Research Article   **Open Access**

Parameter Estimation for Stochastic Models of Biochemical Reactions

Christoph Zimmer* and Sven Sahle

BioQuant, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

Abstract

Parameter estimation is very important for the analysis of models in systems biology. Stochastic models are of increasing importance. However, parameter estimation of stochastic models is still in the early phase of development and there is need for efficient methods to estimate model parameters from time course data, which is intrinsically stochastic, only partially observed and has measurement noise.

In this article, a fast and efficient method that is well established in the field of parameter estimation for systems of Ordinary Differential Equations (ODE) is adapted to stochastic models. The focus is on the objective function which is shown to have advantageous properties, that make it directly applicable to problems in systems biology. The proposed method can deal with stochastic systems, where the behaviour qualitatively differs from the corresponding deterministic description. It works with measurements from a single realization of the stochastic process, and with partially observed processes, including measurement errors. The objective function is deterministic, therefore a wide range of optimization methods, from derivative based methods to global optimization to Bayesian techniques can be applied. The computational effort required is comparable to similar methods for parameter estimation in deterministic models. To construct the objective function, a multiple shooting procedure is used, in which the continuity constraints are relaxed to allow for stochasticity. Unobserved states are treated by enlarging the optimization vector and using resulting values from the forward integration. Test functions are suggested that allow to monitor the validity of the approximations involved in this approach. The quality of the method is evaluated for some example models, with a statistic of 50 estimates from 50 stochastic realizations. It is shown that the method performs well compared to established approaches.

**Keywords:** Parameter estimation; Calcium oscillation model

Introduction

Parameter estimation is very important for the analysis of models in systems biology. Computational modeling is a central approach in systems biology, for studying increasingly complex biochemical systems. Progress in experimental techniques, e.g. the possibility to measure small numbers of molecules in single cells [1], highlights the need for stochastic modeling approaches. Simulation methods for stochastic processes are being developed for decades since [2], and nowadays exist with a lot of variants [3]. Parameter estimation methods for stochastic models, however, are still in the early phase of development.

We have identified a number of desirable properties that a parameter estimation procedure should have, in order to be useful in actual biological applications: It should work with measurements from few or even a single realization of the physical process, i.e. it should not require that enough measurements are made to reconstruct reliable probability distributions, or other statistical measures for all time points. It should work in cases where only partial measurements of the systems' state are possible, obviously also including measurement errors. It should not assume that the stochastic system approaches the large volume limit, i.e. the approach should also be valid, if the stochastic system exhibits a behavior that is qualitatively different from what would be covered by a deterministic model. On a more technical side, it is beneficial if the actual target function is deterministic (and differentiable), and if it can be computed efficiently. This means that a wide range of numerical optimization techniques can be employed, including gradient based local methods, numerous global optimization schemes, or Bayesian approaches.

The problem for parameter estimation for stochastic models can be written in the following form: We assume that the experimental or simulated data measures at each time point t_i, $i=1,\dots,n$, the number

of molecules of species v_i in the system, hence $v_i \in \mathbb{N}_0$. Each data set $v=(v_1,\dots v_n)$ is stochastic. This means that even two simulated realization of the system without measurement noise can be different, see figure 1. Figure 1 does also indicate that the mean of stochastic simulations is not close to the ODE solution. Furthermore, it is assumed that the systems behavior depends on some unknown parameters $\theta \in \mathbb{R}^q$ which are to be estimated. To this goal, the data v is compared to simulation with respect to some objective function F, which measures the quality of the fit. Parameter estimation means finding the optimal value $\hat{\theta}$ for
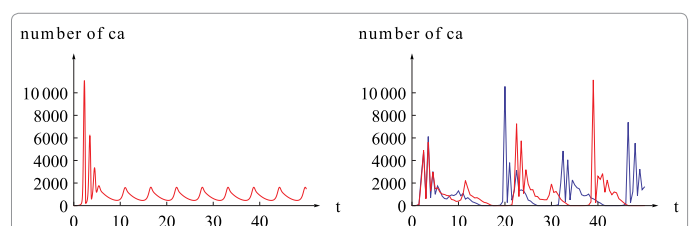
Figure 1: ODE and stochastic simulation of the Calcium oscillation model. The graphics on the left side shows an ODE simulation of the Calcium oscillation model. The graphics on the right side shows two stochastic realizations of the Calcium oscillation system. Parameters and initial values as given in the results section.

*Corresponding author: Christoph Zimmer, BioQuant, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany, E-mail: christoph.zimmer@bioquant.uni-heidelberg.de

the parameter. For models of ODE, the choice of the objective function has been widely discussed [4], and in case of normally distributed measurement error, the common choice is a least squares function. The focus of this article will be the objective function *F*, for stochastic models.

Approaches exist for time series data using stochastic simulations. Due to the Markov property of the time series, the likelihood function factorizes into the product of transition probabilities. These transition probabilities are generally unknown in stochastic modeling. They can be estimated using stochastic simulations. This can be done with density estimation methods [5,6]. Another approach is the use of a reversible jump algorithm [7]. The parameter estimation is then performed with Bayesian methods. An alternative to that is the use of a stochastic gradient descent [8], and the use of a reversible jump Markov chain, Monte Carlo method, for the estimation of the transition probabilities. Using a surrogate probabilistic model as an approximation, is faster from a computational point of view [9]. Another approximation is suggested in form of an approximate maximum likelihood method [10], where also a singular value decomposition likelihood method is described. A use of approximate Bayesian methods is suggested [11].

A second class of methods focuses on a numerical solution of the Chemical Master Equation (CME), which describes the probability for each state, as a function of the time. These systems are generally high dimensional. To address this problem, a state space truncation can be used [12], or moment-closure methods, which are an approximation focusing on a finite number of moments of the probability distribution [13,14]. Parameter estimation with the CME and an approximation for the likelihood function for small systems, or a hybrid method for large systems is suggested in [15]. Deuflhard et al. [16] and Engblom [17] use an adaptive Galerkin method for the solution of the CME. If distribution information is available from measurement, a finite state projection [18], can be used to solve the CME without simulations. The common challenge is the fact that the solution of the CME, as well as simulation-based methods become very time-consuming, as the number of states in the state space becomes larger.

Moment matching methods use repeated measurements to estimate moments of the distribution for the parameter estimation [19,20]. Further approaches can be taken from stochastic epidemic models, and are based on expectation maximization algorithms or Martingale based approaches [21]. In the case of large numbers of molecules, there are theoretical results available, describing the properties of a least-squares estimate [22].

This article suggests a new method for estimating parameters of stochastic models. The method is able to cope with data from a single stochastic trajectory, which only measures some of the species in the system at discrete time points. The behavior of this single trajectory might be well different from the mean or an ODE solution (Figure 1). The approach is based on a method proposed by Bock [23], for the parameter estimation in systems of Ordinary Differential Equations (ODE), further developed [24], and already successfully applied to deterministic systems with chaotic behavior [25,26]. The method can tackle models with fully observed and partially observed data sets. The fact that it works without stochastic simulations and without solving a high dimensional CME means that it is possible to approach systems of a size, as large as realistic models being tackled with ODEs. As the the evaluation of the objective function and the numerical optimization procedure can be separated, the article focuses on the objective function. For the numerical optimization of this deterministic function,

all methods from derivative based methods to global optimization techniques to Bayesian methods can be applied. These techniques then also allow for important other features, such as confidence intervals for the parameters or experimental design. The article restricts itself to the test of the objective function, especially with respect to different stochastic realizations. The objective function is based on short time ODE integration and performs successfully, even in models which behave qualitatively different modeled stochastically (see [27] for an example of such a model). The advantage is very high speed, since neither solving a high dimensional CME system nor lots of stochastic simulations are required.

As the objective function is deterministic, the estimator for a single stochastic time course is unique. Using the same model with the same parameters and initial conditions, the estimation will result in a different value due to the intrinsic stochasticity. To test the quality of the objective function, it is therefore, necessary to estimate more than one stochastic realization of the same model, with the same parameters and initial conditions. This article uses 50 different realizations for each model. A statistic over the 50 estimations can then be used to asses the quality of the objective function. It has to be underlined that the 50 realizations are only used to test the quality of the objective function. For an estimation one realization is sufficient.

The method is described with equidistant time points of measurements here, for simplicity. It is possible to apply it without changes to non equidistant time points of measurements, which is very important for the applicability of optimum experimental design. Due to its structure, the method is easily able to handle measurement noise ,although it will of course reduce information. As the method uses an ODE approximation on short time intervals for stochastic data, test functions will be suggested to check how well this approximation works, and it will be demonstrated that this is not problematic, even in models with irregular stochastic oscillations. Furthermore, it is demonstrated that the method can even be successful in models which are structurally not identifiable, using single shooting ODE methods. As the objective function of the method is completely deterministic it depends on the user's choice, whether to apply derivative based methods or methods without derivatives for the optimization.

The article is structured as follows:

The method section will show how the objective function is composed. This objective function will be extended to partially observed models.

The results section will show the performance of the objective function for different models: first an Immigration-Death model, which is on the one hand an instructive example, but allows on the other hand to compare the performance of the method to an exact analytical solution. Second is a Lotka-Volterra model also treated by Boys et al. [7], followed by a Calcium oscillation model [27]. To evaluate the method, it is applied to 50 different data sets of each model and statistics of the 50 resulting estimates are presented. The conclusion section will give some concluding remarks and an outlook.

Method

Stochastic models describe the systems dynamics with a discrete continuous time Markov jump process. The time dependent probabilities of this process are described by the CME. Stochastic modeling can lead to system's behavior, which is qualitatively different from the behavior of the ODE-modeled system. Qualitatively, different

means that it is not just the behavior of the ODE system plus some noise. This can be illustrated by the Calcium oscillation model of Kummer et al. [27]. This suggests that the methods for parameter estimation in ODE models might not be suited for stochastic models.

It is assumed that the experimental or simulated data measures at each time point t_i, $i=1,...,n$, the number of molecules of species v_i in the system, hence $v_i \in \mathbb{N}_0$. Furthermore, it is assumed that the systems behavior depends on some unknown parameters $\theta \in \mathbb{R}^q$ which are to be estimated. To this goal, the data v is compared to simulation, with respect to some objective function F, which measures the quality of the fit. Parameter estimation means finding the optimal value $\hat{\theta}$ for the parameter:

$$\hat{\theta} = \arg\min_\theta F(v,\theta)$$

As mentioned in the introduction for ODE models, the choice of the objective function has been widely discussed. The focus of this article will be the objective function F for stochastic models.

Due to the Markov property, the likelihood function factorizes into the product of transition probabilities:

$$L(v,\theta) = \prod_{i=1}^{n} p_\theta(v_i, t_i \mid v_{i-1}, t_{i-1})$$

The transition probability p_θ is generally not known, and might either be estimated by means of simulations, or calculated by using the solution of a high dimensional CME system, both of which is very time consuming. The approach approximates the system on the short time interval $[t_{i-1}, t_i]$ with an ODE model. Short in this context means relatively short with respect to the systems dynamics, as an experiment should be designed in a way that it covers the systems dynamics. The fact that the approximation is done only on a very short time interval is crucial. An ODE approximation requires stronger assumptions than a linear noise approximation, but as this approximation only has to hold on a short time interval, it is much less restrictive than the usual linear noise approximation [28]. Further advantages of the relatively short integration interval have already been described in detail for parameter estimation for ODEs [29]: a prior information on state variables from measurement can be included in the multiple shooting scheme. This damps influence of parameter guesses. Numerical stability is increased as splitting the intervals reduces error propagation.

Fully Observed Case

The proposed approach is motivated by the methods for parameter estimation in ODE systems introduced by Bock [23], and uses short time ODE integration: starting at time point t_{i-1} at state v_{i-1}, the initial value problem of the systems of ODEs is solved with initial value, v_{i-1} at time t_{i-1} until time t_i. The result, $h(t_i, \theta, v_{i-1}, t_{i-1})$, is compared to the data point v_i, and the residuum is defined as

$$\varepsilon_i = v_i - h(t_i, \theta, v_{i-1}, t_{i-1})$$

which for the model leads to the description

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} h(t_1, \theta, v_0, t_0) \\ h(t_2, \theta, v_1, t_1) \\ \vdots \\ h(t_n, \theta, v_{n-1}, t_{n-1}) \end{pmatrix} + \begin{pmatrix} \in_1 \\ \in_2 \\ \vdots \\ \in_n \end{pmatrix}$$

If the residuals ε are independent normally distributed random

variables with mean zero and constant variance, the least squares estimator

$$\hat{\theta} = \arg\min_\theta F(v,\theta)$$

with

$$F(v,\theta) = \sum_{i=1}^{n} \| h(t_i, \theta, v_{i-1}, t_{i-1}) - v_i \|_2^2 \qquad (1)$$

is a Maximum Likelihood estimator. In general, the distribution for a time point in a stochastic model is not known, and not necessarily Gaussian. Hence, theoretically the properties of a Maximum Likelihood estimator can not be guaranteed. In practice, it is possible to test if the $\hat{\varepsilon}_i$ perform approximately like independent normally distributed random variables, with mean zero. If this is the case, the estimator might still be quite powerful. Later in this section, test functions will be suggested which describe the properties of the residuals. The structure of equation (1) allows handling normally distributed measurement noise, as well without loosing the desired properties.

As equation (1) corresponds to the multi-experiment setting, described by Schlöder and Bock [24], it is possible to use the efficient methods suggested there to solve the optimization problem.

Partially Observed Models

Assume that only the first d components of the D dimensional vector of species v can be observed. At time t_0, the unobserved states $v_0^{(d+1)},...,v_0^{(D)}$ are also used as optimization variables now. In fact, the unobserved states are discrete numbers, but for the optimization purpose, this condition is relaxed and they are optimized as real numbers. For the solution of the initial value problem on $[t_{i-1}, t_i]$ instead of the full measurement v, use the observed states $v_{i-1}^{(1)},...,v_{i-1}^{(d)}$ from the measurement and for the unobserved states, the result of the initial value problem on the previous time interval, hence $h^{(d+1,...,D)}(t_{i-1}, \theta, \cdot \mid t_{i-2})$. It is possible to enlarge the optimization vector even more, and to include also further unobserved states. Formally, define an index set $\{t_0\} \subset K$ which contains all time points at which the unobserved states will be included in the optimization variable. Denote with $v_K^{(j)} = (v_j^{(d+1)},...,v_j^{(D)})$ the unobserved states in the optimization vector, at time t_j and with $v_K = (v_K^{(j)})_{j \in K}$, the union of unobserved states at different time points. Now define the completion of the observed measurement as $\tilde{v}$ with

$$\tilde{v}_j = \begin{cases} v_j^{(1)},...,v_j^{(d)}, v_K^{(j)} & :t_j \in K \\ v_j^{(1)},...,v_j^{(d)}, h^{(d+1,...D)}(t_j, \theta, \tilde{v}_{j-1}, t_{j-1}) & :t_j \notin K. \end{cases}$$

Then again, as in the fully observed case, a distance measure is used to compare the result of the integration with the data point:

$$F_K(v,\theta,\theta_K) = \sum_{i=1}^{n} \left\| h^{(1,...,d)}(t_i, \theta, \tilde{v}_{i-1}, t_{i-1}) - v_i^{(1,...,d)} \right\|_2^2 \qquad (2)$$

The case $K=\{t_0\}$ means that just the first unobserved state is included in the optimization, and the case $K=\{t_0, t_1,...,t_n\}$ means that all unobserved states are included in the optimization. The results section will show examples for different K.

Comments on the Approximation

As the method uses an approximation, it is very important to investigate if a model satisfies the approximation. Therefore, this subsection suggests test functions to see how well the approximation works. The mean of the residuals $\bar{\epsilon} = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i$ can be calculated. If the model is well approximated, this should be small, in comparison with the range of the residuals.

To see if the residuals are approximately normally distributed, calculate the Kullback-Leibler divergence of a density estimate from the ϵ_i and a centered normal distribution of variance σ_{KL}^2 restricted to the support of the density estimate. The Kullback-Leibler divergence is then minimized over σ_{KL}^2. If the Kullback-Leibler divergence for the optimal σ_{KL}^2 is small, this means that the ϵ_i can be approximated well by a normal distribution with constant variance. The point that the variance is constant is important, as it is not possible to estimate the variance from only one observation per time point.

Instead of calculating the correlation between the time points, it will be calculated how long it takes until the residuals are uncorrelated. Estimate the autocorrelation of the residuals:

$$\hat{R}(k,\epsilon) = \frac{1}{(n-k)\hat{\sigma}_{\epsilon}^2} \sum_{i=1}^{n-k} (\epsilon_{t_i} - \bar{\epsilon})(\epsilon_{t_{i+k}} - \bar{\epsilon})$$

where k represents the time step and $\hat{\sigma}_{\epsilon}^2$ an estimate for the variance, within the residuals ε. If ε is such that $\tilde{k} := \min(k \mid \hat{R}(k,\epsilon) < 0) > 0$ exists define the autocorrelation time as $act(\epsilon) = t_{\tilde{k}}$. If the total horizon of measurements $t_n - t_0$ is only of the same size as the autocorrelation time of the residuals $act(\epsilon)$ his indicates correlated residuals, which indicate that the approximation might not be suitable. Furthermore, it is an important question for the quality of the estimation how much information can be found with the MSS method in the (intrinsic) noisy system. A signal to noise ratio is defined componentwise as

$$SNR(v^{(l)},\theta) = \frac{\sum_{i=1}^{n} \left| h^{(l)}(t_t,\theta,v_{i-1},t_{i-1}) - v_{i-1}^{(l)} \right|}{\sum_{i=1}^{n} \left| \epsilon_i^{(l)} \right|} \quad for \ l=1,\dots,d$$

The higher the *SNR* value is the more information is contained in the data. If the *SNR* is small (<1), many measurements are needed.

Optimization

In case of the fully observed system, the dimension of the optimization problem is exactly the number of parameters, namely q. In case of the partially observed system, the dimension of the optimization problem is the number of parameters plus the number of unobserved species times, the number of time points included in the optimization, $q+(D-(d+1)) \times length(K)$.

As the objective function is completely deterministic–meaning, it is calculated without the use of stochastic simulations–the optimization procedure is the same, as in parameter estimation for ODE systems: it is possible to apply derivative based methods [29], or global methods [30], as well as including it in an approximate Bayesian framework, for example [11].

Certainly the optimization of (2) with larger K can be much more challenging than with $K=\{0\}$, due to the increased dimensionality of the optimization problem. However, the numerical optimization techniques, especially questions of local minima and identifiability issues, as well as the question, whether derivative based or global method are preferable shall not be the focus of this article. Nevertheless, these points are of importance and suggested for further research.
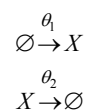
Results

The result section will test the performance of the MSS objective function, with respect to the stochastic realizations. To this aim, 50 stochastic realizations for each scenario are obtained from simulations, using the Gillespie method [2], with the software COPASI [31]. It is important to note that each of the stochastic realizations might look differently. For each single time course, data set the MSS function is optimized and a parameter estimate obtained. To investigate the influence of the intrinsic stochasticity, the mean of the 50 estimates is calculated to see whether the MSS functional is biased. The standard deviation of the 50 estimates is calculated to see how much the estimates are spreading around the mean, due to the stochasticity and the mean relative error for simpler comparison. These values will all be shown in result tables.

Calculating confidence intervals for each single estimate is desirable as well. In general, this is possible with the MSS function, applying more advanced numerical optimization techniques, as mentioned in the methods section. In connection with optimal design of experiment, this is planned for further research. As the focus of the article is the investigation of intrinsic stochasticity, data is assumed to be noise free. The Lotka-Volterra model shows that adding measurement noise leads to the expected behavior–namely that the accuracy of the estimates reduces with increasing level of noise. Very important for practical applicability are partially observed models, which are treated in detail for the models.

Immigration-Death Model

The first example is an Immigration-Death model:

$$\varnothing \xrightarrow{\theta_1} X$$
$$X \xrightarrow{\theta_2} \varnothing$$

where X is the substance and θ_1, θ_2 are parameters with a representation in ODEs

$$\frac{dx}{dt} = \theta_1 - \theta_2 x, \ x(0) = x_0$$

The investigation of this system is instructive, as it allows a comparison between the performance of the MSS functional (1), and the result with the exact transition probability, which can only be calculated exactly in very simple cases.

The exact transition probability from a state v_{i-1} at time point t_{i-1} to a state v_i at time point t_i can be calculated as a solution of the CME, or with the probability generating function [32].

To evaluate the performance of the MSS functional (1), it is compared to an estimation using Exact transition probabilities (EL). To that aim, a statistics is calculated of 50 data sets, which are obtained from simulations using the Gillespie method [2], with the software COPASI [31]. The initial condition is always the steady state of the system. This scenario is of special interest as the system is structurally unidentifiable, using traditional ODE methods. Therefore, it is also the most difficult scenario for parameter estimation. For each of the data sets, the MSS objective function and the objective function with

the exact transition probabilities are optimized using the software Mathematica [33]. Then, the mean and the standard deviation of both estimators, as well as the relative error are calculated. This procedure is done for different parameters and designs. The results are shown in table 1. To check the assumptions, the mean and autocorrelation time of the residuals, as well as the *SNR* are calculated (The letters in brackets indicate the chosen design, see for information table 1):

$$(A): \overline{\in}=1\cdot10^{-8}, \hat{\sigma}_{KL}=1, \ KL:0.14, \ act(\in)=1.5, \ SNR:0.21$$

$$(B): \overline{\in}=9\cdot10^{-9}, \hat{\sigma}_{KL}=1, \ KL:0.18, \ act(\in)=1, \ SNR:0.17$$

$$(C): \overline{\in}=4\cdot10^{-8}, \hat{\sigma}_{KL}=2.6, \ KL:0.01, \ act(\in)=10, \ SNR:0.50$$

$$(D): \overline{\in}=1\cdot10^{-8}, \hat{\sigma}_{KL}=2.7, \ KL:0.03, \ act(\in)=20, \ SNR:0.55$$

While all other test functions give acceptable values, the SNR is very small in table 1 (A,B). This indicates that the systems dynamics is not well approximated by the MSS method. Using 100 observations does in this situation not give enough information, and leads to a biased estimate. Using many more observations as in table 1B resolves the problem.

The results given in table 1 lead to two conclusions: An estimation is possible and unbiased, if there are enough measurements, table 1 (B,C). If the trajectory is very short table 1A, the estimator might be biased. The reason for that is that for a low *SNR*, more measurements are needed to sum up enough information for the estimation-see figure 2, which for each interval $[t_{i-1}, t_i]$ shows the ODE dynamics, in form of the solution of the corresponding initial value problem $h(t_i, \theta, v_{i-1}, t_{i-1})$ as well as the residuals (dotted red line). One can see that for this situation, the system's dynamics are not well represented by the ODE solutions.

The exact method, which is only possible in this simple example model, makes much better use of the intrinsic fluctuations, and therefore, results in more accurate estimates. The mean and autocorrelation time

of the residuals behave very well, with respect to the comments on the residuals given in the methods section.

This example is a proof of concept, example as it shows that an estimation with the MSS method is possible, even in a case which would be structurally unidentifiable for traditional methods [30].

The results of table 1D can be compared to Wang et al. [8], where a different method is applied to the same setting. As there is only the result of one estimation procedure given a comparison with respect to unbiasedness and variance is not possible. However, 4/5 of the relative errors of table 1D are smaller than the relative error of Wang et al. [8].

Table 1 (B-D) also suggest that using this methods models become identifiable, which are unidentifiable using traditional ODE methods.

Lotka-Volterra Model

Second example which is still small, but allows investigation of behavior in partially observed models is a Lotka-Volterra model:

$$Y^{(1)} \xrightarrow{\theta_1} 2Y^{(1)}$$
$$Y^{(1)} + Y^{(2)} \xrightarrow{\theta_2} 2Y^{(2)} \qquad (3)$$
$$Y^{(2)} \xrightarrow{\theta_3} \varnothing$$

where $Y^{(1)}$ is the prey and $Y^{(2)}$ the predator and $\theta_1, \theta_2, \theta_3$ parameters. The first reaction of (3) is the prey reproduction, the second the predator reproduction, and the third is the predator death. In terms of ODEs, this system reads as

$$\frac{d[Y^{(1)}]}{dt} = \theta_1[Y^{(1)}](t) - \theta_2[Y^{(1)}](t)[Y^{(2)}](t)$$

$$\frac{d[Y^{(2)}]}{dt} = \theta_2[Y^{(1)}](t)[Y^{(2)}](t) - \theta_3[Y^{(2)}](t).$$

	Estimation Results				Rel err		Test function SNR
	EL		MSS		EL	MSS	
(A) 100 observations, Δt=0.5, $\Theta^{(0)}$ =(1,0.1)							
Θ_1	1.00	±0.19	1.56	±0.92	15%	71%	SNR: 0.21
Θ_2	0.101	±0.02	0.156	±0.09	16%	72%	
(B) 2000 observations, Δt=0.5, $\Theta^{(0)}$ =(1,0.1)							
Θ_1	1.00	±0.04	1.00	±0.15	2%	13%	SNR: 0.17
Θ_2	0.101	±0.00	0.101	±0.015	2%	13%	
(C) 500 observations, Δt=5, $\Theta^{(0)}$ =(1,0.1)							
Θ_1	1.01	±0.09	1.02	±0.11	7%	9%	SNR: 0.50
Θ_2	0.101	±0.01	0.102	±0.01	7%	9%	
(D)100 observations, Δt=10, $\Theta^{(0)}$ =(0.6,0.06)							
Θ_1	0.60	±0.12	0.65	±0.17	15%	21%	SNR: 0.55
Θ_2	0.061	±0.01	0.065	±0.018	16%	22%	

Table 1: Statistics of the estimation results for Immigration-Death model. The table shows that the MSS method works well in this example for three out of four experimental designs and the test functions identify the problematic design.

50 data sets are simulated using the Gillespie algorithm with initial value v_0=10. For each of the 50 data sets an estimation is performed with an exact method (EL) and the MSS method. The table shows a statistic over the 50 estimates for both the exact and the MSS method: Parameter name (column 1), averages for exact method (column 2), standard deviation of exact method (column 3), averages of MSS method (column 4), standard deviations of MSS method (column 5), relative errors of exact method (column 6) and MSS method (column 7) and the signal to noise (SNR) test function (column 7).

The SNR test function indicates a weak SNR in situation (A) and therefore identifies the case in which the MSS approximation is problematic. If this is the case many more observations are needed to resolve the problem despite still low SNR, situation (B), or another experimental design (C).

This chapter also includes an investigation of the estimation with noisy data.

Fully observed case

To investigate the behavior of the MSS objective functional (1), a statistics is calculated from 50 data sets which are simulated with the Gillespie algorithm, using the software COPASI [31]. The true parameter was $\theta^{(0)}$ =(0.5, 0.0025, 0.3) and 40 observations were taken with Δt=1. The initial conditions were $(v_0^{(1)}, v_0^{(2)})$ =(71, 79). The setting is chosen in a way such that it is identically with Boys et al. [7]. A comparison of the results will be given later in this section.

Measurement noise is simulated as follow: For each time point, a normally distributed random variable is generated with the given variance. Then, it is rounded to the next integer. This is done because measurements are assumed to be integer counts. This is not necessary for theoretical reasons or performance of the estimation. For σ=10 or σ=25 respectively, some negative measurements occur due to the measurement error. For σ=10, these are not corrected, which shows that the method can even handle negative data to some amount. For σ=25 measurements which are negative are set to zero, which leads to an measurement error, which is not centered around zero. The results show that the method is still able to handle that situation. Note that setting negative measurements to zero only implies the simulated ODE dynamics of the next time interval. For all later time intervals, the ODE is again initialized with a measurement, which well can be larger
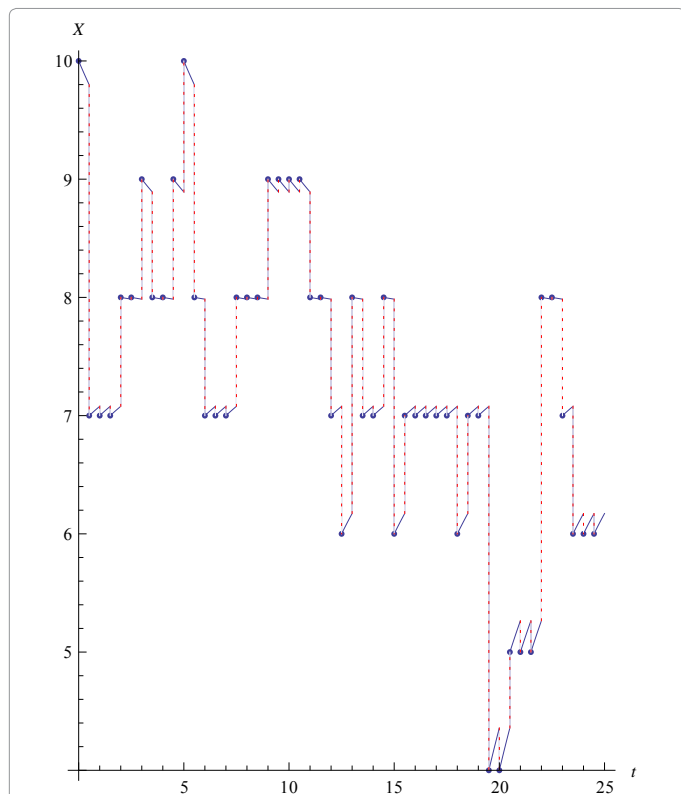
Figure 2: Representation of Immigration-Death system's dynamics with MSS method. The blue points are the data points. Blue curves are the ODE dynamics for each time interval [t_{i-1},t] namely $h(t, \hat{\Theta}, v_{i-1}, t_{i-1})$ with the estimated parameter. Red dotted lines show the residuals. 100 simulated observations from Immigration-Death model with Θ=(1,0.1) and T=50 for the estimation of $\hat{\Theta}$, dynamics shown until T=25.

	Estimation Results		Rel err
	exact measurements		
Θ_1	0.501	±0.016	2.5%
Θ_2	0.00250	±7·10⁻⁵	2.2%
Θ_2	0.301	±0.011	3.1%
	noise: σ=10		
Θ_1	0.490	±0.019	3.2%
Θ_2	0.00248	±9·10⁻⁵	2.9%
Θ_2	0.302	±0.012	3.4%
	noise: σ=25		
Θ_1	0.454	±0.031	9.7%
Θ_2	0.00243	±15·10⁻⁵	5.2%
Θ_2	0.301	±0.021	5.6%

Table 2: Statistics of the estimation results for Lotka-Volterra model. 50 data sets are simulated using the Gillespie algorithm with 40 observations with Δt=1 and true parameter Θ$^{(0)}$=(0.5, 0.0025,0.3) and v_0=(71,79). For each of the data sets an estimation is performed with the MSS method. The table shows a statistic of the 50 estimates: Parameter name (column 1), averages of estimates (column 2), standard deviation of estimates (column 3) and relative errors (column 4).

zero again. Further studies (without measurement noise) for cases in which the species die out show an acceptable estimation. Of course, the amount of information for the estimation is depending on the time point of die out.

For all data sets, the objective function is optimized with the software Mathematica [33]. Table 2 gives averages and standard deviation for the 50 estimation results using (1). For each estimation result, the relative error is calculated. The test functions give the following results: for exact measurements $\overline{\in}$ =(0.16, 0.07), $\hat{\sigma}^2_{KL}$ =(12, 11) KL:(0.2, 0.1), $act(\in)$=(2, 2) and SNR:(4.3, 5.8). For measurements with noise level σ=10, they are $\overline{\in}$ =(0.89, 0.62), $\hat{\sigma}^2_{KL}$ =(19,18), KL:(0.1, 0.1), $act(\in)$=(1,1) and SNR:(2.2, 2.9). For noise level σ=25, they are $\overline{\in}$ =(3.48, 2.52), $\hat{\sigma}^2_{KL}$ =(40, 40), KL:(0.04, 0.04), $act(\in)$=(1, 1) and SNR:(1.1,1.3). For σ=25, the mean of the residuals is not close to zero, but considering the means of the 10%− and 90%− quantiles, −90 and 47, one sees that the residuals still are small. This shows that the approximation with the MSS method works well in this model.

The MSS method estimates parameters very well in this example. The relative error of the estimation is in the range of the relative error of the method proposed in Boys et al. [7], where only one data set is estimated. The mean of the residuals is close to zero and the autocorrelation time small, compared to the total duration of observations. The signal to noise ratio is much better than in the Immigration-Death case.

Partially observed Lotka-Volterra model

Now assume that only prey can be observed. As in the completely observed case simulated data with true parameter $\theta^{(0)}$ =(0.5, 0.0025, 0.3) and 40 observation with Δt=1 is used. Initial condition is again $(v_0^{(1)}, v_0^{(2)})$ =(71,79), but as only prey can be observed, only $v_0^{(1)}$ =71 will be used for parameter estimation. The optimization variable is now $(\theta_1, \theta_2, \theta_3, v_0^{(2)})$. Table 3 gives averages and standard deviation for the 50 estimation results using (2) with K={0}. For each estimation result, the relative error is calculated. The estimation performs still quite well, even if only one species is observed. The values of the test functions

are for exact measurements $\bar{\bar{\epsilon}}$ =-0.1, $\hat{\sigma}^2_{KL}$ =12, KL:0.1, $act(\in)$=2 and SNR:4.4. For noise with σ=10, they are $\bar{\bar{\epsilon}}$ =0.7 $\hat{\sigma}^2_{KL}$ =20, KL:0.04, $act(\in)$=1 and SNR:2.3. For noise with σ=2 they are $\bar{\bar{\epsilon}}$ =3.1, $\hat{\sigma}^2_{KL}$ =19, KL:0.04, $act(\in)$=1 and SNR:1.1. The residuals are still approximately normally distributed as in the fully observed case, see also figure 3.

The same model is used for parameter estimation in stochastic models from Boys et al. [7]. For comparison in this article, the same true parameter values and initial conditions are used. Although the method in Boys et al. [7] is exact the estimate will be a random variable depending on the intrinsic noisy data. This is the same in the MSS method, which has additionally an approximation error. The relative error with the MSS method lies in 2/3 of the cases, below the smallest relative error of the methods suggested by Boys et al. [7], who gives one estimation result. This shows that the approximation error of the MSS method does not influence the estimation accuracy noticeable.
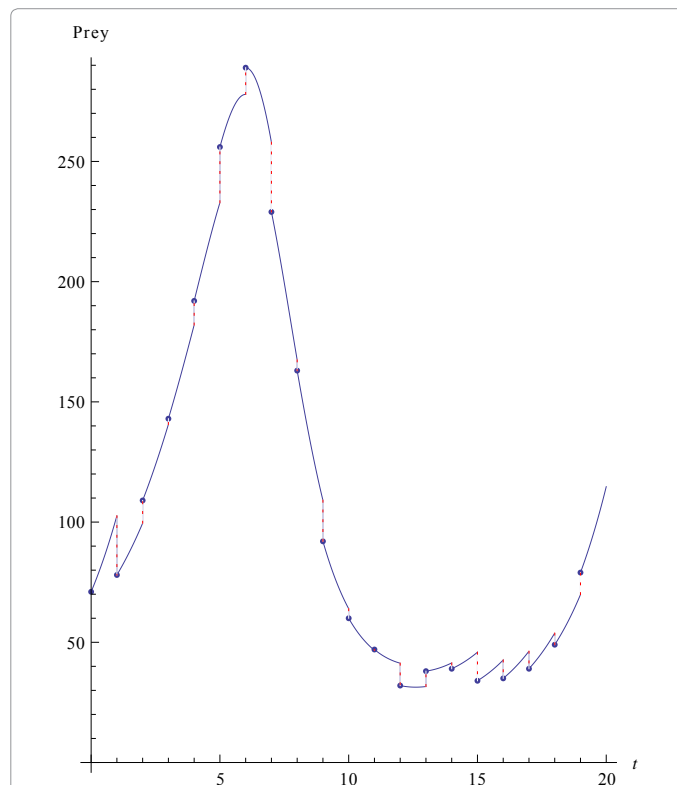
Calcium Oscillation Model

Third model is a Calcium oscillation model [27]

$$\frac{d\,g}{dt} = \theta_1 + \theta_2 g(t) - \frac{\theta_5\,g(t)\,plc(t)}{g(t)+\theta_6} - \frac{\theta_7\,g(t)\,plc(t)}{g(t)+\theta_8}$$

$$\frac{d\,plc}{dt} = \theta_3\,g(t) - \frac{\theta_9\,plc(t)}{plc(t)+\theta_{10}} \qquad (4)$$

$$\frac{d\,ca}{dt} = \theta_4\,g(t) - \frac{\theta_{11}\,ca(t)}{ca(t)+\theta_{12}}$$

This model shows a qualitatively different behavior between stochastic and ODE modeling for small particle numbers as shown in Kummer et al. [27]. In this model, the estimation of the transition probabilities is challenging due to the large state space.

The ODEs and parameters correspond to reactions and rate laws. These are interpreted stochastically using the software COPASI. To investigate the behavior of the MSS objective functional (1), 50 data sets are simulated from the stochastic model, corresponding to the ODEs given in (4) with the Gillespie algorithm using the software COPASI. The true parameter and initial conditions were $\theta^{(0)}$ =(212, 2.95, 1.52,

	Estimation Results		Rel err
	exact measurements		
Θ_1	0.501	±0.054	8.8%
Θ_2	0.0026	±3·10⁻⁴	11.1%
Θ_2	0.312	±0.048	13.4%
	noise: σ=10		
Θ_1	0.478	±0.050	9.3%
Θ_2	0.0026	±3·10⁻⁴	10.6%
Θ_2	0.318	±0.045	13.4%
	noise: σ=25		
Θ_1	0.430	±0.070	16.6%
Θ_2	0.0028	±4·10⁻⁴	16.8%
Θ_2	0.336	±0.048	16.5%

Table 3: Statistics of the estimation results for partially observed Lotka-Volterra model: prey Settings as in Table 2, but only prey can be observed.

Figure 3: Representation of Lotka-Volterra system's dynamics with MSS method. The blue points are the data points. Blue curves are the ODE dynamics for each time interval $[t_{i-1}, t_i]$ namely $h(t, \hat{\theta}, \tilde{v}_{i-1}, t_{i-1})$ with the estimated parameter. Red dotted lines show the residuals. 40 simulated observations from partially observed Lotka-Volterra model (prey) with Θ=(0.5,0.0025,0.3), T=40 and v_0=(71,79) for the estimation of Θ using (2) with K={0}. Dynamics shown until T=20.

190, 4.88, 1180, 1.24, 32240, 29090, 13.58, 153000, 160)

(Ca_0, g_0, plc_0) =(10, 10, 10).

Note that for this set of parameters, the system shows irregular oscillations with large amplitudes modeled stochastically, but shows regular oscillations with small amplitudes modeled with ODEs. 100 observations were taken with Δt=0.5, which cover about 4 oscillation cycles. For all data sets, the objective function (1) is optimized using a program implemented in the software Mathematica [33]. Table 4 gives averages and standard deviation for the 50 estimation results. For each estimation result, the relative error is calculated. The estimation performs quite successfully. The parameter θ_2 which determines the oscillatory behavior of the system has very small relative error.

Calculating for one example trajectory, the residuals for all three species yields the following numbers for 10%–Quantile, mean and 90%–quantile: Ca, {−197, −10, 127}; g, {−244, −1, 171}; plc, {−111, −1, 106}. The estimated variances are $\hat{\sigma}^2_{KL}$ =(178, 187, 83), with a KL divergence value of (0.7, 0.2, 0.05). The calculations for the autocorrelation time, Ca:1, g:0.5, plc:1, show that these are much smaller than the total observation time. The SNR are (10,11,19), which states that the system's dynamic is well represented, see also figure 4, which shows for calcium that the residuals (red dotted lines) are small compared to the ODE system's dynamics (blue line). For eg, see supplementary file 1 and supplementary file 2.

The hybrid approach of Mikeev and Wolf [15] would be of interest for a comparison. But as the species change the regime from very small molecule numbers to relatively high numbers, an automatic switching would be required. The MSS method handles this point technically less involved.

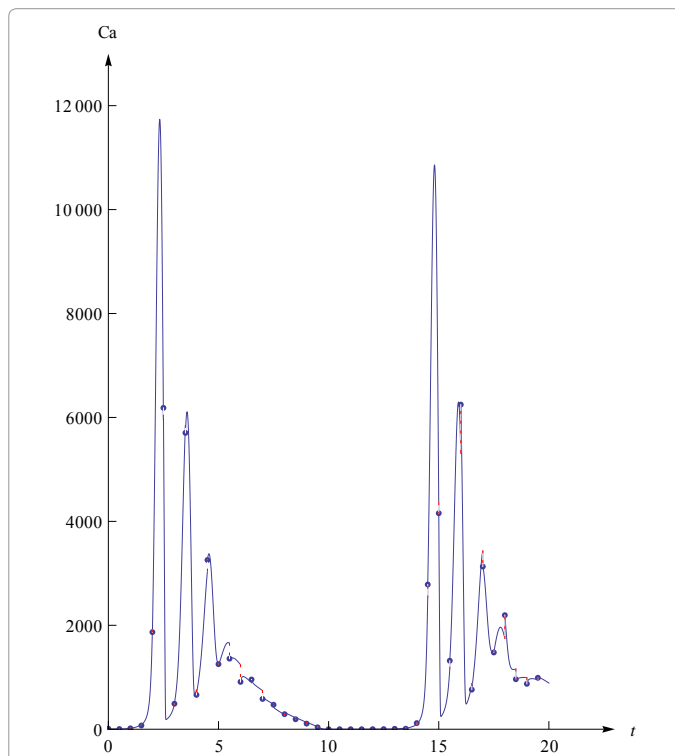### Partially observed Calcium oscillation model

Observing g: At first, assume that only g can be observed. As in the completely observed case, simulated data with true parameter $\theta^{(0)}$ and 100 observation with $\Delta t = 0.5$ is used. Initial condition is again (Ca_0, g_0, plc_0)=(10, 10, 10), but as only g can be observed only g_0=10 will be used as data input for the parameter estimation. The system is structurally unidentifiable now on a two dimensional manifold. The reason is that a multiple of the true amount of Ca_0 or Plc_0 can be compensated by adjusting the parameter values, see supplementary file 3. Therefore, for the optimization (Ca_0, plc_0) is fixed and the optimization problem still has 12 optimization variables, although the system is partially observed. Table 5 gives averages and standard deviation for the 50 estimation results, using (2) with $K=\{0\}$. For each estimation result, the relative error is calculated.

Calculating the averages of the residuals yields, the following numbers for 10%−quantile, mean and 90%−quantile: g, {−285,−2.7,270}. The estimated variance is $\hat\sigma^2_{KL}$ =232, with a KL divergence value of 0.07. The calculations for the autocorrelation time, g: 0.5, show that these are much smaller than the total observation time. Figure 5 shows that the system's dynamic is well represented, SNR is 7.1 which shows that the estimation is possible even with only one observed species.

Observing ca: This situation leads to difficulties. An estimation performed with the functional (2) with $K=\{0\}$, as it is done for g observed results for one example, in an estimate of $\hat\theta$ =(498, 2.76, 1.57, 14,1.89, 187, 5.49, 1183, 26244, 34848, 160653, 200) which is far from the true parameter. The calculation of the residuals yields as above with 10%−Quantile, mean and 90%−quantile: ca, {−187, 424, 1552}, which

Figure 4: Representation of Calcium oscillation system's dynamics with MSS method. The blue points are the data points. Blue curves are the ODE dynamics for each time interval [t_{i-1},t_i] namely $h(t, \hat\theta, v_{i-1}, t_{i-1})$ with the estimated parameter. Red dotted lines show the residuals. 100 observation from fully observed Calcium oscillation model with T=50 for the estimation of $\hat\Theta$, dynamics shown until T=20.

|	true param	Estimation Results		Rel err
Θ_1	212	293	±100	52%
Θ_2	2.95	2.97	±0.05	1.8%
Θ_3	1.52	1.88	±0.6	36%
Θ_4	190	410.6	±178	123%
Θ_5	4.88	4.86	±0.67	10.7%
Θ_6	1180	1610	±1332	86%
Θ_7	1.24	1.10	±0.39	28.4%
Θ_8	32240	48216	±11676	51%
Θ_9	29090	56510	±17299	94%
Θ_{10}	13.58	14.26	±1.67	10.3%
Θ_{11}	153000	160625	±18698	10.3%
Θ_{12}	160	167.0	±29.7	15.4%

Table 5: Statistics of the estimation results for partially observed Calcium oscillation model: g Settings as in Table 4, but only g observed.

	true param	Estimation Results		Rel err
Θ_1	212	210	±51	18.1%
Θ_2	2.95	2.95	±0.03	0.79%
Θ_3	1.52	1.52	±0.02	1.06%
Θ_4	190	197	±76	27.56%
Θ_5	4.88	4.88	±0.34	5.83%
Θ_6	1180	1198	±829	57%
Θ_7	1.24	1.24	±0.01	0.58%
Θ_8	32240	32735	±1518	3.4%
Θ_9	29090	29849	±2270	5.21%
Θ_{10}	13.58	13.56	±0.38	2.20%
Θ_{11}	153000	152754	±4439	2.30%
Θ_{12}	160	161	±7	3.09%

Table 4: Statistics of the estimation results for fully observed Calcium oscillation model. 50 data sets are simulated using the Gillespie algorithm with 100 observations with T=50 for the Calcium oscillation model, (Ca_0,g_0,plc_0)=(10,10,10). For each of the data sets an estimation is performed with the MSS method. The table shows a statistic of the 50 estimates: Parameter name (column 1), true parameter (column 2), averages of estimates (column 3), standard deviation of estimates (column 4) and relative errors (column 5).

seems to indicate a bias. SNR 1.5 and the system's dynamics are not at all represented, see also supplementary file 4. With the knowledge of the true parameter, the problem can be identified: For a short time interval, the system's dynamics are well represented, but then due to the development of the unobserved species, it is not well represented any longer, see figure 6. Hence, in this case, it is better to enlarge the optimization vector and use functional (2) with $K=\{0,5,10,\ldots,45\}$, which means that also unobserved states at other time points than

zero are included in the optimization vector. The results show that an estimation is possible with this functional, see table 6.
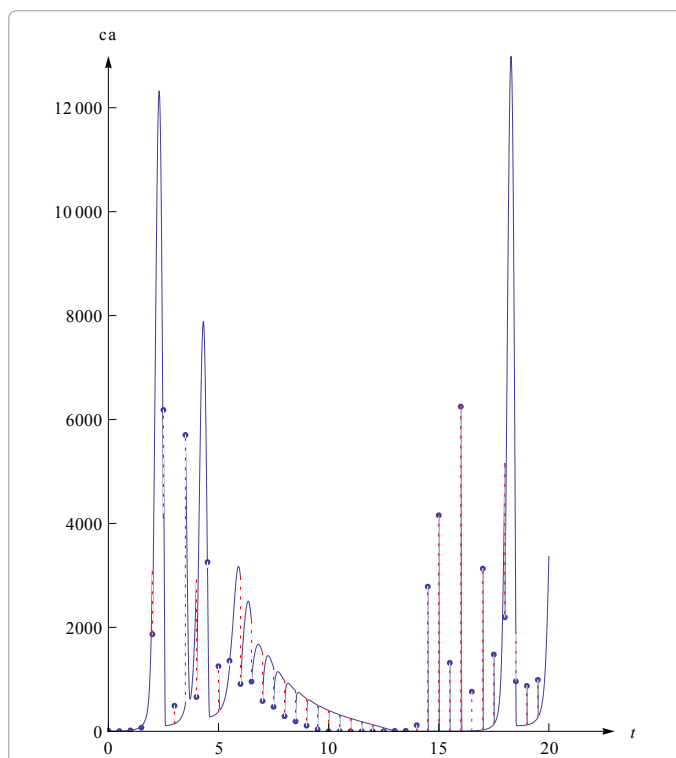
Now, the situation has improved. Averages of the test functions: 10%–quantile, mean and 90%–quantile of the residuals are $(-1997, 162, 1214)$. The estimated variance is $\hat{\sigma}_{KL}^2 = 1823$, with a KL divergence value of 0.67. The autocorrelation time is 1.16 and the SNR is 1.7. Figure 7 shows that the system's dynamic is well represented. This example is highly important, as it shows that the short time ODE integration method can still be used, if the system's behavior is very
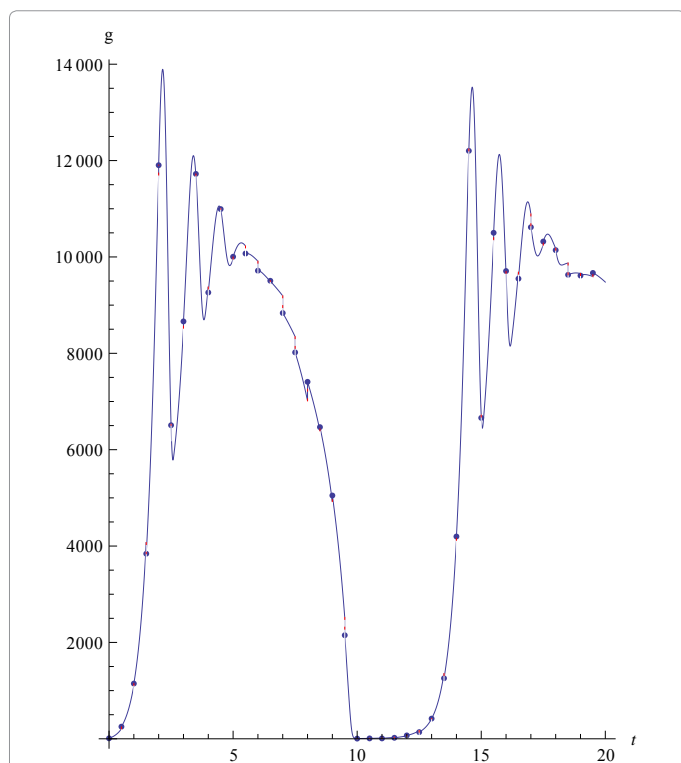
Figure 5: Representation of partially observed Calcium oscillation system's dynamics, *g*. The blue points are the data points. Blue curves are the ODE dynamics for each time interval $[t_{i-1}, t]$ namely $h(t, \hat{\theta}, \tilde{v}_{i-1}, t_{i-1})$ with the estimated parameter. Red dotted lines show the residuals. 100 observation from from partially observed Calcium oscillation model (*g*) with $T=50$ and $v_0 = (10, 10, 10)$ for the estimation of $\hat{\Theta}$ using (2) with $K=\{0\}$. Dynamics shown until $T=20$.
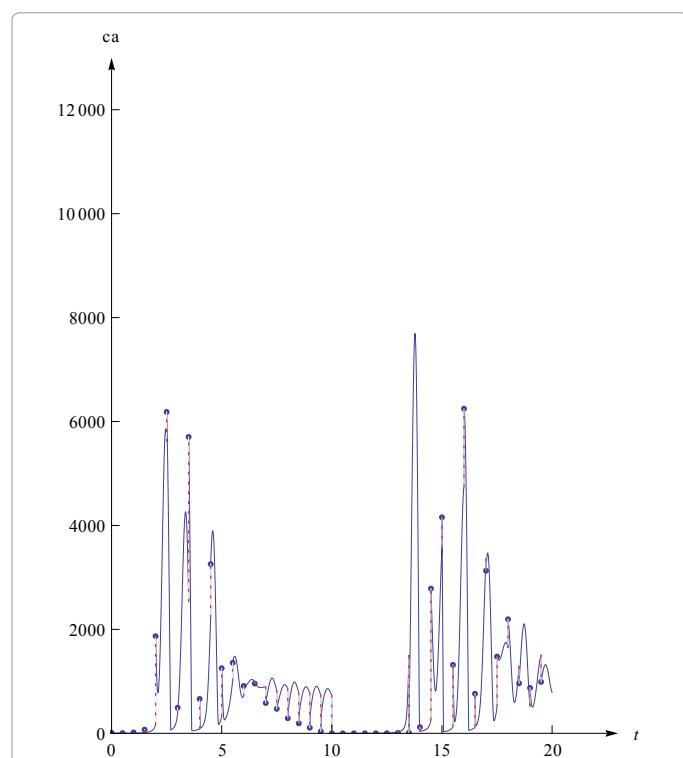

Figure 6: Representation of partially observed Calcium oscillation system's dynamics, *ca*. The blue points are the data points. Blue curves are the ODE dynamics for each time interval $[t_{i-1}, t]$ namely $h(t, \hat{\theta}, \tilde{v}_{i-1}, t_{i-1})$ with the true parameter. Red dotted lines show the residuals. 100 observation from partially observed Calcium oscillation model (*ca*) with $T=20$ using (2) with $K=\{0\}$.

different in ODE modeling or stochastic modeling. The method is even applicable to partially observed models.

Discussion

The article presents a method for parameter estimation in stochastic models, based on short time ODE integration. The method is able to estimate parameters, even in models which behave qualitatively different in stochastic modeling than in ODE modeling [27]. The data which is used as input for the method is a single stochastic trajectory, which might be well different from the mean behavior or other stochastic realizations. The advantage of the objective function is that it does not need stochastic simulations, nor a solution of a high dimensional CME, which increases its speed. Hence, it is well applicable in larger realistic size models which would be very time consuming in simulations based methods.

For the numerical optimization, this allows the user to choose between various numerical optimization methods, such as derivative based methods, global optimization techniques or Bayesian methods. As these are well established techniques, the article restricts its focus to the objective function. To test this objective function, it was optimized for three different examples. A more rigorous analysis including the calculation of confidence intervals and experimental design would be desirable and is planned for further research. This article focuses on the impact of the stochasticity of the models, and its influence on the estimation. The approximation with an ODE model on a short time interval is not a restriction for applicability, as the test functions show that the approximation works, even in models with qualitatively

	true param	Estimation Results		Rel err
Θ_1	212	145.4	±115.1	56.2%
Θ_2	2.95	3.06	±0.39	9.9%
Θ_3	1.52	1.49	±0.22	11.5%
Θ_4	190	212.3	±200.5	76.5%
Θ_5	4.88	7.85	±5.58	70.4%
Θ_6	1180	1944	±2405	144.4%
Θ_7	1.24	1.48	±0.20	21.7%
Θ_8	32240	33860	±8578	20.9%
Θ_9	29090	23670	±6278	23.7%
Θ_{10}	13.58	15.12	±5.55	24.3%
Θ_{11}	153000	161206	±22965	13.0%
Θ_{12}	160	101	±56.9	43.0%

Table 6: Statistics of the estimation results for partially observed Calcium oscillation model: *ca* Settings as in Table 4, but only *ca* observed.

Figure 7: Representation of partially observed Calcium oscillation system's dynamics, *ca*. The blue points are the data points. Blue curves are the ODE dynamics for each time interval $[t_{i-1}, t_i]$ namely $h(t, \hat{\theta}, \tilde{v}_{i-1}, t_{i-1})$ with the estimated parameter. Red dotted lines show the residuals. 100 observations from partial observed Calcium oscillation model (*ca*) with $T=50$ for the estimation of $\hat{\Theta}$ using (2) with $K=\{0,5,\ldots,45\}$. Dynamics shown until $T=20$.

different behavior in stochastic modeling than in ODE modeling. For that it is important that the approximation is only done on a short time interval. In models with very few reactions per time interval, such as the Immigration-Death model, the signal to noise ratio is bad, so that many observations are necessary. But in many realistic models, it is not possible to measure fast enough to capture every single reaction, so this fact does not reduce the applicability of the method much. The MSS method proposed in this article performs very well, even in oscillatory system such as Lotka-Volterra or the Calcium oscillation model, in which a larger state space makes other approaches more time consuming.

Instead of using the solution of the initial value problem of the ODE system in (1), one could also use the mean of the stochastic simulations-hence the first moment-which could be calculated with a moment-closure [13], with only the first moment. Although calculating the moment-closure should be fast compared to other calculations, this would be slightly slower than the suggested method. The models in the results section suggest that it is in many cases not necessary. If it would be more stable, especially in cases of partially observed models, still has to be investigated. The results for the Immigration-Death model suggest that the method yields an unbiased result compared to an exact analytical estimation, if there are enough measurements. The standard deviation seems to be slightly higher. Compared to a stochastic approach by [8], the suggested method performed well: the results for sample data sets were more accurate in 80% of the 50 sample data sets, which still might be due to the stochastic effects as [8], provided a single data set. The case of table 1A, in which the method is biased

is a situation when only very few reactions happen per time interval, between two points of measurements. This is in reality generally not the case. If the systems which is under investigation has that behavior, it is suggested to use CME based methods, which should work fast in that case due to the very small state space.

The benefit of the MSS method using the measurement points as initialization for the next interval is that it resolves the identifiability problem, which is present in this scenario of the Immigration-Death model, using traditional ODE methods.

The results of the Lotka-Volterra model show that the method is well able to provide estimates for fully and partially observed models, comparable to the accuracy of methods using simulations [7].

The Calcium oscillation model was a very important test case as the method showed good performance in estimating the parameters, although the system's behavior is completely different in stochastic and deterministic modeling. For the partially observed case, the amount of information depends on the question which species is observed. Observing g the estimation is possible; observing only ca, the objective function has to be modified to (2) with K={0,5,…,45}.

The examples show that although the Maximum Likelihood property of the method is theoretically lost, the estimation is still quite precise. The proposed test functions work fine as they identify those situations, which lead to a bias and "accept" the others. This argumentation also allows for the extension to the case of measurement noise, in which even data points with negative resulting negative molecule counts can be used. An example with measurement noise is given for the Lotka-Volterra example.

To conclude the article, presents a method for parameter estimation for stochastic models. The method is based on short time ODE integration. It is shown that the approach exhibits the desirable properties mentioned in the introduction, that makes it applicable to realistic problems in systems biology: it works in models which behave qualitatively different in stochastic modeling than in ODE modeling. It is able to cope with partially observed models with measurement noise. As the method is deterministic and works without stochastic simulations, it is computationally efficient.

Points for future research are confidence intervals for the parameters, as well as optimum experimental design. The first is crucial for the quality assessment of the estimates, and the second helps in saving experimental cost, whilst increasing the estimation accuracy.

Acknowledgements

References

1. Raj A, van Oudenaarden A (2009) Single-molecule approaches to stochastic gene expression. Annu Rev Biophys 38: 255-270.

2. Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22: 403-434.

3. Pahle J (2009) Biochemical simulations: stochastic, approximate stochastic and hybrid approaches. Brief Bioinform 10: 53-64.

4. Jagaman K, Danuser G (2006) Linking data to models: data regression. Nat Rev Mol Cell Biol 7: 813-819.

5. Tian T, Xu S, Gao J, Burrage K (2007) Simulated maximum likelihood method for estimating kinetic rates in gene expression. Bioinformatics 23: 84-91.

6. Poovathingal SK, Gunawan R (2010) Global parameter estimation methods for stochastic biochemical systems. BMC Bioinformatics 11: 414.

7. Boys RJ, Wilkinson DJ, Kirkwood TBL (2008) Bayesian inference for a discretely observed stochastic kinetic model. Stat Comput 18: 125-135.

8. Wang Y, Christley S, Mjolsness E, Xie X (2010) Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent. BMC Syst Biol 4: 99.

9. Henderson D, Boys R, Proctor C, Wilkinson D (2010) Linking systems biology models to data: a stochastic kinetic model of p53 oscillations. In: The Oxford Handbook of Applied Bayesian Analysis, O' Hagan A, West M (Eds). OUP Oxford, Oxford.

10. Reinker S, Altman RM, Timmer J (2006) Parameter estimation in stochastic biochemical reactions. Syst Biol (Stevenage) 153: 168-178.

11. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface 6: 187-202.

12. Andreychenko A, Mikeev L, Spieler D, Wolf V (2011) Parameter identification for markov models of biochemical reactions.

13. Gillespie CS (2009) Moment-closure approximations for mass-action models. IET Syst Biol 3: 52-58.

14. Gillespie CS, Golightly A (2010) Bayesian inference for generalized stochastic population growth models with application to aphids. J R Stat Soc Ser C Appl Stat 59: 341-357.

15. Mikeev L, Wolf V (2012) Parameter estimation for stochastic hybrid models of biochemical reaction networks. HSCC 12, Beijing.

16. Deuflhard P, Huisinga W, Jahnke T, Wulkow M (2008) Adaptive discrete galerkin methods applied to the chemical master equation. SIAM J Sci Comput 30: 2990-3011.

17. Engblom S (2006) A discrete spectral method for the chemical master equation. Technical Report, Uppsala University 2006-036.

18. Munsky B, Khammash M (2010) Identification from stochastic cell-to-cell variation: a genetic switch case study. IET Syst Biol 4: 356-366.

19. Zechner C, Ruess J, Krenn P, Pelet S, Peter M, et al. (2012) Moment-based inference predicts bimodality in transient gene expression. Proc Natl Acad Sci U S A 109: 8340-8345.

20. Kügler P (2012) Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. PLoS One 7: e43001.

21. Andersson H, Britton T (2000) Stochastic epidemic models and their statistical analysis, Vol 151 of Lecture notes in statistics. Springer, Heidelberg.

22. Rempala GA (2012) Least squares estimation in stochastic biochemical networks. Bull Math Biol 74: 1938-1955.

23. Bock H (1981) Numerical treatment of inverse problems in chemical reaction kinetics. In: Ebert K, Deuhard P, Jäger W (eds), Modelling of Chemical Reaction Systems, Volume 18 of Springer Series in Chemical Physics. Springer, Heidelberg.

24. Schlöder J, Bock H (1983) Identification of Rate Constants in Bistable Chemical Reactions. In: Deuhard P, Hairer E (eds), Numerical Treatment of Inverse Problems in Differential and Integral. Birkhäuser, Berlin.

25. Baake E, Baake M, Bock HG, Briggs KM (1992) Fitting ordinary differential equations to chaotic data. Phys Rev A 45: 5524-5529.

26. Kallrath J, Schlöder J, Bock HG (1993) Least squares estimation in chaotic differential equations. Celestial Mechanics and Dynamical Astronomy 56: 353-371.

27. Kummer U, Krajnc B, Pahle J, Green AK, Dixon CJ, et al. (2005) Transition from stochastic to deterministic behavior in calcium oscillations. Biophys J 89: 1603-1611.

28. Komorowski M, Finkenstädt B, Harper CV, Rand DA (2009) Bayesian inference of biochemical kinetic parameters using the linear noise approximation. BMC Bioinformatics 10: 343.

29. Bock HG, Kostina E, Schlöder JP (2007) Numerical methods for parameter estimation in nonlinear differential algebraic equations. GAMM Mitteilungen 30: 376-408.

30. Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparision of global optimization methods. Genome Res 13: 2467-2474.

31. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, et al. (2006) COPASI--a COmplex PAthway SImulator. Bioinformatics 22: 3067-3074.

32. Cox DR, Miller HD (1965) The Theory of Stochastic Processes. Wiley, London.

33. Wolfram Research (2010) Mathematica, version 8.0. Wolfram Research, Inc Champaign, IL.