, we have used the Nelder-Mead simplex method [15] for parameter optimization which works without computing derivatives.

Figure 1 shows results of state inference for a single Lotka-Volterra process which is based on the maximum likelihood parameter estimates. Similar results for the genetic autoregulatory network are shown in Fig. 2. Here, it is even possible to predict the strongly nonlinear behavior of the system for time points after the last observation qualitatively. Further analysis of the approximate posterior distribution  $p_t(\mathbf{x}|D)$  indicates that the prediction is indeed well calibrated; i.e.,  $\mathbf{m}(t)$  and  $\sigma_j(t) = \sqrt{C_{jj}(t)}$  are good estimators for  $\mathbf{x}(t)$  and its fluctuations. Consequently, our method gives reliable information about the state of a reaction system, although the internal noise is rather large due to the small number of individuals (molecules) in the examples given here.

Up to now we have only considered complete observations, which provide a measurement for each component of  $\mathbf{x}$ . But often it is difficult or even impossible to observe all species of a reaction system. In this case the last partial observation  $\mathbf{y}_N$  is not sufficient for initializing  $\mathbf{b}$  and  $\mathbf{S}$ . Therefore we define an additional virtual observation

TABLE II: Maximum likelihood estimates of reaction constants based on 11 observations with  $\sigma = 1$ . Mean values and standard deviations have been obtained by averaging over 1000 samples from the genetic autoregulatory network model.

| parameter | inference result               | true value         |
|-----------|--------------------------------|--------------------|
| $\alpha$  | $(1.8 \pm 1.2) \times 10^{-1}$ | $2 \times 10^{-1}$ |
| $\beta$   | $(5.7 \pm 1.9) \times 10^{-3}$ | $6 \times 10^{-3}$ |
| $\gamma$  | $(4.6 \pm 1.2) \times 10^{-2}$ | $5 \times 10^{-2}$ |
| $\delta$  | $(7.4 \pm 2.3) \times 10^{-3}$ | $7 \times 10^{-3}$ |
| $x_c$     | $17.6 \pm 3.6$                 | 20                 |

$\hat{\mathbf{y}}$  at  $t_N$  which contains only noise-free data for the hidden species. Then, the backward integration can be started with  $r_{t_N}(\mathbf{x}) = p(\mathbf{y}_N|\mathbf{x})p(\hat{\mathbf{y}}|\mathbf{x})$  and the total likelihood is given by  $p(D|\theta, \hat{\mathbf{y}}) = p(D, \hat{\mathbf{y}}|\theta)/p(\hat{\mathbf{y}}|\theta)$ . Here, we choose  $p_{t_N}(\mathbf{x})$ , the marginal distribution of the prior process, as prior distribution for our virtual observation  $\hat{\mathbf{y}}$  so that it and the denominator of the likelihood cancel out exactly. Therefore, it is possible to estimate the parameters and to reconstruct the dynamics of unobserved species by just maximizing  $p(D, \hat{\mathbf{y}}|\theta)$  with respect to  $\hat{\mathbf{y}}$  and  $\theta$ .

We are planning to combine our method with a Monte Carlo approach in order to sample from the exact posterior. Using our approximation, we expect to generate good proposal paths for an efficient Metropolis sampler.

- 
- [1] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, 1981).
  - [2] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1996), 2nd ed.
  - [3] D. J. Wilkinson, *Stochastic modelling for systems biology* (CRC Press, Boca Raton, FL, 2006).
  - [4] *Physica D* **230** (2007).
  - [5] S. Alfarano, T. Lux, and F. Wagner, *Computational Economics* **26**, 19 (2004).
  - [6] A. Golightly and D. J. Wilkinson, *Biometrics* **61**, 781 (2005).
  - [7] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood, *Statistics and Computing* **18**, 125 (2008).
  - [8] A. Doucet, *Ann. Inst. Statist. Math.* **55**, 409 (2003).
  - [9] G. M. Guldberg and P. Waage, *Journal für Praktische Chemie* **19**, 69 (1879).
  - [10] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).
  - [11] S. S. Haykin, *Kalman filtering and neural networks* (John Wiley and Sons, New York, 2001).
  - [12] G. Gilioli, S. Pasquali, and F. Ruggeri, *Bulletin of Mathematical Biology* **70**, 358 (2008).
  - [13] M. Opper and G. Sanguinetti, in *Advances in Neural Information Processing Systems 20* (Curran Associates, Red Hook, NY, 2007).
  - [14] D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976).
  - [15] J. A. Nelder and R. Mead, *Comput. J.* **7**, 308 (1965).