Chapter 28
Global Parameter Identification of Stochastic Reaction Networks from Single Trajectories

Christian L. Müller*, Rajesh Ramaswamy*, and Ivo F. Sbalzarini

Abstract We consider the problem of inferring the unknown parameters of a stochastic biochemical network model from a single measured time-course of the concentration of some of the involved species. Such measurements are available, e.g., from live-cell fluorescence microscopy in image-based systems biology. In addition, fluctuation time-courses from, e.g., fluorescence correlation spectroscopy (FCS) provide additional information about the system dynamics that can be used to more robustly infer parameters than when considering only mean concentrations. Estimating model parameters from a single experimental trajectory enables single-cell measurements and quantification of cell–cell variability. We propose a novel combination of an adaptive Monte Carlo sampler, called Gaussian Adaptation (GaA), and efficient exact stochastic simulation algorithms (SSA) that allows parameter identification from single stochastic trajectories. We benchmark the proposed method on a linear and a non-linear reaction network at steady state and during transient phases. In addition, we demonstrate that the present method also provides an ellipsoidal volume estimate of the viable part of parameter space and is able to estimate the physical volume of the compartment in which the observed reactions take place.

1 Introduction

Systems biology implies a holistic research paradigm, complementing the reductionist approach to biological organization [15, 16]. This frequently has the goal of mechanistically understanding the function of biological entities and processes in

*Authors C.L. Müller and R. Ramaswamy contributed equally to this work.

C.L. Müller • R. Ramaswamy • I.F. Sbalzarini (✉)
Institute of Theoretical Computer Science and Swiss Institute of Bioinformatics,
ETH Zurich, CH–8092 Zurich, Switzerland
e-mail: christian.mueller@inf.ethz.ch; rajeshr@ethz.ch; ivos@ethz.ch

I.I. Goryanin and A.B. Goryachev (eds.), *Advances in Systems Biology*,
Advances in Experimental Medicine and Biology 736,
DOI 10.1007/978-1-4419-7210-1_28, © Springer Science+Business Media, LLC 2012

interaction with the other entities and processes they are linked to or communicate with. A formalism to express these links and connections is provided by network models of biological processes [1, 4]. Using concepts from graph theory [26] and dynamic systems theory [44], the organization, dynamics, and plasticity of these networks can then be studied.

Systems biology models of molecular reaction networks contain a number of parameters. These are the rate constants of the involved reactions and, if spatiotemporal processes are considered, the transport rates, e.g., diffusion constants, of the chemical species. In order for the models to be predictive, these parameters need to be inferred. The process of inferring them from experimental data is called *parameter identification*. If in addition also the network structure is to be inferred from data, the problem is called *systems identification*. Here, we consider the problem of identifying the parameters of a biochemical reaction network from a single, noisy measurement of the concentration time-course of some of the involved species. While this time series can be long, ensemble replicas are not possible, either because the measurements are destructive or one is interested in variations between different specimens or cells. This is particularly important in *molecular systems biology*, where cell–cell variations are of interest or large numbers of experimental replica are otherwise not feasible.

This problem is particularly challenging and traditional genomic and proteomic techniques do not provide single-cell resolution. Moreover, in individual cells the molecules and chemical reactions can only be observed indirectly. Frequently, fluorescence microscopy is used to observe biochemical processes in single cells. Fluorescently tagging some of the species in the network of interest allows measuring the spatiotemporal evolution of their concentrations from video microscopy and fluorescence photometry. In addition, fluorescence correlation spectroscopy (FCS) allows measuring fluctuation time-courses of molecule numbers [23].

Using only a single trajectory of the mean concentrations would hardly allow identification of network parameters. There could be several combinations of network parameters that lead to the same mean dynamics. A stochastic network model, however, additionally provides information about the fluctuations of the molecular abundances. The hope is that there is then only a small region of parameter space that produces the correct behavior of the mean *and* the correct spectrum of fluctuations [31]. Experimentally, fluctuation spectra can be measured at single-cell resolution using FCS.

The stochastic behavior of biochemical reaction networks can be due to low copy numbers of the reacting molecules [10, 39]. In addition, biochemical networks may exhibit stochasticity due to extrinsic noise. This can persist even at the continuum scale, leading to continuous–stochastic models. Extrinsic noise can, e.g., arise from environmental variations or variations in how the reactants are delivered into the system. Also measurement uncertainties can be accounted for in the model as extrinsic noise, modeling our inability to precisely quantify the experimental observables.

We model stochastic chemical kinetics using the chemical master equation (CME). Using a CME forward model in biological parameter identification amounts to tracking the evolution of a probability distribution, rather than just of a single value. This prohibits predicting the state of the system and only allows statements about the probability for the system to be in a certain state, hence requiring sampling-based parameter identification methods. In the stochastic–discrete context, a number of different approaches have been suggested. Boys et al. proposed a fully Bayesian approach for parameter estimation using an explicit likelihood for data/model comparison and a Markov Chain Monte Carlo (MCMC) scheme for sampling [5]. Zechner et al. developed a recursive Bayesian estimation technique [45] to cope with cell–cell variability in experimental ensembles. Toni and co-workers used an approximate Bayesian computation (ABC) ansatz, as introduced by Marjoram and co-workers [25], that does not require an explicit likelihood [43]. Instead, sampling is done in a sequential Monte Carlo (or particle filter) framework. Reinker et al. used a hidden Markov model where the hidden states are the actual molecule abundances, and state transitions model chemical reactions [40]. Inspired by Prediction Error Methods [24], Cinquemani et al. identified the parameters of a hybrid deterministic–stochastic model of gene expression from multiple experimental time courses [7]. Randomized optimization algorithms have been used, e.g., by Koutroumpas et al. who applied a Genetic Algorithm to a hybrid deterministic–stochastic network model [21]. More recently, Poovathingal and Gunawan used another global optimization heuristic, the Differential Evolution algorithm [32]. A variational approach for stochastic two-state systems has been proposed by Stock and co-workers based on Maximum Caliber [41], an extension of Jaynes' Maximum Entropy principle [14] to non-equilibrium systems. If estimates are to be made based on a single trajectory, the stochasticity of the measurements and of the model leads to noisy similarity measures, requiring optimization and sampling schemes that are robust against noise in the data.

Here, we propose a novel combination of exact stochastic simulations for a CME forward model and an adaptive Monte Carlo sampling technique, called Gaussian Adaptation (GaA), to address the single-trajectory parameter estimation problem for monostable stochastic biochemical reaction networks. Evaluations of the CME model are done using exact partial-propensity stochastic simulation algorithms (SSA) [35]. Parameter optimization uses GaA. The method iteratively samples model parameters from a multivariate normal distribution and evaluates a suitable objective function that measures the distance between the dynamics of the forward model output and the experimental measurements. In addition to estimates of the kinetic parameters in the network, the present method also provides an ellipsoidal volume estimate of the viable part of parameter space and is able to estimate the physical volume of the compartment in which the reactions take place.

We assume that quantitative experimental time series of either a transient or the steady state of the concentrations of some of the molecular species in the network are available. This can, for example, be obtained from single-cell fluorescence microscopy by translating fluorescence intensities to estimated chemical

concentrations. Accurate methods that account for the microscope's point-spread function and the camera noise model are available to this end [6, 12, 13]. Additionally, FCS spectra can be analyzed in order to quantify molecule populations, their intrinsic fluctuations, and lifetimes [23, 34, 39]. The present approach requires only a *single* stochastic trajectory from each cell. Since the forward model is stochastic and only a single experimental trajectory is used, the objective function needs to robustly measure closeness between the experimental and the simulated trajectories. We review previously considered measures and present a new distance function in Sect. 5. First, however, we set out the formal stochastic framework and problem description below. We then describe GaA and its applicability to the current estimation task. The evaluation of the forward model is outlined in Sect. 4. We consider a linear cyclic chain and a non-linear colloidal aggregation model as benchmark test cases in Sect. 6 and conclude in Sect. 7.

2 Background and Problem Statement

We consider a network model of a biochemical system given by M coupled chemical reactions

$$\sum_{i=1}^{N} \nu_{i,j}^{-} \mathrm{S}_i \xrightarrow{k_j} \sum_{i=1}^{N} \nu_{i,j}^{+} \mathrm{S}_i \quad \forall j = 1, \ldots, M \tag{28.1}$$

between N species, where $\underline{\underline{\nu}}^{-} = [\nu_{i,j}^{-}]$ and $\underline{\underline{\nu}}^{+} = [\nu_{i,j}^{+}]$ are the stoichiometry matrices of the reactants and products, respectively, and S_i is the ith species in the reaction network. Let n_i be the population (molecular copy number) of species S_i. The reactions occur in a physical volume Ω and the macroscopic reaction rate of reaction j is k_j. This defines a dynamic system with integer-valued state $\underline{n}(t) = [n_i(t)]$ and $M + 1$ parameters $\underline{\theta} = [k_1, \ldots, k_M, \Omega]$.

The state of such a system can be interpreted as a realization of a random variable $\underline{n}(t)$ that changes over time t. Every one can know about the system is the probability for it to be in a certain state at a certain time t_j given the system's state history, hence

$$\begin{aligned} &P(\underline{n}(t_j) \,|\, \underline{n}(t_{j-1}), \ldots, \underline{n}(t_1), \underline{n}(t_0))\, \mathrm{d}^N n \\ &\quad = \mathrm{Prob}\{\underline{n}(t_j) \in [\underline{n}(t_j), \underline{n}(t_j) + \mathrm{d}\underline{n}) \,|\, \underline{n}(t_i), \quad i = 0, \ldots, j-1\}. \end{aligned} \tag{28.2}$$

A frequently made model assumption, substantiated by physical reasoning, is that the probability of the current state depends solely on the previous state, i.e.,

$$P(\underline{n}(t_j) \,|\, \underline{n}(t_{j-1}), \ldots, \underline{n}(t_1), \underline{n}(t_0)) = P(\underline{n}(t_j) \,|\, \underline{n}(t_{j-1})). \tag{28.3}$$

The system is then modeled as a first-order Markov chain where the state $\underline{n}$ evolves as:

$$\underline{n}(t + \Delta t) = \underline{n}(t) + \underline{\Xi}(\Delta t; \underline{n}, t). \tag{28.4}$$

This is the *equation of motion* of the system. If $\underline{n}$ is real-valued, it defines a continuous–stochastic model in the form of a continuous-state Markov chain. Discrete $\underline{n}$, as is the case in chemical kinetics, amount to discrete–stochastic models expressed as discrete-state Markov chains. The Markov propagator $\underline{\Xi}$ is itself a random variable, distributed with probability distribution $\Pi(\underline{\xi} \mid \Delta t; \underline{n}, t) = P(\underline{n} + \underline{\xi}, t + \Delta t \mid \underline{n}, t)$ for the state change $\underline{\xi}$. For continuous-state Markov chains, Π is a continuous probability density function (PDF), for discrete-state Markov chains a discrete probability distribution. If $\Pi(\underline{\xi}) = \delta(\underline{\xi} - \underline{\xi}_0)$, with δ the Dirac delta distribution, then the system's state evolution becomes deterministic with predictable discrete or continuous increments $\underline{\xi}_0$. Deterministic models can hence be interpreted as a limit case of stochastic models [22].

In chemical kinetics, the probability distribution Π of the Markov propagator is a linear combination of Poisson distributions with weights given by the reaction stoichiometry. This leads to the equation of motion for the population $\underline{n}$ given by

$$\underline{n}(t + \Delta t) = \underline{n}(t) + (\underline{\underline{\nu}}^{+} - \underline{\underline{\nu}}^{-}) \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_M \end{bmatrix}, \tag{28.5}$$

where $\psi_i \sim \mathscr{P}(a_i(\underline{n}(t))\Delta t)$ is a random variable from the Poisson distribution with rate $\lambda = a_i(\underline{n}(t))\Delta t$. The second term on the right-hand side of (28.5) follows a probability distribution $\Pi(\underline{\xi} \mid \Delta t; \underline{n}, t)$ whose explicit form is analytically intractable in the general case. The rates a_j, $j = 1, \ldots M$, are called the *reaction propensities* and are defined as:

$$a_j = \prod_{i=1}^{N} \binom{n_i}{\nu_{i,j}^{-}} \frac{k_j}{\Omega^{1+\sum_{i'=1}^{N} \nu_{i',j}^{-}}}. \tag{28.6}$$

They depend on the macroscopic reaction rates and the reaction volume and can be interpreted as the probability rates of the respective reactions. Advancing (28.5) with a Δt such that more than one reaction event happens per time step yields an approximate simulation of the biochemical network as done in approximate SSA [3, 9].

An alternative approach consists in considering the evolution of the state probability distribution $P(\underline{n}, t \mid \underline{n}_0, t_0)$ of the Markov chain described by (28.5), hence:

$$\frac{\partial P}{\partial t} = \sum_{j=1}^{M} \left(\prod_{i=1}^{N} \mathrm{E}_i^{\nu_{i,j}^{-}} \mathrm{E}_j^{-\nu_{i,j}^{+}} - 1 \right) a_j(\underline{n}(t)) P(\underline{n}, t) \tag{28.7}$$

with the *step operator* $\mathrm{E}_i^p f(\underline{n}) = f\left(\underline{n} + p\hat{\underline{i}}\right)$ for any function f, where $\hat{\underline{i}}$ is the N-dimensional unit vector along the ith dimension. This equation is called the *CME*. Directly solving it for P is analytically intractable, but trajectories of the Markov chain governed by the unknown state probability P can be sampled using exact SSA [8]. Exact SSAs are exact in the sense that they sample Markov chain realizations from the exact solution P of the CME, without ever explicitly computing this solution. Since SSAs are Monte Carlo algorithms, however, a sampling error remains.

Assuming that the population $\underline{n}$ increases with the volume Ω, $\underline{n}$ can be approximated as a continuous random variable in the limit of large volumes, and (28.5) becomes

$$\underline{n}(t + \Delta t) = \underline{n}(t) + (\underline{\underline{\nu}}^+ - \underline{\underline{\nu}}^-) \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_M \end{bmatrix}, \tag{28.8}$$

where $\eta_i \sim \mathcal{N}(a_i(\underline{n}(t))\Delta t, a_i(\underline{n}(t))\Delta t)$ are normally distributed random variables. The second term on the right-hand side of this equation is a random variable, that is, distributed according to the corresponding Markov propagator $\Pi(\underline{\xi} \mid \Delta t; \underline{n}, t)$, which is a Gaussian. Equation (28.8) is called the *chemical Langevin equation* with Π given by:

$$\Pi\left(\underline{\xi} \mid \Delta t; \underline{n}, t\right) = (2\pi)^{-N/2} \left|\underline{\underline{\Sigma}}\right|^{-1/2} \mathrm{e}^{-\frac{1}{2}\left(\underline{\xi}-\underline{\mu}\right)^{\mathrm{T}} \underline{\underline{\Sigma}}^{-1}\left(\underline{\xi}-\underline{\mu}\right)}, \tag{28.9}$$

where

$$\underline{\mu} = \Delta t \left(\underline{\underline{\nu}}^+ - \underline{\underline{\nu}}^-\right) \begin{bmatrix} a_1(\underline{n}(t)) \\ \vdots \\ a_M(\underline{n}(t)) \end{bmatrix} \text{and } \underline{\underline{\Sigma}} = \Delta t \left(\underline{\underline{\nu}}^+ - \underline{\underline{\nu}}^-\right) \operatorname{diag} \underline{a}(\underline{n}(t)) \left(\underline{\underline{\nu}}^+ - \underline{\underline{\nu}}^-\right)^{\mathrm{T}}.$$

The corresponding equation for the evolution of the state PDF is the non-linear Fokker–Planck equation, given by:

$$\frac{\partial P}{\partial t} = \underline{\nabla}^{\mathrm{T}} \left(\frac{1}{2}\underline{\underline{D}}\,\underline{\nabla} - \underline{F}\right) P(\underline{n}, t), \tag{28.10}$$

where

$$\underline{\nabla}^{\mathrm{T}} = \left[\frac{\partial}{\partial n_1}, \ldots, \frac{\partial}{\partial n_N}\right], \tag{28.11}$$

$$F_i = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} \mathrm{d}\xi_i \, \xi_i \, \Pi(\underline{\xi} \mid \Delta t; \underline{n}, t), \tag{28.12}$$

and

$$D_{ij} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{d}\xi_i \mathrm{d}\xi_j \, \xi_i \, \xi_j \, \Pi(\underline{\xi} \mid \Delta t; \underline{n}, t) - F_i F_j . \quad (28.13)$$

At much larger Ω, when the population $\underline{n}$ is on the order of Avogadro's number, (28.8) can be further approximated as:

$$\underline{n}(t + \Delta t) = \underline{n}(t) + \left(\underline{\underline{\nu}}^{+} - \underline{\underline{\nu}}^{-}\right) \begin{bmatrix} \phi_1(\underline{n}(t))\Delta t \\ \vdots \\ \phi_M(\underline{n}(t))\Delta t \end{bmatrix}, \quad (28.14)$$

where $\phi_j(\underline{n}) = k_j \Omega^{1-\sum_{i'=1}^{N} \nu_{i',j}^{-}} \prod_{i=1}^{N} n_i^{\nu_{i,j}^{-}} (\nu_{i,j}^{-}!)^{-1}$. Note that the second term on the right-hand side of this equation is a random variable whose probability distribution is the Dirac delta

$$\Pi\left(\underline{\xi} \mid \Delta t; \underline{n}, t\right) = \delta\left(\underline{\xi} - \left(\underline{\underline{\nu}}^{+} - \underline{\underline{\nu}}^{-}\right) \begin{bmatrix} \phi_1(\underline{n}(t))\Delta t \\ \vdots \\ \phi_M(\underline{n}(t))\Delta t \end{bmatrix}\right). \quad (28.15)$$

Equation (28.14) hence is a deterministic equation of motion. In the limit $\Delta t \to 0$ this equation can be written as the ordinary differential equation

$$\frac{\mathrm{d}\underline{x}}{\mathrm{d}t} = \left(\underline{\underline{\nu}}^{+} - \underline{\underline{\nu}}^{-}\right) \begin{bmatrix} \phi_1(\underline{x}(t)) \\ \vdots \\ \phi_M(\underline{x}(t)) \end{bmatrix} \quad (28.16)$$

for the *concentration* $\underline{x} = \underline{n}\Omega^{-1}$. This is the classical *reaction rate equation* for the system in (28.1).

By choosing the appropriate probability distribution Π of the Markov propagator, one can model reaction networks in different regimes: small population $\underline{n}$ (small Ω) using SSA over (28.7), intermediate population (intermediate Ω) using (28.8), and large population (large Ω) using (28.16). The complete model definition therefore is $\mathscr{M}(\underline{\theta}) = \left\{\underline{\underline{\nu}}^{-}, \underline{\underline{\nu}}^{+}, \Pi\right\}$.

The problem considered here can then be formalized as follows: Given a forward model $\mathscr{M}(\underline{\theta})$ and a single noisy trajectory of the population of the chemical species $\hat{\underline{n}}(t_0 + (q-1)\Delta t_{\mathrm{exp}})$ at K discrete time points $t = t_0 + (q-1)\Delta t_{\mathrm{exp}}$, $q = 1, \ldots, K$, we wish to infer $\underline{\theta} = [k_1, \ldots, k_M, \Omega]$. The time between two consecutive measurements Δt_{exp} and the number of measurements K are given by the experimental technique used. As a forward model we use the full CME as given in (28.7) and sample trajectories from it using the partial-propensity formulation of Gillespie's exact SSA as described in Sect. 4.

3 Gaussian Adaptation for Global Parameter Optimization, Approximate Bayesian Computation, and Volume Estimation

Gaussian Adaptation, introduced in the late 1960s by Gregor Kjellström [17,19], is a Monte Carlo technique that has originally been developed to solve design-centering and optimization problems in analog electric circuit design. Design-centering solves the problem of determining the nominal values (resistances, capacitances, etc.) of the components of a circuit such that the circuit output is within specified design bounds and is maximally robust against random variations in the circuit components with respect to a suitable criterion or objective function. This problem is a superset of general optimization, where one is interested in finding a parameter vector that minimizes (or maximizes) an objective function without any additional robustness criterion. GaA has been specifically designed for scenarios where the objective function $f(\underline{\theta})$ is only available in a black-box (or oracle) model that is defined on a real-valued domain $\mathscr{A} \subseteq \mathbb{R}^n$ and returns scalar real-valued output. The black-box model assumes that gradients or higher-order derivatives of the objective function may not exist or may not be available, hence including the class of discontinuous and noisy functions. The specific objective function used here is presented in Sect. 5.

The principle idea behind GaA is the following: Starting from a user-defined point in parameter space, GaA explores the space by iteratively sampling single parameter vectors from a multivariate Gaussian distribution $\mathscr{N}(\underline{m}, \underline{\underline{\Sigma}})$ whose mean $\underline{m} \in \mathbb{R}^n$ and covariance matrix $\underline{\underline{\Sigma}} \in \mathbb{R}^{n \times n}$ are dynamically adapted based on the information from previously accepted samples. The acceptance criterion depends on the specific mode of operation, i.e., whether GaA is used as an optimizer or as a sampler [27,28]. Adaptation is performed such as to maximize the entropy of the search distribution under the constraint that acceptable search points are found with a predefined, fixed hitting (success) probability $p < 1$ [19]. Using the definition of the entropy of a multivariate Gaussian distribution $\mathscr{H}(\mathscr{N}) = \log\left(\sqrt{(2\pi \mathrm{e})^n \det(\underline{\underline{\Sigma}})}\right)$ shows that this is equivalent to maximizing the determinant of the covariance matrix $\underline{\underline{\Sigma}}$. GaA thus follows Jaynes' Maximum Entropy principle [14].

GaA starts by setting the mean $\underline{m}^{(0)}$ of the multivariate Gaussian to an initial acceptable point $\underline{\theta}^{(0)}$ and the Cholesky factor $\underline{\underline{Q}}^{(0)}$ of the covariance matrix to the identity matrix $\underline{\underline{I}}$. At each iteration $g > 0$, the covariance $\underline{\underline{\Sigma}}^{(g)}$ is decomposed as: $\underline{\underline{\Sigma}}^{(g)} = \left(r \cdot \underline{\underline{Q}}^{(g)}\right)\left(r \cdot \underline{\underline{Q}}^{(g)}\right)^{\mathrm{T}} = r^2 \left(\underline{\underline{Q}}^{(g)}\right)\left(\underline{\underline{Q}}^{(g)}\right)^{\mathrm{T}}$, where r is the scalar step size that controls the scale of the search. The matrix $\underline{\underline{Q}}^{(g)}$ is the normalized square root of $\underline{\underline{\Sigma}}^{(g)}$, found by eigen- or Cholesky decomposition of $\underline{\underline{\Sigma}}^{(g)}$. The candidate parameter

vector in iteration $g+1$ is sampled from a multivariate Gaussian according to $\underline{\theta}^{(g+1)} = \underline{m}^{(g)} + r^{(g)}\underline{\underline{Q}}^{(g)}\underline{\eta}^{(g)}$, where $\underline{\eta}^{(g)} \sim \mathcal{N}(\underline{0}, \underline{\underline{I}})$. The parameter vector is then evaluated by the objective function $f(\underline{\theta}^{(g+1)})$.

Only if the parameter vector is accepted, the following adaptation rules are applied: The step size r is increased as $r^{(g+1)} = f_{\mathrm{e}} \cdot r^{(g)}$, where $f_{\mathrm{e}} > 1$ is termed the *expansion factor*. The mean of the proposal distribution is updated as:

$$\underline{m}^{(g+1)} = \left(1 - \frac{1}{N_{\mathrm{m}}}\right)\underline{m}^{(g)} + \frac{1}{N_{\mathrm{m}}}\underline{\theta}^{(g+1)}. \tag{28.17}$$

N_{m} is a weighting factor that controls the learning rate of the method. The successful search *direction* $\underline{d}^{(g+1)} = \left(\underline{\theta}^{(g+1)} - \underline{m}^{(g)}\right)$ is used to perform a rank-one update of the covariance matrix: $\underline{\underline{\Sigma}}^{(g+1)} = \left(1 - \frac{1}{N_{\mathrm{C}}}\right)\underline{\underline{\Sigma}}^{(g)} + \frac{1}{N_{\mathrm{C}}}\underline{d}^{(g+1)}\underline{d}^{(g+1)\mathrm{T}}$. N_{C} weights the influence of the accepted parameter vector on the covariance matrix. In order to decouple the volume of the covariance (controlled by $r^{(g+1)}$) from its orientation, $\underline{\underline{Q}}^{(g+1)}$ is normalized such that $\det\left(\underline{\underline{Q}}^{(g+1)}\right) = 1$.

In case $\underline{\theta}^{(g+1)}$ is not accepted at the current iteration, only the step size is adapted as $r^{(g+1)} = f_{\mathrm{c}} \cdot r^{(g)}$, where $f_{\mathrm{c}} < 1$ is the *contraction factor*.

The behavior of GaA is controlled by several strategy parameters. Kjellström analyzed the information-theoretic optimality of the acceptance probability p for GaA in general regions [19]. He concluded that the efficiency E of the process and p are related as $E \propto -p \log p$, leading to an optimal $p = \frac{1}{\mathrm{e}} \approx 0.3679$, where e is Euler's number. A proof is provided in [18]. Maintaining this optimal hitting probability corresponds to leaving the volume of the distribution, measured by $\det\left(\underline{\underline{\Sigma}}\right)$, constant under stationary conditions. Since $\det\left(\underline{\underline{\Sigma}}\right) = r^{2n}\det\left(\underline{\underline{Q}}\,\underline{\underline{Q}}^{\mathrm{T}}\right)$, the expansion and contraction factors f_{e} and f_{c} expand or contract the volume by a factor of f_{e}^{2n} and f_{c}^{2n}, respectively. After S accepted and F rejected samples, a necessary condition for constant volume thus is: $\prod_{i=1}^{S}(f_{\mathrm{e}})^{2n}\prod_{i=1}^{F}(f_{\mathrm{c}})^{2n} = 1$. Using $p = \frac{S}{S+F}$, and introducing a small $\beta > 0$, the choice $f_{\mathrm{e}} = 1 + \beta(1-p)$ and $f_{\mathrm{c}} = 1 - \beta p$ satisfies the constant-volume condition to first order. The scalar rate β is coupled to N_{C}. N_{C} influences the update of $\underline{\underline{\Sigma}} \in \mathbb{R}^{n\times n}$, which contains n^2 entries. Hence, N_{C} should be related to n^2. We suggested using $N_{\mathrm{C}} = (n+1)^2/\log(n+1)$ as a standard value, and coupling $\beta = \frac{1}{N_{\mathrm{C}}}$ [29]. A similar reasoning is also applied to N_{m}. Since N_{m} influences the update of $\underline{\underline{m}} \in \mathbb{R}^n$, it is reasonable to set $N_{\mathrm{m}} \propto n$. We propose $N_{\mathrm{m}} = \mathrm{e}n$ as a standard value.

Depending on the specific acceptance rule used, GaA can be turned into a global optimizer [29], an adaptive MCMC sampler [27,28], or a volume estimation method [30], as described next.

3.1 GaA for Global Black-Box Optimization

In a minimization scenario, GaA uses an adaptive-threshold acceptance mechanism. Given an initial scalar cutoff threshold $c_{\mathrm{T}}^{(0)}$, we accept a parameter vector $\underline{\theta}^{(g+1)}$ at iteration $g+1$ if $f\left(\underline{\theta}^{(g+1)}\right) < c_{\mathrm{T}}^{(g)}$. Upon acceptance, the threshold c_{T} is lowered as $c_{\mathrm{T}}^{(g+1)} = \left(1 - \frac{1}{N_{\mathrm{T}}}\right) c_{\mathrm{T}}^{(g)} + \frac{1}{N_{\mathrm{T}}} f\left(\underline{\theta}^{(g+1)}\right)$, where N_{T} controls the weighting between the old threshold and the objective-function value of the *accepted* sample. This sample-dependent threshold update renders the algorithm invariant to linear transformations of the objective function. The standard strategy parameter value is $N_{\mathrm{T}} = \mathrm{e}n$ [28]. We refer to [28] for further information about convergence criteria and constraint handling techniques in GaA.

3.2 GaA for Approximate Bayesian Computation and Viable Volume Estimation

Replacing the threshold acceptance-criterion by a probabilistic Metropolis criterion, and setting $N_{\mathrm{m}} = 1$, turns GaA into an adaptive MCMC sampler with global adaptive scaling [2]. We termed this method *Metropolis-GaA* [27, 28]. Its strength is that GaA can automatically adapt to the covariance of the target probability distribution while maintaining the fixed hitting probability. For standard MCMC, this cannot be achieved without fine-tuning the proposal using multiple MCMC runs. We hypothesize that GaA might also be an effective tool for ABC [43]. In essence, the ABC ansatz is MCMC without an explicit likelihood function [25]. The likelihood is replaced by a distance function – which plays the same role as our objective function – that measures closeness between a parameterized model simulation and empirical data $\mathscr{D}$, or summary statistics thereof. When a uniform prior over the parameters and a symmetric proposal are assumed, a parameter vector in ABC is unconditionally accepted if its corresponding distance function value $f(\underline{\theta}^{(g+1)}) < c_{\mathrm{T}}$ [25]. The threshold c_{T} is a problem-dependent constant that is fixed prior to the actual computation. Marjoram and co-workers have shown that samples obtained in this manner are approximately drawn from the posterior parameter distribution given the data $\mathscr{D}$. While Pritchard et al. used a simple rejection sampler [33], Marjoram and co-workers proposed a standard MCMC scheme [25]. Toni and co-workers used sequential MC for sample generation [43]. To the best of our knowledge, however, the present work presents the first application of an adaptive MCMC scheme for ABC in biochemical network parameter inference. Finally, we emphasize that when GaA's mean, covariance matrix, and hitting probability p stabilize during ABC, they provide direct access to an ellipsoidal estimation of the volume of the viable parameter space as defined by the threshold c_{T} [30]. Hafner and co-workers have shown how to use such viable volume estimates for model discrimination [11].

4 Evaluation of the Forward Model

In each iteration of the GaA algorithm, the forward model of the network needs to be evaluated for the proposed parameter vector $\underline{\theta}$. This requires an efficient and exact SSA for the chemical kinetics of the reaction network, used to generate trajectories $\underline{n}(t)$ from $\mathscr{M}(\underline{\theta})$. Since GaA could well propose parameter vectors that lead to low copy numbers for some species, it is important that the SSA be exact since approximate algorithms are not appropriate at low copy number.

In its original formulation, Gillespie's SSA has a computational cost that is linearly proportional to the total number M of reactions in the network. If many model evaluations are required, as in the present application, this computational cost quickly becomes prohibitive. While more efficient formulations of SSA have been developed for weakly coupled reaction networks, their computational cost remains proportional to M for strongly coupled reaction networks [35]. A reaction network is weakly coupled if the number of reactions that are influenced by any other reaction is bounded by a constant. If a network contains at least one reaction whose firing influences the propensities of a fixed proportion (in the worst case all) of the other reactions, then the network is strongly coupled [35]. Scale-free networks as seem to be characteristic for systems biology models [1, 42] are by definition strongly coupled. This is due to the existence of *hubs* that have a higher connection probability than other nodes. These hubs frequently correspond to chemical reactions that produce or consume species that also participate in the majority of the other reactions, such as water, ATP, or CO_2 in metabolic networks.

We use partial-propensity methods [35, 36] to simulate trajectories according to the solution of the chemical master (28.7) of the forward model. Partial-propensity methods are exact SSAs whose computational cost scales at most linearly with the number N of species in the network [35]. For large networks, this number is usually much smaller than the number of reactions. Depending on the network model at hand, different partial-propensity methods are available for its efficient simulation. Strongly coupled networks where the rate constants span only a limited spectrum of values are best simulated with the partial-propensity direct method (PDM) [35]. Multi-scale networks where the rate constants span many orders of magnitude are most efficiently simulated using the sorting partial-propensity direct method (SPDM) [35]. Weakly coupled reaction networks can be simulated at constant computational cost using the partial-propensity SSA with composition-rejection sampling (PSSA-CR) [37]. Lastly, reaction networks that include time delays can be exactly simulated using the delay partial-propensity direct method (dPDM) [38]. Different combinations of the algorithmic modules of partial-propensity methods can be used to constitute all members of this family of SSAs [36]. We refer to the original publications for algorithmic details, benchmarks of the computational cost, and a proof of exactness of partial-propensity methods.

5 Objective Function

In the context of parameter identification of stochastic biochemical networks, a number of distance or objective functions have previously been suggested. Reinker et al. proposed an approximate maximum-likelihood measure under the assumption that only a small number of reactions fire between two experimental measurement points, and a likelihood based on singular value decomposition that works when many reactions occur per time interval [40]. Koutroumpas et al. compared objective functions based on least squares, normalized cross-correlations, and conditional probabilities using a Genetic Algorithm [21]. Koeppl and co-workers proposed the Kantorovich distance to compare experimental and model-based probability distributions [20]. Alternative distance measures include the Earth Mover's distance or the Kolomogorov–Smirnov distance [32]. These distance measures, however, can only be used when many experimental trajectories are available. In order to measure the distance between a *single* experimental trajectory $\underline{\hat{n}}(t)$ and a *single* model output $\underline{n}(t)$, we propose a novel cost function $f(\underline{\theta}) = f\left(\mathscr{M}(\underline{\theta}), \underline{\hat{n}}\right)$ that reasonably captures the kinetics of a monostable system. We define a compound objective function $f(\underline{\theta}) = f_1(\underline{\theta}) + f_2(\underline{\theta})$ with

$$f_1(\underline{\theta}) = \sum_{i=1}^{4} \gamma_i \ , \qquad f_2(\underline{\theta}) = \sum_{i=1}^{N} \frac{\sum_{l=0}^{z_x} \left|\mathrm{ACF}_l(\hat{n}_i) - \mathrm{ACF}_l(n_i)\right|}{\sum_{l=0}^{z_x} \mathrm{ACF}_l(\hat{n}_i)} \ , \tag{28.18}$$

where

$$\gamma_i = \sum_{j=1}^{N} \sqrt{\left(\frac{\mu_i(n_j) - \mu_i(\hat{n}_j)}{\mu_i(\hat{n}_j)}\right)^2} \tag{28.19}$$

with the central moments given by:

$$\mu_i(n_j) = \begin{cases} \sum_{p=1}^{K} n_j\left(t_0 + (p-1)\Delta t_{\mathrm{exp}}\right) & \text{if} \quad i = 1 \\ \left(\left|\sum_{q=1}^{K}\left(n_j\left(t_0 + (q-1)\Delta t_{\mathrm{exp}}\right) - \mu_1(n_j)\right)^i\right|\right)^{1/i} & \text{otherwise} \end{cases} \tag{28.20}$$

and the time–autocorrelation function (ACF) at lag l given by:

$$\mathrm{ACF}_l(n_i) = \frac{n_i(t_0)n_i(t_0 + l\ \Delta t_{\mathrm{exp}}) - (\mu_1(n_i))^2}{\mu_2(n_i)}.$$

The variable z_x is the lag at which the experimental ACF crosses 0 for the first time. The function $f_1(\underline{\theta})$ measures the difference between the first four moments of $\underline{n}$ and $\underline{\hat{n}}$. This function alone would, however, not be enough to capture the kinetics since it lacks information about correlations in time. This is taken into account by $f_2(\underline{\theta})$, measuring the difference in the lifetimes of all chemical species. These lifetimes are systematically modulated by the volume Ω [39], hence enabling volumetric measurements of intra-cellular reaction compartments along with the identification of the rate constants.

The present objective function allows inclusion of experimental readouts from image-based systems biology. The moment-matching part is a typical readout from fluorescence photometry, whereas the autocorrelation of the fluctuations can directly be measured using, e.g., FCS.

6 Results

We estimate the unknown parameters $\underline{\theta}$ for two reaction networks: a weakly coupled cyclic chain and a strongly coupled non-linear colloidal aggregation network. For the cyclic chain we estimate $\underline{\theta}$ at steady state. For the aggregation model we estimate $\underline{\theta}$ both at steady state and in the transient phase. Every kinetic parameter is allowed to vary in the interval $[10^{-3},\ 10^{3}]$ and the reaction volume Ω in $[1,\ 500]$. Each GaA run starts from a point selected uniformly at random in logarithmic parameter space.

6.1 Weakly Coupled Reaction Network: Cyclic Chain

The cyclic chain network is given by:

$$\begin{aligned} \mathrm{S}_i &\xrightarrow{k_i} \mathrm{S}_{i+1} \qquad && i = 1, \ldots, N-1, \\ \mathrm{S}_i &\xrightarrow{k_N} \mathrm{S}_1 \qquad && i = N. \end{aligned} \tag{28.21}$$

In this linear network, the number of reactions M is equal to the number of species N. The maximum degree of coupling of this reaction network is 2, irrespective of the size of the system (length of the chain), rendering it weakly coupled [35]. We hence use PSSA-CR to evaluate the forward model with a computational complexity of $O(1)$ [37]. In the present test case, we limit ourselves to 3 species and 3 reactions, i.e., $N = M = 3$. The parameter vector for this case is given by $\underline{\theta} = [k_1, k_2, k_3]$, since the kinetics of linear reactions is independent of the volume Ω [39].

We simulate steady-state "experimental" data $\underline{\hat{n}}$ using PSSA-CR with ground truth $k_1 = 2$, $k_2 = 1.5$, $k_3 = 3.2$ (see Fig. 28.1a). We set the initial population of the species to $n_1(t = 0) = 50$, $n_2(t = 0) = 50$, and $n_3(t = 0) = 50$ and sample a single CME trajectory at equi-spaced time points with $\Delta t_{\text{exp}} = 0.1$ between $t = t_0$ and $t = t_0 + (K-1)\Delta t_{\text{exp}}$ with $t_0 = 2000$ and $K = 1001$ for each of the 3 species S_1, S_2, and S_3. For the generated data we find $z_x = 7$.

We generate trajectories from the forward model for every parameter vector $\underline{\theta}$ proposed by GaA using PSSA-CR between $t = 0$ and $t = (K-1)\Delta t_{\text{exp}} = 100$, starting from the initial population $n_i(t = 0) = \hat{n}_i(t = t_0)$.

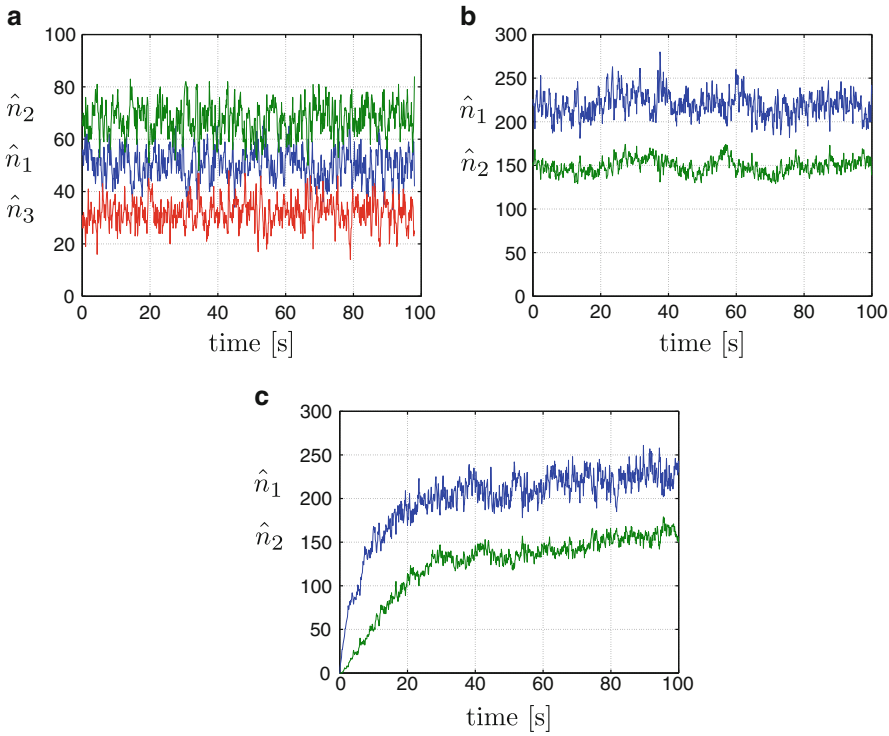

Fig. 28.1 In silico data for all test cases. (**a**) Time evolution of the populations of three species in the cyclic chain model at steady state (starting at $t_0 = 2000$). (**b**) Time evolution of the populations of two species in the aggregation model at steady state (starting at $t_0 = 5000$). (**c**) Same as (**b**), but during the transient phase (starting at $t_0 = 0$)

Before turning to the actual parameter identification, we illustrate the topography of the objective function landscape for the present example. We fix $k_3 = 3.2$ to its optimal value and perform a two-dimensional grid sampling for k_1 and k_2 over the full search domain. We use 40 logarithmically spaced sample points per parameter, resulting in 40^2 parameter combinations. For each combination we evaluate the objective function. The resulting landscapes of $f_1(\underline{\theta})$, $f_2(\underline{\theta})$, and $f(\underline{\theta})$ are depicted in Fig. 28.2a. Figure 28.2b shows refined versions around the global optimum. We see that the moment-matching term $f_1(\underline{\theta})$ is largely responsible for the global single-funnel topology of the landscape. The autocorrelation term $f_2(\underline{\theta})$ sharpens the objective function near the global optimum and renders it locally more isotropic.

We perform both global optimization and ABC runs using GaA. In each of the 15 independent optimization runs the number of objective function evaluations (FES) is limited to MAX_FES $= 1000M = 3000$. We set the initial step size to $r^{(0)} = 1$ and perform all searches in logarithmic scale of the parameters. Independent restarts from uniform random points are performed when the step size r drops below

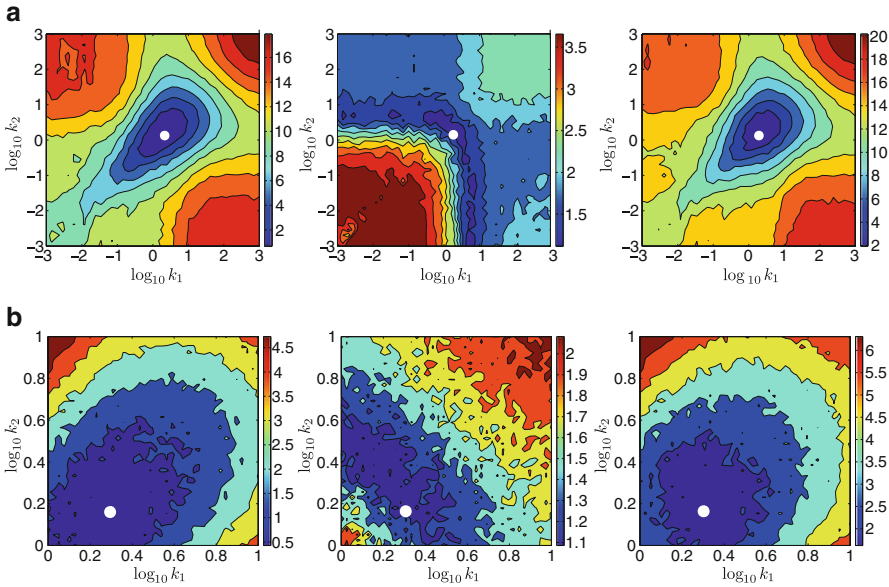

Fig. 28.2 (**a**) Global objective function landscape for the cyclic chain over the complete search domain for optimal $k_3 = 3.2$. The three panels from left to right show $f_1(\underline{\theta})$, $f_2(\underline{\theta})$, and $f(\underline{\theta})$, respectively. (**b**) A refined view of the global objective function landscape near the global optimum. The three panels from left to right show $f_1(\underline{\theta})$, $f_2(\underline{\theta})$, and $f(\underline{\theta})$, respectively. The white dots mark the ground truth parameters

10^{-4} [29]. For each of the 15 independent runs, the 30 parameter vectors with the smallest objective function value are collected and displayed in the box plot shown in the left panel of Fig. 28.3a. All 450 collected parameter vectors have objective function values smaller than 1.6. These results suggest that the present method is able to accurately determine the correct scale of the kinetic parameters from a single experimental trajectory, although an overestimation of the rates is apparent.

We use the obtained optimization results for subsequent ABC runs. We conduct 15 independent ABC runs using $c_{\mathrm{T}} = 2$. The starting points for the ABC runs are selected uniformly at random from the 450 collected parameter vectors in order to ensure stable initialization. For each run we again set MAX_FES $= 1000M = 3000$. The initial step size $r^{(0)}$ is set to 0.1, and the parameters are again explored in logarithmic scale. For all runs we observe rapid convergence of the empirical hitting probability p_{emp} to the optimal $p = \frac{1}{\mathrm{e}}$ (see Sect. 3). We collect the ABC samples along with the means and covariances of GaA as soon as $|p_{\mathrm{emp}} - p| < 0.05$. As an example we show the histograms of the posterior samples for a randomly selected run in Fig. 28.3b. The means of the posterior distributions are again larger than the true kinetic parameters. Using GaA's means, covariance matrices, and the corresponding hitting probabilities that generated the posterior

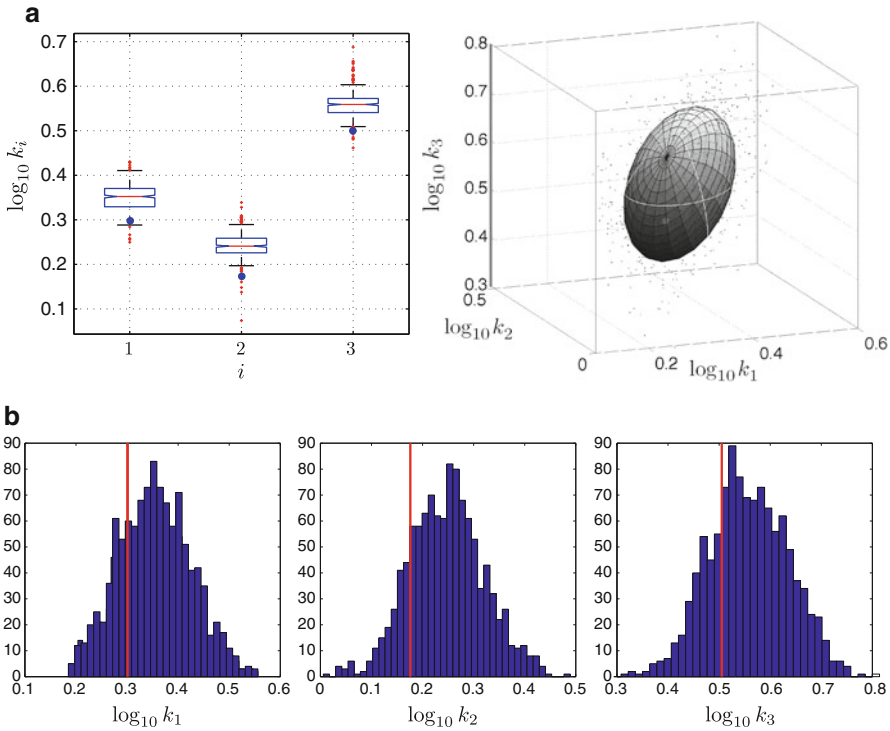

Fig. 28.3 (**a**) *Left panel*: Box plot of the 30 best parameter vectors from each of the 15 independent optimization runs. The blue dots mark the true parameter values. *Right panel*: Ellipsoidal volume estimate of the parameter space below an objective-function threshold $c_{\mathrm{T}} = 2$ from a single ABC run. (**b**) Empirical posterior distributions of the kinetic parameters from the same single ABC run with $c_{\mathrm{T}} = 2$. The red lines indicate the true parameters

samples, we can construct an ellipsoidal volume estimation [30]. This is done by multiplying each eigenvalue of the average of the collected covariance matrices with $c_{p_{\mathrm{emp}}} = \mathrm{inv}\,\chi^2_n(p_{\mathrm{emp}})$, the n-dimensional inverse Chi-square distribution evaluated at the empirical hitting probability. The product of these scaled eigenvalues and the volume of the n-dimensional unit sphere, $|S(n)| = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$, then yields the ellipsoid volume with respect to a uniform distribution (see [30] for details). The resulting ellipsoid contains the optimal kinetic parameter vector and is depicted in the right panel of Fig. 28.3a. It has a volume of 0.045 in log-parameter space. This constitutes only 0.0208% of the initial search space volume, indicating that GaA significantly narrows down the viable parameter space around the true optimal parameters despite the noise in the forward model and in the data.

6.2 Strongly Coupled Reaction Network: Colloidal Aggregation

The colloidal aggregation network is given by:

$$\begin{aligned}
\emptyset &\xrightarrow{k_1^{\text{on}}} \mathrm{S}_1 \\
\mathrm{S}_i + \mathrm{S}_j &\xrightarrow{k_{ij}} \mathrm{S}_{i+j} \qquad i+j = 1,\ldots,N \\
\mathrm{S}_{i+j} &\xrightarrow{\bar{k}_{ij}} \mathrm{S}_i + \mathrm{S}_j \qquad i+j = 1,\ldots,N \\
\mathrm{S}_i &\xrightarrow{k_i^{\text{off}}} \emptyset \qquad i = 1,\ldots,N.
\end{aligned} \tag{28.22}$$

For this network of N species, the number of reactions is $M = \left\lfloor \frac{N^2}{2} \right\rfloor + N + 1$. The maximum degree of coupling of this reaction network is proportional to N, rendering the network strongly coupled [35]. We hence use SPDM to evaluate the forward model with a computational complexity of $O(N)$ [35]. We use SPDM instead of PDM since the search path of GaA is unpredictable and could well generate parameters that lead to multi-scale networks. For this test case, we limit ourselves to two species, i.e., $N = 2$ and $M = 5$. The parameter vector for this case is $\underline{\theta} = \left[k_{11}, \bar{k}_{11}, k_1^{\text{on}}, k_1^{\text{off}}, k_2^{\text{off}}, \Omega\right]$.

We perform GaA global optimization runs following the same protocol as for the cyclic chain network with MAX_FES = $1000(M+1) = 6000$.

6.2.1 At Steady State

We simulate "experimental" data $\underline{\hat{n}}$ using SPDM with ground truth $k_{11} = 0.1$, $\bar{k}_{11} = 1.0$, $k_1^{\text{on}} = 2.1$, $k_1^{\text{off}} = 0.01$, $k_2^{\text{off}} = 0.1$, and $\Omega = 15$ (see Fig. 28.1b). We set the initial population of the species to $n_1(t=0) = 0$, $n_2(t=0) = 0$, and $n_3(t=0) = 0$ and sample $K = 1001$ equi-spaced data points between $t = t_0$ and $t = t_0 + (K-1)\Delta t_{\text{exp}}$ with $t_0 = 5000$ and $\Delta t_{\text{exp}} = 0.1$.

We generate trajectories from the forward model for every parameter vector $\underline{\theta}$ proposed by GaA using SPDM between $t = 0$ and $t = (K-1)\Delta t_{\text{exp}} = 100$, stating from the initial population $n_i(t=0) = \hat{n}_i(t=t_0)$.

The optimization results are summarized in the left panel of Fig. 28.4a. For each of the 15 independent runs, the 30 lowest-objective parameter vectors are collected and shown in the box plot. We observe that the true parameters corresponding to $\theta_2 = \bar{k}_{11}$, $\theta_3 = k_1^{\text{on}}$, $\theta_4 = k_1^{\text{off}}$, and $\theta_5 = k_2^{\text{off}}$ are between the 25th and 75th percentiles of the identified parameters. Both the first parameter and the reaction volume are, on average, overestimated. Upon rescaling the kinetic rate constants with the estimated volume, we find $\underline{\theta}^{\text{norm}} = [\theta_1/\theta_6, \theta_2, \theta_3\,\theta_6, \theta_4, \theta_5]$, which are the specific probability rates of the reactions. The identified values are shown in the

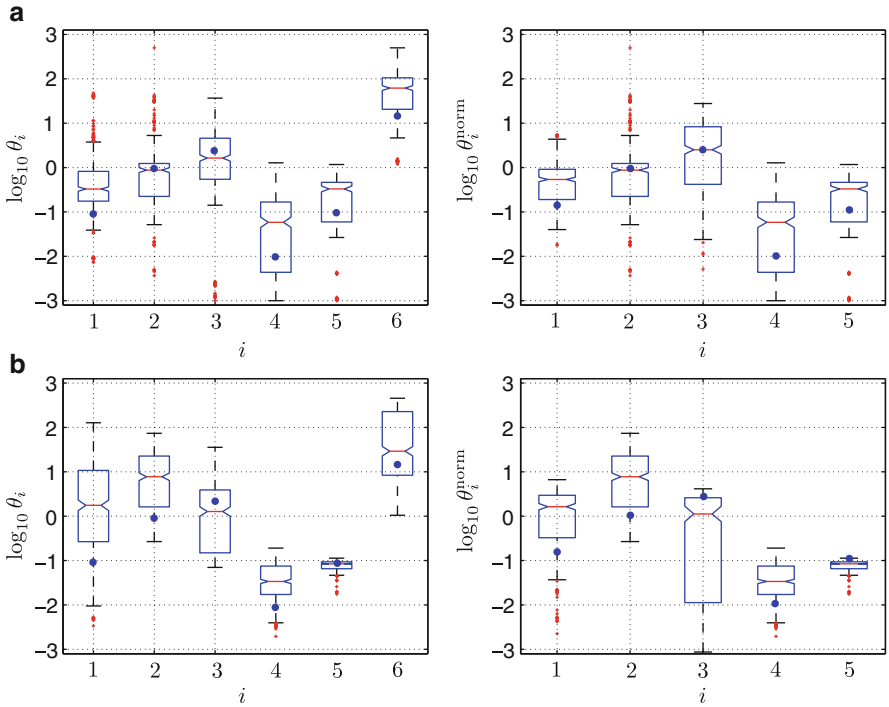

Fig. 28.4 (**a**) *Left panel*: Box plot of the 30 best parameter vectors from each of the 15 independent optimization runs for the steady-state data set. *Right panel*: Box plots of the normalized parameters (see main text for details). (**b**) *Left panel*: Box plot of the 30 best parameter vectors from each of the 15 independent optimization runs for the transient data set. *Right panel*: Box plot of the normalized parameters (see main text for details). The blue dots indicate the true parameter values

right panel of Fig. 28.4a. The median of the identified θ_3^{norm} coincides with the true specific probability rate. Likewise, θ_1^{norm} is closer to the 25th percentile of the parameter distribution. This suggests a better estimation performance of GaA in the space of specific probability rates, at the expense of not obtaining an estimate of the reactor volume.

6.2.2 In the Transient Phase

We simulate "experimental" data in the transient phase of the network dynamics using the same parameters as above between $t = t_0$ and $t = (K-1)\Delta t_{\text{exp}}$ with $t_0 = 0$, $\Delta t_{\text{exp}} = 0.1$, and $K = 1001$ (see Fig. 28.1c). We evaluate the forward model with $n_i(t = 0) = \hat{n}_i(t = t_0)$ to obtain trajectories between $t = 0$ and $t = (K-1)\Delta t_{\text{exp}}$ for every proposed parameter vector $\underline{\theta}$.

The optimization results for the transient case are summarized in Fig. 28.4b. We observe that the true parameters corresponding to $\theta_3 = k_1^{\text{on}}$, $\theta_5 = k_2^{\text{off}}$, and $\theta_6 = \Omega$ are between the 25th and 75th percentiles of the identified parameters. The remaining parameters are, on average, overestimated. In the space of rescaled parameters $\underline{\theta}^{\text{norm}}$ we do not observe a significant improvement of the estimation.

7 Conclusions and Discussion

We have considered parameter estimation in monostable stochastic biochemical networks from single experimental trajectories. Parameter identification from single time series is desirable in image-based systems biology, where per-cell estimates of the fluorescence evolution and its fluctuations are available. This enables quantifying cell–cell variability on the level of network parameters. The histogram of the parameters identified for different cells provides a biologically meaningful way of assessing phenotypic variability beyond simple differences in the fluorescence levels.

We have proposed a novel combination of a flexible Monte Carlo method, the GaA algorithm, and efficient exact stochastic simulation algorithms, the partial-propensity methods. The presented method can be used for global parameter optimization, approximate Bayesian inference under uniform prior, and ellipsoidal volume estimation of the viable parameter space. We have introduced an objective function that measures closeness between a single experimental trajectory and a single trajectory generated by the forward model. The objective function comprises a moment-matching and a time-autocorrelation part. This allows including experimental readouts from, e.g., fluorescence photometry and FCS.

We have applied the method to estimate the parameters of two monostable reaction networks from a single simulated temporal trajectory each, both at steady state and during transient phases. We considered the linear cyclic chain network and a non-linear colloidal aggregation network. For the linear model we were able to robustly identify a small region of parameter space containing the true kinetic parameters. In the non-linear aggregation model, we could identify several parameter vectors that fit the simulated experimental data well. There are two possible reasons for this reduced parameter identifiability: either GaA cannot find the globally optimal region of parameter space due to high ruggedness and noise in the objective function, or the non-linearity of the aggregation network modulates the kinetics in a non-trivial way [10,39]. Both cases are not accounted for in the current objective function, thus leading to reduced performance for non-linear reaction networks.

We also used GaA as an adaptive MCMC method for approximate Bayesian inference of the posterior parameter distributions in the linear chain network. This enabled estimating the volume of the viable parameter space below a given objective-function value threshold. We found these volume estimates to be stable

across independent runs. We thus believe that GaA might be a useful tool for exploring the parameter spaces of stochastic systems.

Future work will include (1) alternative objective functions that include temporal cross-correlations between species and the derivative of the autocorrelation; (2) longer experimental trajectories; (3) multi-stable and oscillatory systems; and (4) alternative global optimization schemes. Moreover, the applicability of the present method to large-scale, non-linear biochemical networks, and real-world experimental data will be tested in future work.

Acknowledgments RR was financed by a grant from the Swiss SystemsX.ch initiative (grant WingX), evaluated by the Swiss National Science Foundation. This project was also supported with a grant from the Swiss SystemsX.ch initiative, grant LipidX-2008/011, to IFS.

References

1. Albert R (2005) Scale-free networks in cell biology. J Cell Sci 118(21):4947–4957
2. Andrieu C, Thoms J (2008) A tutorial on adaptive MCMC. Stat Comput 18(4):343–373
3. Auger A, Chatelain P, Koumoutsakos P (2006) *R*-leaping: accelerating the stochastic simulation algorithm by reaction leaps. J Chem Phys 125:084103
4. Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2):101–113
5. Boys RJ, Wilkinson DJ, Kirkwood TBL (2008) Bayesian inference for a discretely observed stochastic kinetic model. Stat Comput 18(2):125–135
6. Cardinale J, Rauch A, Barral Y, Székely G, Sbalzarini IF (2009) Bayesian image analysis with on-line confidence estimates and its application to microtubule tracking. In: Proc. IEEE Int Symp Biomedical Imaging (ISBI). IEEE, Boston, USA, pp 1091–1094
7. Cinquemani E, Milias-Argeitis A, Summers S, Lygeros J (2008) Stochastic dynamics of genetic networks: modelling and parameter identification. Bioinformatics 24(23):2748–2754
8. Gillespie DT (1992) A rigorous derivation of the chemical master equation. Phys A 188:404–425
9. Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys 115(4):1716–1733
10. Grima R (2009) Noise-induced breakdown of the Michaelis–Menten equation in steady-state conditions. Phys Rev Lett 102(21):218103 DOI 10.1103/PhysRevLett.102.218103
11. Hafner M, Koeppl H, Hasler M, Wagner A (2009) 'Glocal' robustness analysis and model discrimination for circadian oscillators. PLoS Comput Biol 5(10):e1000534
12. Helmuth JA, Burckhardt CJ, Greber UF, Sbalzarini IF (2009) Shape reconstruction of subcellular structures from live cell fluorescence microscopy images. J Struct Biol 167:1–10
13. Helmuth JA, Sbalzarini IF (2009) Deconvolving active contours for fluorescence microscopy images. In: Proc Int Symp Visual Computing (ISVC) (Lecture notes in computer science), vol 5875. Springer, Las Vegas, USA, pp 544–553
14. Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106(4):620–630 DOI 10.1103/PhysRev.106.620
15. Kitano H (2002) Computational systems biology. Nature 420(6912):206–210
16. Kitano H (2002) Systems biology: a brief overview. Science 295(5560):1662–1664
17. Kjellström G (1969) Network optimization by random variation of component values. Ericsson Tech 25(3):133–151
18. Kjellström G (1991) On the efficiency of Gaussian Adaptation. J Optim Theor Appl 71(3):589–597

19. Kjellström G, Taxen L (1981) Stochastic optimization in system design. IEEE Trans Circ Syst 28(7):702–715
20. Koeppl H, Setti G, Pelet S, Mangia M, Petrov T, Peter M (2010) Probability metrics to calibrate stochastic chemical kinetics. In: Proc IEEE Int Symp Circuits and Systems, Paris, France, pp 541–544
21. Koutroumpas K, Cinquemani E, Kouretas P, Lygeros J (2008) Parameter identification for stochastic hybrid systems using randomized optimization: a case study on subtilin production by *Bacillus subtilis*. Nonlin Anal Hybrid Syst 2(3):786–802
22. Kurtz TG (1972) Relationship between stochastic and deterministic models for chemical reactions. J Chem Phys 57(7):2976–2978
23. Lakowicz JR (2006) Principles of fluorescence spectroscopy. Springer USA DOI 10.1007/978-0-387-46312-4
24. Ljung L (2002) Prediction error estimation methods. Circ Syst Signal Process 21(1):11–21
25. Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. Proc Natl Acad Sci USA 100(26):15324–15328
26. Mason O, Verwoerd M (2007) Graph theory and networks in biology. Syst Biol IET 1(2):89–119
27. Müller CL (2010) Exploring the common concepts of adaptive MCMC and Covariance Matrix Adaptation schemes. In: Auger A, Shapiro JL, Whitley D, Witt C (eds.) Theory of evolutionary algorithms, Dagstuhl Seminar Proceedings, no. 10361. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany, Dagstuhl, Germany URL http://drops.dagstuhl.de/opus/volltexte/2010/2813
28. Müller CL, Sbalzarini IF (2010) Gaussian Adaptation as a unifying framework for continuous black-box optimization and adaptive Monte Carlo sampling. In: Proc IEEE Congress on Evolutionary Computation (CEC). Barcelona, Spain, pp 2594–2601
29. Müller CL, Sbalzarini IF (2010) Gaussian Adaptation revisited – an entropic view on covariance matrix adaptation. In: Proc EvoStar (Lecture notes computer science), vol 6024. Springer, Istanbul, Turkey, pp 432–441
30. Müller CL, Sbalzarini IF (2011) Gaussian Adaptation for robust design centering. In: Proc EuroGen Int Conf Evolutionary and Deterministic Methods for Design, Optimization and Control. Capua, Italy, pp 736–742
31. Munsky B, Trinh B, Khammash M (2009) Listening to the noise: random fluctuations reveal gene network parameters. Mol Sys Biol 5(1):318
32. Poovathingal SK, Gunawan R (2010) Global parameter estimation methods for stochastic biochemical systems. BMC Bioinformatics 11(1):414
33. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16(12):1791–1798
34. Qian H, Elson EL (2004) Fluorescence correlation spectroscopy with high-order and dual-color correlation to probe nonequilibrium steady states. Proc Natl Acad Sci USA 101(9):2828–2833
35. Ramaswamy R, González-Segredo N, Sbalzarini IF (2009) A new class of highly efficient exact stochastic simulation algorithms for chemical reaction networks. J Chem Phys 130(24):244104
36. Ramaswamy R, Sbalzarini IF (2010) Fast exact stochastic simulation algorithms using partial propensities. In: Proc ICNAAM, numerical analysis and applied mathematics, international conference. AIP, Rhodes, Greece, pp 1338–1341
37. Ramaswamy R, Sbalzarini IF (2010) A partial-propensity variant of the composition-rejection stochastic simulation algorithm for chemical reaction networks. J Chem Phys 132(4):044102
38. Ramaswamy R, Sbalzarini IF (2011) A partial-propensity formulation of the stochastic simulation algorithm for chemical reaction networks with delays. J Chem Phys 134:014106
39. Ramaswamy R, Sbalzarini IF, González-Segredo N (2011) Noise-induced modulation of the relaxation kinetics around a non-equilibrium steady state of non-linear chemical reaction networks. PLoS ONE 6(1):e16045

40. Reinker S, Altman RM, Timmer J (2006) Parameter estimation in stochastic biochemical reactions. IEE Proc Syst Biol 153(4):168
41. Stock G, Ghosh K, Dill KA (2008) Maximum Caliber: a variational approach applied to two-state dynamics. J Chem Phys 128(19):194102
42. Strogatz SH (2001) Exploring complex networks. Nature 410:268–276
43. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface 6(31):187–202
44. Wolkenhauer O (2001) Systems biology: the reincarnation of systems theory applied in biology? Briefings Bioinform 2(3):258
45. Zechner C, Pelet S, Peter M, Koeppl H (2011) Recursive Bayesian estimation of stochastic rate constants from heterogeneous cell populations. In: Proc 50th IEEE CDC, Conference on Decision and Control. Orlando, Florida, USA