BMC
Bioinformatics

METHODOLOGY ARTICLE

Open Access

Identification of models of heterogeneous cell populations from population snapshot data

Jan Hasenauer[1*], Steffen Waldherr[1], Malgorzata Doszczak[2], Nicole Radde[1], Peter Scheurich[2] and Frank Allgöwer[1]

Abstract

Background: Most of the modeling performed in the area of systems biology aims at achieving a quantitative description of the intracellular pathways within a "typical cell". However, in many biologically important situations even clonal cell populations can show a heterogeneous response. These situations require study of cell-to-cell variability and the development of models for heterogeneous cell populations.

Results: In this paper we consider cell populations in which the dynamics of every single cell is captured by a parameter dependent differential equation. Differences among cells are modeled by differences in parameters which are subject to a probability density. A novel Bayesian approach is presented to infer this probability density from population snapshot data, such as flow cytometric analysis, which do not provide single cell time series data. The presented approach can deal with sparse and noisy measurement data. Furthermore, it is appealing from an application point of view as in contrast to other methods the uncertainty of the resulting parameter distribution can directly be assessed.

Conclusions: The proposed method is evaluated using artificial experimental data from a model of the tumor necrosis factor signaling network. We demonstrate that the methods are computationally efficient and yield good estimation result even for sparse data sets.

Background

The main goals of research in systems biology are the development of quantitative models of intracellular pathways and the development of tools to support the modeling process. Thereby, most of the available methods and models consider only a single "typical cell" whereas most experimental data used to calibrate the models are obtained using cell population experiments, e.g. western blotting. This yields problems in particular if the studied population shows a large cell-to-cell variability. In such situations inferring a single cell model from cell population data can lead to biologically meaningless results. In order to understand the dynamical behavior of heterogeneous cell populations, it is crucial to develop cell population models, describing the whole population and not only a single individual [1-4].

This has already been realized by several authors, and it has been shown that stochasticity in biochemical

reactions and unequal partitioning of cell material at cell division can lead to complex population dynamics [1-5], such as bimodal distributions. Besides these sources for heterogeneity also genetic and epigenetic differences have to be considered [6].

For the purpose of this paper, heterogeneity in populations is modeled by differences in parameter values and initial conditions of the model describing the single cell dynamics [4,7,8]. The network structure is assumed to be identical in all cells. The distribution of the parameter values within the cell population is described by a multivariate probability density function, which is part of the population model. This modeling framework is well suited for modeling genetic and epigenetic differences among cells [2,4,7].

In the following, the problem of estimating the probability density of the parameters is studied. Therefore, we employ population snapshot data (PSD), which provide single cell measurements at every time instance but which do not provide single cell time series data. A typical experimental setup which provides PSD is flow

* Correspondence: Jan.Hasenauer@ist.uni-stuttgart.de
[1]Institute for Systems Theory and Automatic Control, University of Stuttgart, Germany
Full list of author information is available at the end of the article

cytometric analysis. In general, PSD are a common data type in the experimental analysis of biological systems.

So far, there are not many methods available for the estimation of parameter distributions. In pharmacokinetic studies mixed effect models [9] are used frequently. Unfortunately, as in the problem we consider the number of individuals is very large ($> 10^4$) and the amount of information per individual very limited (often only one data point), these methods are computationally too demanding. Furthermore, as in this study we are particularly interested in intracellular signal transduction, also methods which purely focus on the population balance [10-12] cannot be employed. In [8,13,14] methods are proposed which can in principle deal with the problem at hand. There, the considered estimation problem has been formulated as a convex optimization problem. Unfortunately, these methods either require an extensive amount of measurement data [8,13], and/or do not allow considering prior knowledge [8,13,14]. Additionally, no methods to evaluate the reliability of the estimates are provided.

In this paper a novel Bayesian approach [15,16] for inferring the parameter density will be introduced. The approach is mainly based on the maximum likelihood methods presented in [13,14], but can deal with sparse and noisy single cell data in addition to realistic measurement noise models. Furthermore, one may directly access the remaining uncertainty of the estimation result and the prediction uncertainties via the calculation of Bayesian confidence intervals [17,18]. It is shown that the posterior distribution can be determined efficiently employing a parameterization of the parameter density in combination with commonly used Markov chain Monte Carlo (MCMC) sampling techniques [19].

To illustrate the properties of the proposed methods, a mathematical model of the tumor necrosis factor (TNF) pathway [20] is analyzed using artificial experimental data.

Methods
Problem statement
Cell population model
For the purpose of this work we consider intracellular biochemical reaction networks which are modeled by systems of ordinary differential equations. This modeling framework allows to describe metabolic networks as well as signal transduction pathways, as long as spatial effects and stochasticity of the biochemical reactions can be neglected. Mathematically, the dynamic behavior of each single cell is determined by an ordinary differential equation in state space form

$$\Sigma(\theta^{(i)}) : \begin{cases} \dot{x}^{(i)} = F(x^{(i)}, \theta^{(i)}), & x^{(i)}(0) = x_0(\theta^{(i)}), \\ y^{(i)} = H(x^{(i)}, \theta^{(i)}) \end{cases} \quad (1)$$

with state variables $x^{(i)}(t) \in \mathbb{R}_+^n$, output variables $y^{(i)}(t) \in \mathbb{R}_+^m$, and parameters $\theta^{(i)} \in \mathbb{R}_+^q$. The vector field $F : \mathbb{R}_+^n \times \mathbb{R}_+^q \to \mathbb{R}^n$ is Lipschitz continuous and the functions $H : \mathbb{R}_+^n \times \mathbb{R}_+^q \to \mathbb{R}_+^m$ and $x_0 : \mathbb{R}_+^q \to \mathbb{R}_+^n$ are continuous. If for example the concentration $x_k^{(i)}$ is measured via flow cytometry, we would have $y^{(i)} = H(x^{(i)}, \theta^{(i)}) = c x_k^{(i)}$, where c is a proportionality factor. The index i specifies the individual cells within the population. The parameters $\theta^{(i)}$ can be kinetic constants, e.g. reaction rates or binding affinities.

Employing the definition of the single cell dynamics (1), a cell population model is given by the collection of N cells,

$$\Sigma_{\text{pop}} = \left\{ \Sigma(\theta^{(i)}) | i = \{1, \ldots, N\}, \theta^{(i)} \sim \Theta \right\}. \quad (2)$$

The heterogeneity within this population is modeled by differences in parameter values among individual cells. The parameters are distributed according to the probability density function $\Theta : \mathbb{R}_+^q \to \mathbb{R}_+$, with $\int_{\mathbb{R}_+^q} \Theta(\theta) d\theta = 1$. This density function is part of the population model specification. The parameter vector of cell i is subject to the probability distribution

$$\Pr(\theta_1^{(i)} \leq \theta_1, \cdots, \theta_q^{(i)} \leq \theta_q) = \int_0^{\theta_1} \cdots \int_0^{\theta_q} \Theta(\tilde{\theta}) d\tilde{\theta}_1 \cdots d\tilde{\theta}_q. \quad (3)$$

Note that interactions among individual cells influencing the analyzed pathway are not allowed. This is a restriction but indeed fulfilled in many *in vitro* lab experiments.
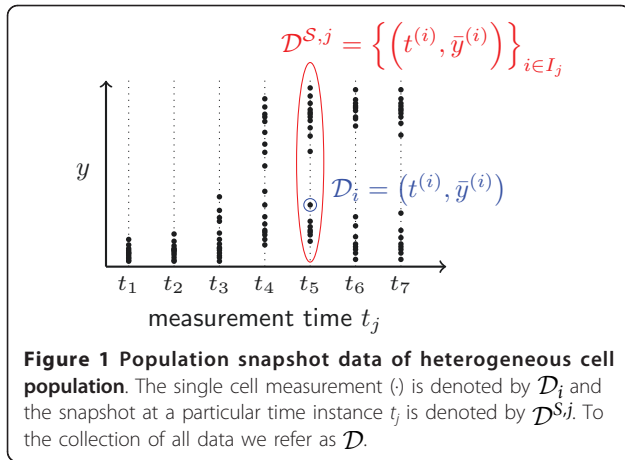
Measurement data and noise
In this paper we consider experimental setups where measurement data are obtained in the form of population snapshots, e.g. via flow cytometry. Population snapshots are taken at different times t_j, and the jth snapshot contains measurements for the output y of M_j cells. Due to experiment setup, it can be assumed that any cell is present at most in one snapshot.

The cells in the individual snapshots are referenced through index sets: snapshot j contains all cells from the index set $I_j = \left\{ \sum_{k=0}^{j-1} M_k + 1, \ldots, \sum_{k=0}^{j} M_k \right\}$, with $M_0 = 0$. Thereby, $j = 1, \ldots, n_\mathcal{S}$ in which $n_\mathcal{S}$ is the number of snapshots.

Let the data point for the cell with index i be denoted as

$$\mathcal{D}_i = \left(t^{(i)}, \bar{y}^{(i)}(t^{(i)}) \right), \quad (4)$$

where $t^{(i)}$ is the time at which the measurement was taken, and $\bar{y}^{(i)}$ is the measured, noise-distorted output as defined below. If cell i has been measured as part of the jth snapshot, then $t^{(i)} = t_j$. The snapshot j is the set of all data points $\mathcal{D}_i$ with $i \in I_j$, as depicted in Figure 1:

Figure 1 Population snapshot data of heterogeneous cell population. The single cell measurement (·) is denoted by $\mathcal{D}_i$ and the snapshot at a particular time instance t_j is denoted by $\mathcal{D}^{S,j}$. To the collection of all data we refer as $\mathcal{D}$.

$$\mathcal{D}^{S,j} = \bigcup_{i \in I_j} \mathcal{D}_i = \left\{ \left(t^{(i)}, \bar{\gamma}^{(i)}(t^{(i)}) \right) \right\}_{i \in I_j}. \quad (5)$$

In the parameter estimation, only the union of all snapshots is considered, and the parameter density function Θ is fitted to all snapshots simultaneously. To this end, we introduce the collection of all data, denoted as

$$\mathcal{D} = \bigcup_{j=1}^{n_S} \mathcal{D}^{S,j} = \left\{ \left(t^{(i)}, \bar{\gamma}^{(i)}(t^{(i)}) \right) \right\}_{i=1}^{M}, \quad (6)$$

in which $M = M_1 + M_2 + \ldots + M_{n_S}$ is the total number of measured cells.

We emphasize that experimental setups are considered in which cells are not tracked over time. These setups are very common in studies on the population scale. Classical examples for measurement techniques yielding such data are flow cytometric analysis and cytometric fluorescence microscopy. These measurement techniques allow to determine protein concentrations within single cells. As the population is well mixed when the measurement is performed and no cell is measured more than once, the individual single cell measurements $\mathcal{D}_i$ are independent. This independence of $\mathcal{D}_{i_1}$ and $\mathcal{D}_{i_2}$ (respectively $\bar{\gamma}^{(i_1)}(t^{(i_1)})$ and $\bar{\gamma}^{(i_2)}(t^{(i_2)})$), $i_1 \neq i_2$, holds if both cells are measured during one snapshot ($t^{(i_1)} = t^{(i_2)}$) as well as if the cells are measured within different snapshots ($t^{(i_1)} \neq t^{(i_2)}$).

Like most other experiments also the considered single cell experiments are subject to noise. We consider the noise model

$$\bar{\gamma}^{(i)}(t^{(i)}) = \text{diag}\left(\eta^{\times} \right) \gamma^{(i)}(t^{(i)}) + \eta^{+}, \quad (7)$$

in which $\bar{\gamma}^{(i)}$ is the measured output and $y^{(i)}$ is the actual output from (1). The multiplicative noise is denoted by $\eta^{\times} \in \mathbb{R}^m$ and the additive noise s denoted by $\eta^{+} \in \mathbb{R}^m$. Both, $\eta^{\times}$ and η^{+} are in the following

assumed to be vectors of log-normally distributed random variables with probability density functions

$$p(\eta_k^j | \sigma_k^j, \mu_k^j) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma_k^j \eta_k^j} \exp\left(-\dfrac{(\log \eta_k^j - \mu_k^j)^2}{2(\sigma_k^j)^2} \right) & \eta_k^j > 0 \\ 0 & \eta_k^j \leq 0, \end{cases} \quad (8)$$

for all $j = \{\times, +\}$, $k = 1, \ldots, m$. This noise model allows the study of relative and absolute measurement noise and describes the commonly seen noise distributions in biological experiments [21].

From (8) the conditional probability of measuring $\bar{\gamma}^{(i)}$ given $y^{(i)}$ can be determined. As the different output errors are assumed to be independent the conditional probability density is

$$p(\bar{\gamma}^{(i)} | \gamma^{(i)}) = \prod_{k=1}^{m} p(\bar{\gamma}_k^{(i)} | \gamma_k^{(i)}), \quad (9)$$

with $p(\bar{\gamma}_k^{(i)} | \gamma_k^{(i)})$ being the value of the line integral

$$p(\bar{\gamma}_k^{(i)} | \gamma_k^{(i)}) = \int_0^{\bar{\gamma}_k^{(i)} / \gamma_k^{(i)}} p(s | \sigma_k^{\times}, \mu_k^{\times}) p(\bar{\gamma}_k^{(i)} - \gamma_k^{(i)} s | \sigma_k^{+}, \mu_k^{+}) ds, \quad (10)$$

which is illustrated in Figure 2. For this line integral no explicit solution can be given. In this paper its value is determined numerically using the adaptive Simpson quadrature method [22] implemented in `MATLAB`.

Problem formulation

As mentioned previously, when studying heterogeneous populations the density of the parameters Θ is in general unknown but necessary to gain an in-depth understanding of the population dynamics. Therefore, we are concerned with the problems:

Problem 1 *Given the measurement data $\mathcal{D}$, the cell population model Σ_{pop}, and the noise model (8), infer the parameter density Θ^*.*

Figure 2 Line integral (10). Set (*red line*) of measurement noise realization $\eta^{\times}$ and η^{+} which yield for a given output $\gamma_k^{(i)}$ the measured, noise corrupted output $\bar{\gamma}_k^{(i)}$.
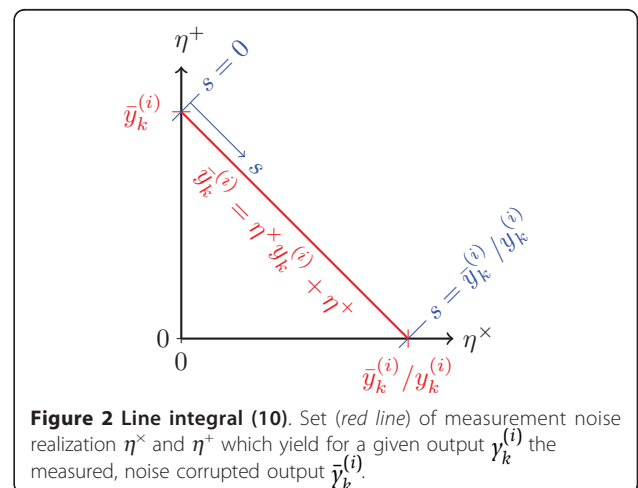
Problem 2 *Given the measurement data $\mathcal{D}$, the cell population model Σ_{pop}, and the noise model (8), determine the uncertainty of the estimated parameter density Θ^*.*

Unfortunately, the number of cells considered in a standard lab experiment is on the order of 10^4 to 10^7. Thus, simulating the population model (2) is computationally expensive. Furthermore, it is hard from a theoretical point of view to deal with ensemble models such as (2). Density-based descriptions of the population dynamics are far more appealing for solving Problem 1 and 2. Therefore, in the next section a density description of the population is introduced.

Density-based modeling of cell populations

To simplify the inference problems on Θ the population description is changed from an ensemble model (2) to a density model [13]. The variables of this new model are no longer states or outputs of the single cells but the density function Υ of the output, with $\Upsilon : \mathbb{R}_+^m \times \mathbb{R} \times \ell^1 \to \mathbb{R}_+ : (\gamma, t, \Theta) \mapsto \Upsilon(\gamma|t, \Theta)$. and $\int_{\mathbb{R}^m} \Upsilon(\gamma|t, \Theta)d\gamma = 1 \; \forall t$ This density function yields the probability of drawing a cell sample from the cell population with output $\gamma^{(i)}(t) \in \mathcal{Y} \subset \mathbb{R}_+^m$,

$$\Pr(\gamma^{(i)}(t) \in \mathcal{Y}) = \int_{\mathcal{Y}} \Upsilon(\gamma|t, \Theta)d\gamma, \tag{11}$$

in which $\mathcal{Y} \subset \mathbb{R}_+^m$ is an arbitrary subset of the output space. Hence, $y^{(i)}(t)$ can be interpreted as a random variable which is distributed according to $\Upsilon(\gamma|t,\Theta)$.

To compute the cell population response $\Upsilon(\gamma|t,\Theta)$ for a given Θ, S independent single cell trajectories $y^{(i)}(t)$ of the cell population (2) are calculated. The parameters for these cells are drawn from Θ and the initial conditions are computed according to $x_0(\theta^{(i)})$. This yields $\{y^{(i)}(t)\}_{i=1}^S$ with distribution of $y^{(i)}(t)$ depending on Θ. Given $\{y^{(i)}(t)\}_{i=1}^S$ an approximation of Υ is

$$\Upsilon(\gamma|t, \Theta) = \frac{1}{S} \sum_{i=1}^S \mathcal{K}\left(\gamma, \gamma^{(i)}(t), H\right), \tag{12}$$

in which $\mathcal{K}(\gamma, \gamma^{(i)}(t), H)$ is the density of the applied kernel density estimator, with $\int_{\mathbb{R}^m} \mathcal{K}d\gamma = 1$. This is illustrated in Figure 3. In this work multivariate log-normal kernels

$$\mathcal{K}\left(\gamma, \gamma^{(i)}(t), H\right) = \begin{cases} \dfrac{\exp\left(-\dfrac{1}{2}v^T H^{-1} v\right)}{(2\pi)^{\frac{m}{2}} |H|^{\frac{1}{2}} \prod_{k=1}^m \gamma_k} & \gamma > 0 \\ 0 & \gamma \le 0 \end{cases} \tag{13}$$

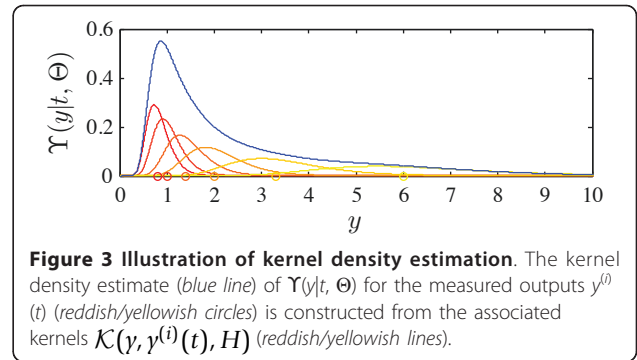$$\text{with}: v = \log \gamma - \log \gamma^{(i)}(t),$$

Figure 3 Illustration of kernel density estimation. The kernel density estimate (*blue line*) of $\Upsilon(y|t, \Theta)$ for the measured outputs $y^{(i)}$ (t) (*reddish/yellowish circles*) is constructed from the associated kernels $\mathcal{K}(\gamma, \gamma^{(i)}(t), H)$ (*reddish/yellowish lines*).

are used to conserve the property that all variables are positive. The positive definite matrix H is used to select the width of the kernel $\mathcal{K}$, and is chosen using the available rule of thumb described in [23]. The selection of H is crucial for achieving a good approximation of the probability density, and depends strongly on S.

Approaches similar to the one we use to approximate $\Upsilon(y|t,\Theta)$ are employed in [13,14], in [8] with a naive density estimator and in [24] in the context of cell migration.

Estimation of the parameter density

In the previous section an approach to determine the output density Υ within the cell population for a given parameter density Θ is presented. Based on this an approach for estimating Θ from the available data $\mathcal{D}$ is developed next.

Bayes' theorem for heterogeneous cell populations

For learning the parameter density from the data Bayes' theorem

$$p(\Theta|\mathcal{D}) = \frac{p(\mathcal{D}|\Theta)p(\Theta)}{p(\mathcal{D})} \tag{14}$$

is used, in which $p(\Theta)$ is the prior probability of Θ, $p(\mathcal{D}|\Theta)$ is the conditional probability of observing $\mathcal{D}$ given Θ, $p(\Theta|\mathcal{D})$ is the posterior probability of Θ given $\mathcal{D}$, and $p(\mathcal{D}) = \int p(\mathcal{D}|\Theta)p(\Theta)d\Theta$ is the marginal probability of $\mathcal{D}$. As the different single cell measurements are independent (14) can be written as

$$p(\Theta|\mathcal{D}) = \frac{\prod_{i=1}^M p(\mathcal{D}_i|\Theta)p(\Theta)}{\int \prod_{i=1}^M p(\mathcal{D}_i|\Theta)p(\Theta)d\Theta}, \tag{15}$$

in which $p(\mathcal{D}_i|\Theta)$ is the conditional probability of observing $\mathcal{D}_i$ given Θ. Note that due to the independence of $\mathcal{D}_{i_1}$ and $\mathcal{D}_{i_2}$, for $i_1 \ne i_2$, it is not necessary to distinguish between the cases that (1) the two cells are measured at the same instance ($t^{(i_1)} = t^{(i_2)}$) and that (2) the two cells are measured at different time instances

$\mathcal{K}(\gamma, \gamma^{(i)}(t), H)$. Hence, merely the conditional probability of each individual single cell measurement has to be determined. For the considered process the $p(\mathcal{D}_i|\Theta)$ can be determined using the output density Υ,

$$
\begin{aligned}
p(\mathcal{D}_i|\Theta) &= p(\bar{\gamma}^{(i)}|\Upsilon(\gamma|t^{(i)}, \Theta)) \\
&= \int_{\mathbb{R}_+^m} p(\bar{\gamma}^{(i)}|\gamma)\Upsilon(\gamma|t^{(i)}, \Theta)d\gamma.
\end{aligned} \tag{16}
$$

As this equation cannot be solved explicitly the integral has to be approximated numerically. This could be realized using importance sampling [19], but as drawing a independent sample from Υ requires knowledge of Υ in the first place, $p(\mathcal{D}_i|\Theta)$ is for computational purposes expressed as an integral over θ,

$$
p(\mathcal{D}_i|\Theta) = \int_{\mathbb{R}_+^q} p(\bar{\gamma}^{(i)}|\theta)\Theta(\theta)d\theta, \tag{17}
$$

in which $p(\bar{\gamma}^{(i)}|\theta) = p(\bar{\gamma}^{(i)}|\gamma(t^{(i)}, \theta))$, with $\gamma(t^{(i)}, \theta)$ being the output at time $t^{(i)}$ for a cell having parameters θ. This reformulation of (16) is possible as Υ directly depends on Θ. This step simplifies the evaluation of $p(\mathcal{D}_i|\Theta)$ tremendously.

Based on (15) and (17), the calculation of the posterior probability for a given probability density of the parameters Θ is possible. Unfortunately, the inference problem nevertheless cannot be solved directly, as Θ is element of a function space, and hence further steps are necessary.

Parameterization of parameter density

In order to avoid the infinite dimensional inference problem the parameter density is parameterized. Θ is modeled by a finite weighted sum of multivariate ansatz functions Λ_j,

$$
\Theta_\varphi(\theta) = \sum_{j=1}^{n_\varphi} \varphi_j \Lambda_j(\theta). \tag{18}
$$

The ansatz functions $\Lambda_j : \mathbb{R}_+^q \to \mathbb{R}_+$ are themselves probability densities with $\int_{\mathbb{R}_+^q} \Lambda_j(\theta)d\theta = 1$. The weighting vector is denoted by $\varphi \in [0, 1]^{n_\varphi}$, where n_φ is the number of ansatz functions and $\sum_{j=1}^{n_\varphi} \varphi_j = 1$ to guarantee that Θ_φ is a probability density. The weightings φ_j can be interpreted as parameters determining the probability density Θ_φ and are for the remainder also called density parameters.

Note that the method presented in the following is independent of the choice of ansatz functions. Nevertheless, a clever choice of the ansatz functions may improve the approximation of the true parameter density dramatically. In this work, the ansatz functions are chosen to be multivariate Gaussians.

Given a parameterization of Θ_φ, the output density can be simplified to

$$
\Upsilon(\gamma|t, \Theta_\varphi) = \Upsilon\left(\gamma \,\middle|\, t, \sum_{j=1}^{n_\varphi} \varphi_j \Lambda_j\right), \tag{19}
$$

$$
= \sum_{j=1}^{n_\varphi} \varphi_j \Upsilon(\gamma|t, \Lambda_j), \tag{20}
$$

in which $\Upsilon(\gamma|t, \Lambda_j)$ is the output density obtained for single cell parameters distributed according to Λ_j. This representation of the response is possible as the output density fulfills the superposition principle with respect to the parameter distribution Θ_φ. This reformulation has the advantage that computing the output density for arbitrary density parameters φ only requires the non-recurring computation of the responses $\Upsilon(\gamma|t, \Lambda_j)$ and summation of those.

Reformulation of posterior probability

Having parameterized Θ_φ and $\Upsilon(\gamma|t, \Theta_j)$ the conditional probability $p(\mathcal{D}_i|\Theta_\varphi)$ may be parameterized and expressed as the weighted sum,

$$
p(\mathcal{D}_i|\Theta_\varphi) = \int_{\mathbb{R}_+^q} p(\bar{\gamma}^{(i)}|\theta) \left(\sum_{j=1}^{n_\varphi} \varphi_j \Lambda_j(\theta)\right) d\theta, \tag{21}
$$

$$
= \sum_{j=1}^{n_\varphi} \varphi_j \int_{\mathbb{R}_+^q} p(\bar{\gamma}^{(i)}|\theta) \Lambda_j(\theta) d\theta, \tag{22}
$$

$$
= \sum_{j=1}^{n_\varphi} \varphi_j p(\mathcal{D}_i|\Lambda_j), \tag{23}
$$

in which $p(\mathcal{D}_i|\Lambda_j)$ is the conditional probability of observing $\mathcal{D}_i$ given the parameter density Λ_j. As the ansatz functions are predefined the conditional probability $p(\mathcal{D}_i|\Lambda_j)$ can be evaluated,

$$
c_j^{(i)} := p(\mathcal{D}_i|\Lambda_j) = \int_{\mathbb{R}_+^q} p(\bar{\gamma}^{(i)}|\theta) \Lambda_j(\theta) d\theta. \tag{24}
$$

This in general high-dimensional integral is approximated employing Monte Carlo integration, yielding

$$
c_j^{(i)} \approx \frac{1}{S_c} \sum_{k=1}^{S_c} p(\bar{\gamma}^{(i)}|\theta^{(k)}), \tag{25}
$$

in which $\theta^{(k)}$ is drawn from Λ_j, $\theta^{(k)} \sim \Lambda_j$, and S_c is the total size of the Monte Carlo sample $\{\theta^{(k)}\}_k$. If Λ_j allows for an efficient drawing of samples, the computational cost of determining $c_j^{(i)}$ is reasonable, requiring S_c simulations of the single cell model (1).

Given these precomputed $c_j^{(i)}$'s, which are independent of the density parameters φ, the conditional probability can be simplified to

$$
p(\mathcal{D}_i|\Theta_\varphi) = \sum_{j=1}^{n_\varphi} \varphi_j c_j^{(i)} = \left\langle \varphi, c^{(i)} \right\rangle, \tag{26}
$$

in which $c^{(i)} = [c_1^{(i)}, \ldots, c_{n_\varphi}^{(i)}]^T \in \mathbb{R}_+^{n_\varphi}$ and $\langle \cdot, \cdot \rangle$ denotes the scalar product. Employing (26) the posterior probability can be written as

$$p(\Theta_\varphi | \mathcal{D}) = \frac{\prod_{i=1}^M \langle \varphi, c^{(i)} \rangle p(\Theta_\varphi)}{\int_{[0,1]^{n_\varphi}} \prod_{i=1}^M \langle \varphi, c^{(i)} \rangle p(\Theta_\varphi) d\varphi}, \qquad (27)$$

where the prior probability,

$$p(\Theta_\varphi) = \begin{cases} p(\Theta_\varphi) & \text{for } \Theta_\varphi \geq 0 \ \wedge \sum_{j=1}^{n_\varphi} \varphi_j = 1 \\ 0 & \text{otherwise} \end{cases}, \qquad (28)$$

enforces the satisfaction of the constraint of Θ_φ being a probability density. Note that for parameter estimation often only the shape of the posteriori probability density is of interest, and not the normalization. Therefore, we only consider

$$p(\Theta_\varphi | \mathcal{D}) \propto q(\Theta_\varphi | \mathcal{D}) = \prod_{i=1}^M \langle \varphi, c^{(i)} \rangle p(\Theta_\varphi), \qquad (29)$$

in which $q(\Theta_\varphi | \mathcal{D})$ is the unnormalized posterior probability. Sampling from $q(\Theta_\varphi | \mathcal{D})$ and $p(\Theta_\varphi | \mathcal{D})$ will yield the same distribution of sample members. Furthermore, $q(\Theta_\varphi | \mathcal{D})$ and $p(\Theta_\varphi | \mathcal{D})$ have the same extrema.

Computation of maximum posterior probability estimate

Given the simplified unnormalized posterior probabilities $q(\Theta_\varphi | \mathcal{D})$ one important question is which parameter density Θ_φ maximizes $q(\Theta_\varphi | \mathcal{D})$. This is the most likely parameter density for the measured data and the prior knowledge.

This optimal parameter density Θ_{φ^*} can be computed by solving the optimization problem

$$\varphi^* = \arg\max_\varphi \left[q(\Theta_\varphi | \mathcal{D}) = \prod_{i=1}^M \langle \varphi, c^{(i)} \rangle p(\Theta_\varphi) \right]$$

$$s.t. \quad \mathbf{1}^T \varphi = 1 \qquad (30)$$

$$\sum_{j=1}^{n_\varphi} \varphi_j \Lambda_j(\theta) \geq 0, \ \forall \theta \in \mathbb{R}_+^q,$$

in which the two constraints ensure that the obtained density is positive and has integral one. Note that as Λ_j is a probability density, Θ_{φ^*} is positive if all φ_j are positive. Employing this and optimizing $-\log(q(\Theta_\varphi | \mathcal{D}))$ instead of $q(\Theta_\varphi | \mathcal{D})$, (30) can be simplified to

$$\varphi^* = \arg\min_\varphi \ - \left(\sum_{i=1}^M \log\left(\langle \varphi, c^{(i)} \rangle \right) + \log(p(\Theta_\varphi)) \right)$$

$$s.t. \quad \mathbf{1}^T \varphi = 1 \qquad (31)$$

$$\varphi \geq 0.$$

This minimization problem can for concave $p(\Theta_\varphi)$ be solved rather efficiently, as in such case (31) is a convex optimization problem [25]. For this problem solvers exist which guarantee convergence to the global optimum in polynomial time, e.g the interior point method [25].

Uncertainty of parameter density

In the previous section a method is presented which allows the computation of the maximum posterior probability estimate Θ_{φ^*}. As measurement data are limited and noise corrupted this estimate will not be the true parameter density. Hence, the uncertainty of the parameter density has to be evaluated.

Sampling of posterior probability density In order to analyze the uncertainty of the estimate, a sample of the posterior probability density $q(\Theta_\varphi | \mathcal{D})$ is generated. This is possible, as the unnormalized posterior probability of a distribution $\Theta_{\varphi^{(k)}}$ can be evaluated efficiently given (24) - (28). In this work the sampling is performed with a classical Metropolis-Hastings method [19]. Also Gibbs or slice sampling approaches may be employed. Compared to importance and rejection sampling these methods are well suited as they do not require the selection of an appropriate proposal density, a task which is difficult in this case.

The main point of concern when using MCMC sampling for the problem at hand is that the prior probability and the posterior probability respectively are only non-zero on a $(n_\varphi - 1)$-dimensional subset of the density parameter space (28). This is due to the fact that the sum over the elements of φ has to be one for Θ_φ being a probability density. If a standard proposal step was used, the acceptance rate would have been zero.

This problem can be overcome by performing the sampling in the $(n_\varphi - 1)$-dimensional space, $[\varphi_1, \ldots, \varphi_{n_\varphi-1}]^T \in \mathbb{R}^{n_\varphi-1}$, and computing the remaining density parameter via the closing condition $\varphi_{n_\varphi} = 1 - \sum_{j=1}^{n_\varphi-1} \varphi_j$. According to this the update step for φ consists of two steps:

1. Draw proposals $[\varphi_1^{k+1}, \ldots, \varphi_{n_\varphi-1}^{k+1}]^T$ from the $(n_\varphi - 1)$-dimensional reduced proposal density T_r,

$$\begin{bmatrix} \varphi_1^{k+1} \\ \vdots \\ \varphi_{n_\varphi-1}^{k+1} \end{bmatrix} \sim T_r \left(\begin{bmatrix} \varphi_1^{k+1} \\ \vdots \\ \varphi_{n_\varphi-1}^{k+1} \end{bmatrix} \middle| \begin{bmatrix} \varphi_1^{k} \\ \vdots \\ \varphi_{n_\varphi-1}^{k} \end{bmatrix} \right). \qquad (32)$$

2. Determine $\varphi_{n_\varphi}^{k+1}$ such that $\sum_{j=1}^{n_\varphi} \varphi_j^{k+1} = 1$,

$$\varphi_{n_\varphi}^{k+1} = 1 - \sum_{j=1}^{n_\varphi-1} \varphi_j^{k+1}. \qquad (33)$$

In this work, the reduced proposal density is chosen to be a multivariate normal distribution, $T_r = \mathcal{N}([\varphi_1^k, \ldots, \varphi_{n_\varphi-1}^k]^T, \Sigma_{T_r})$, with covariance matrix Σ_{T_r}.

This two-step proposal generation procedure is in the following denoted by $\varphi^{k+1} \sim T(\varphi^{k+1}|\varphi^k)$. The proposed density parameter vector φ^{k+1} is accepted with probability

$$p_a(\varphi^{k+1}|\varphi^k) = \begin{cases} \min\left\{1, \dfrac{q(\Theta_{\varphi^{k+1}}|\mathcal{D})}{q(\Theta_{\varphi^k}|\mathcal{D})} \dfrac{T(\varphi^k|\varphi^{k+1})}{T(\varphi^{k+1}|\varphi^k)}\right\} & \text{for } \varphi^{k+1} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (34)$$

The distinction of the two cases is hereby crucial to ensure that only probability densities $\Theta_{\varphi^{k+1}}$ which are greater than zero for all $\theta \in \mathbb{R}_+^q$ are accepted.

By combining update and acceptance step one obtains an algorithm which draws a sample of weighting vectors $\{\varphi^k\}_{k=1}^{S_\varphi}$, or respectively parameter densities $\{\Theta_{\varphi^k}\}_{k=1}^{S_\varphi}$, from the posterior distribution. The number of sample members is thereby S_φ. The pseudo code for the routine is given in Algorithm 1. In particular, the facts that

- the conditional probabilities $c_j^{(i)}$ are only computed once in the beginning, and that
- every evaluation of the acceptance probability p_a requires only a small number of algebraic operations,

ensure hereby an efficient sampling. Without this reformulation the integral defining the conditional probability $p(\mathcal{D}|\Theta)$ would have to be computed in each update step. The resulting computational effort would be very large.

Algorithm 1 Sampling of posteriori distribution $p(\Theta_\varphi|\mathcal{D})$

   Require: data $\mathcal{D}$, prior $p(\Theta_\varphi)$, model $p(y|\theta)$, ansatz functions $\{\Lambda_j\}_{j=1}^{n_\varphi}$, initial point φ^0.

       Calculation of conditional probabilities $c_j^{(i)}$ employing $p(y|\theta)$.

       Initialize the Markov Chain with φ^0.

       for $k = 1$ to S_φ **do**

           Given φ^i propose φ^{k+1} from proposal density $T(\varphi^{k+1}|\varphi^k)$.

           Calculate posterior probability $q(\Theta_{\varphi^{k+1}}|\mathcal{D}) = \prod_{i=1}^M \langle \varphi, c^{(i)} \rangle p(\Theta_\varphi)$.

           Generate uniformly distributed random number $r \in [0,1]$.

           if $r < p_a(\varphi^{k+1}|\varphi^k)$ **then**

               Accept proposed parameter vector φ^{k+1}.

           else

               Restore previous parameter vector, $\varphi^{k+1} = \varphi^k$.

           end if

       end for

end

Bayesian confidence intervals The sample $\{\Theta_{\varphi^k}\}_{k=1}^{S_\varphi}$ generated by Algorithm 1 contains information about the shape of the posterior density $p(\Theta_\varphi|\mathcal{D})$. This information can be employed to determine the Bayesian confidence intervals, also called credible intervals.

In this work an approach is presented which employs the percentile method [17] to analyze the uncertainty of Θ_φ. The 100α-th percentile of a random variable r is the value $\bar{r}^{(\alpha)}$ below which 100α % of the observations fall. Accordingly, the $100(1-\alpha)$-th percentile interval of r is defined as $[\bar{r}^{(\alpha/2)}, \bar{r}^{(1-\alpha/2)}]$. The Bayesian confidence interval is frequently defined as the 95-th percentile interval [18]. Thus, the parameter is contained in $[\bar{r}^{(0.025)}, \bar{r}^{(0.975)}]$ with a probability of 95% given the measurement data and the prior knowledge.

For the problem of estimating parameter densities, the 100α-th percentile is not simply a number but a function. As we are interested in the confidence intervals of $\Theta_\varphi(\theta)$, the percentiles are defined point-wise for every θ. The 100α-th percentile of $\Theta_\varphi(\theta)$ is thus the function $\bar{\Theta}_\varphi^{(\alpha)}(\theta)$ which gives for every parameter vector θ the value under which 100α % of the observations fall,

$$\forall \theta \in \mathbb{R}_+^q : \quad \Pr(\Theta_\varphi(\theta) \leq \bar{\Theta}_\varphi^{(\alpha)}(\theta)) = \alpha. \quad (35)$$

Consequently, the $100(1-\alpha)\%$ Bayesian confidence interval $\mathrm{CI}_\Theta^{(1-\alpha)}(\theta)$ of $\Theta_\varphi(\theta)$ is defined as

$$\mathrm{CI}_\Theta^{(1-\alpha)}(\theta) = [\bar{\Theta}_\varphi^{(\alpha/2)}(\theta), \bar{\Theta}_\varphi^{(1-\alpha/2)}(\theta)]. \quad (36)$$

As the sample $\{\Theta_{\varphi^k}\}_{k=1}^{S_\varphi}$ is given, an approximation of $\bar{\Theta}_\varphi^{(\alpha/2)}$ and $\bar{\Theta}_\varphi^{(1-\alpha/2)}$ can be obtained by studying the percentiles of the sample [26]. For instance the nearest rank method or linear interpolation between closest ranks can be used to determine $\bar{\Theta}_\varphi^{(\alpha)}$.

Predictions of output density

As the parameter density is not known precisely, also the model predictions show uncertainties. To evaluate the reliability of the population model and its predictive power, these prediction uncertainties have to be quantified. Therefore, the Bayesian confidence interval of the prediction around the output density with the highest a posteriori probability density,

$$\Upsilon^*(\gamma|t, \Theta_{\varphi^*}) = \sum_{j=1}^{n_\varphi} \varphi_j^* \Upsilon(\gamma|t, \Lambda_j), \quad (37)$$

is determined.

The $100(1-\alpha)\%$ confidence intervals of the predictions $\mathrm{CI}_\Upsilon^{(1-\alpha)}(\gamma|t)$ are again defined via the percentile method,

$$\mathrm{CI}_\Upsilon^{(1-\alpha)}(\gamma|t) = [\bar{\Upsilon}^{(\alpha/2)}(\gamma|t, \Theta_\varphi), \bar{\Upsilon}^{(1-\alpha/2)}(\gamma|t, \Theta_\varphi)], \quad (38)$$

in which the 100α-percentile $\bar{\Upsilon}^{(\alpha)}(\gamma|t, \Theta_\varphi)$ of the predicted out put $\Upsilon(y|t,\Theta_\varphi)$ is defined as

$$\forall \gamma \in \mathbb{R}_+^m, \forall t: \quad \Pr(\Upsilon(\gamma|t, \Theta_\varphi) \le \bar{\Upsilon}^{(\alpha)}(\gamma|t, \Theta_\varphi)) = \alpha. \quad (39)$$

Computing $\bar{\Upsilon}^{(\alpha)}(\gamma|t, \Theta_\varphi)$ for an experiment is a three step procedure. At first, the outputs $\Upsilon(y|t,\Lambda_j)$ (12) are computed. Given $\Upsilon(y|t,\Lambda_j)$ and the sample from the posterior density $\{\Theta_{\varphi^k}\}_{k=1}^{S_\varphi}$, a sample from the predicted output density $p(\Upsilon|\mathcal{D})$ is given by

$$\left\{\Upsilon(\gamma|t, \Theta_{\varphi^k})\right\}_{k=1}^{S_\varphi} = \left\{\sum_{j=1}^{n_\varphi} \varphi_j^k \Upsilon(\gamma|t, \Lambda_j)\right\}_{k=1}^{S_\varphi}. \quad (40)$$

This sample can be used to approximate the prediction confidence interval $\text{CI}_\Upsilon^{(1-\alpha)}(\gamma|t)$. As the population model has to be simulated only n_φ times, this calculation is computationally tractable.

To sum up, in this section a method for the estimation of parameter distributions in heterogeneous cell populations from population data has been presented. It has been shown that the optimal value as well as the Bayesian confidence intervals can be computed efficiently employing a parameterization of the parameter density. Also a method to determine prediction uncertainties has been presented. This allows an in-depth analysis of the reliability of the model. A summary of the procedure is shown in Figure 4.

Results and discussion

To illustrate the properties of the proposed methods, artificial measurement data of a cell population responding to a tumor necrosis factor (TNF) stimulus will be analyzed. For illustration purposes, we consider a case where only one parameter is distributed in a first step. In a second step, we show that the method is also applicable in the case of multi-parametric heterogeneity.

In multicellular organisms, the removal of infected, malfunctioning, or no longer needed cells is an important issue. Therefore, multicellular organisms developed different mechanisms to externally enforce cell death. Thereby the signaling molecule TNF is one of the key players.

TNF can bind to specific death receptors in the cell membrane and is able to induce programmed cell death, also called apoptosis, via the activation of the caspase cascade. On the other hand, it promotes cell survival via the inflammatory response, specifically activation of the NF-κB pathway [27]. The proportion of the activation of these two signaling pathways decides about the fate of the single cell. In the following a simple model for the caspase and NF-κB activation is studied which

reproduces the main features of the single cell response to a TNF stimulus.

Model of TNF signaling pathway

The model considered in this study has been introduced in [20] and is based on known activating and inhibitory interactions among key signaling proteins of the TNF pathway. A schematic is shown in Figure 5. Besides active caspase 8 (C8a) and active caspase 3 (C3a), the nuclear transcription factor κB (NF-κB) and its inhibitor I-κB are considered in the model. The model is given by the ODE system

$$\begin{aligned}
\dot{x}_1 &= -x_1 + \frac{1}{2}(\text{inh}_4(x_3)\text{act}_1(u) + \text{act}_3(x_2)) \\
\dot{x}_2 &= -x_2 + \text{act}_2(x_1)\text{inh}_3(x_3) \\
\dot{x}_3 &= -x_3 + \text{inh}_2(x_2)\text{inh}_5(x_4) \\
\dot{x}_4 &= -x_4 + \frac{1}{2}(\text{inh}_1(u) + \text{act}_4(x_3)).
\end{aligned} \quad (41)$$

The state variables x_i, $i = 1, ..., 4$ denote the relative activities of the signaling proteins C8a, C3a, NF-κB and I-κB, in this order. The functions $\text{act}_j(x_i)$ and $\text{inh}_j(x_i)$ represent activating and inhibiting interactions, respectively. They are given by

$$\text{act}_j(x_i) = \frac{x_i^2}{a_j^2 + x_i^2}, j \in \{1, \ldots, 4\}, \quad (42)$$

and

$$\text{inh}_j(x_i) = \frac{b_j^2}{b_j^2 + x_i^2}, j \in \{1, \ldots, 5\}. \quad (43)$$

The parameters a_j and b_j are activation and inhibition thresholds, respectively, and take values between 0 and 1. The external TNF stimulus is denoted by u. Initial conditions of the single cells are the steady states with C3a = 0 for u = 0. All nominal parameter values are given in Table 1.

It is known from experiments that the cellular response to a TNF stimulus is highly heterogeneous within a clonal cell population. Some cells die, others survive. The reasons for this heterogeneous behavior are unclear, but of great interest for biological research in TNF signaling, e.g. the use of TNF or related molecules as anti-cancer agent.

To understand the biological process at the physiological and biochemical level it is crucial to consider this cellular heterogeneity, using for example cell population modeling. Here, we model heterogeneity at the cell level via differences in the parameters b_3 and a_4. The parameter b_3 describes the inhibitory effect of NF-κB via the C3a inhibitor XIAP onto the C3 activity, and the
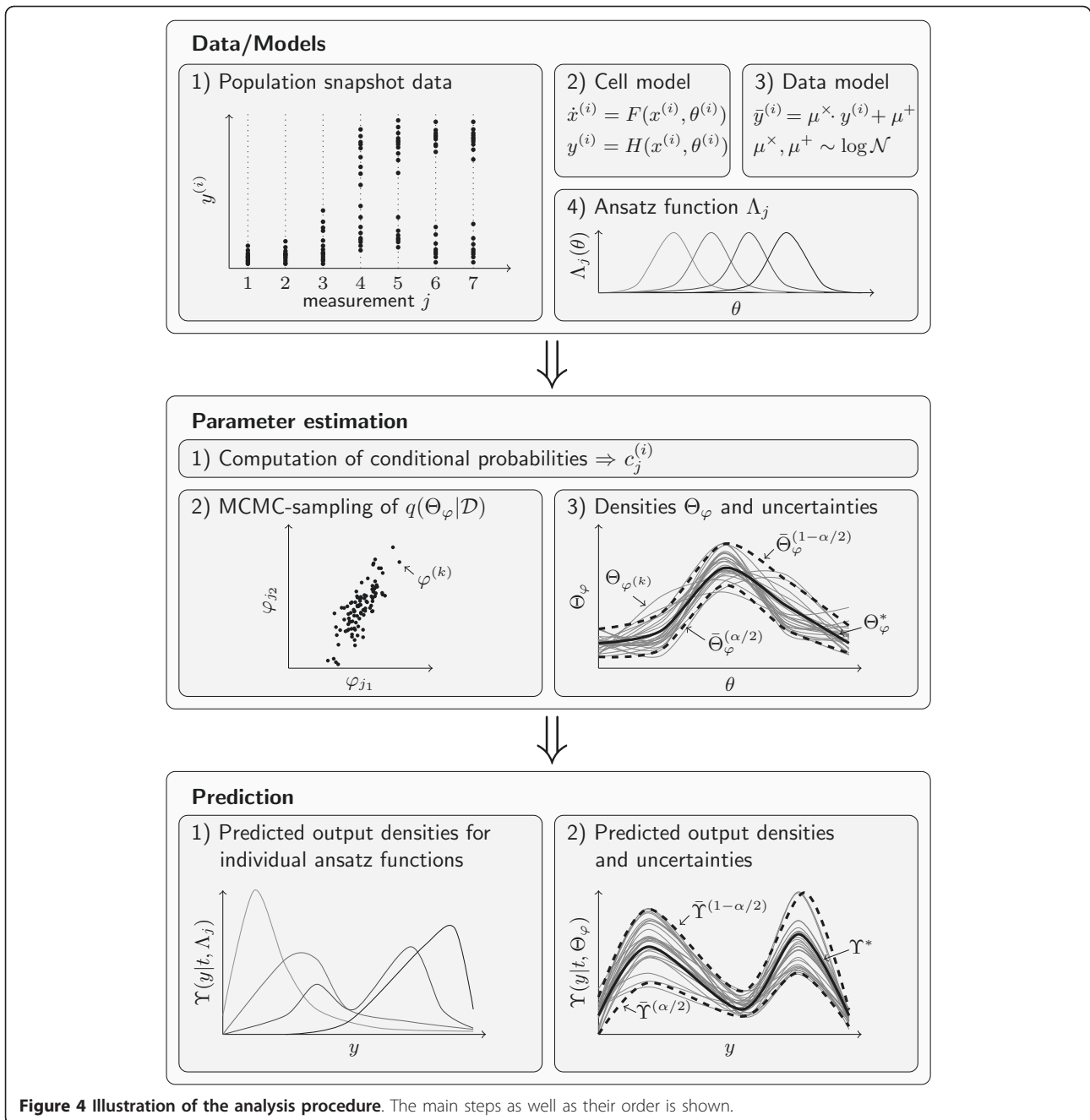
Figure 4 Illustration of the analysis procedure. The main steps as well as their order is shown.
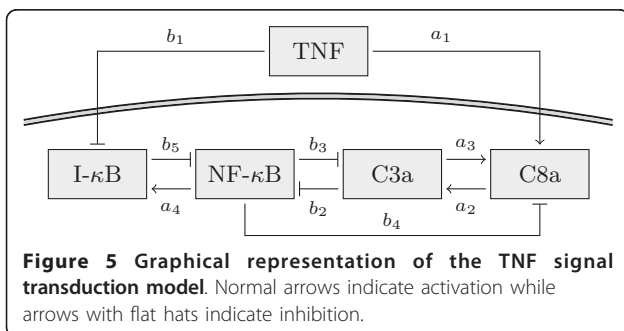

Figure 5 Graphical representation of the TNF signal transduction model. Normal arrows indicate activation while arrows with flat hats indicate inhibition.

Table 1 Nominal parameter values for the TNF signaling model (41)

i	1	2	3	4	5
a_i	0.6	0.2	0.2	0.5	
b_i	0.4	0.7	0.3	0.5	0.4

parameter a_4 models the activation of I-κB via NF-κB. Studies showed that these two interactions show large cell-to-cell variability [4,7,28].

Univariate parameter density

For a first evaluation of the proposed method an artificial experimental setup is considered in which the caspase 3 activity is measured at four different time instances during a TNF stimulus,

$$u(t) = \begin{cases} 1 \text{ for } t \in [0, 2] \\ 0 \text{ otherwise.} \end{cases} \quad (44)$$

At each time instance the C3a concentration in 150 cells is determined, $y = x_2$, with measurement noise according to (7), where $\mu^\times = 0$, $\sigma^\times = 0.1$, $\mu^+ = \log(0.05)$, and $\sigma^+ = 0.3$. This corresponds to an average noise level of about 15%. The generated artificial experimental data for a bimodal distribution in b_3 are depicted in Figure 6.

As ansatz functions Λ_j for the estimation, we use $n_\varphi = 15$ truncated Gaussians

$$\Lambda_j(b_3) = \begin{cases} \dfrac{s_j}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(b_3 - \mu_j)^2}{2\sigma^2}\right) & b_3 \in [0, 1] \\ 0 & \text{otherwise,} \end{cases} \quad (45)$$

where $\sigma = \dfrac{1}{30} =$ and s_j such that $\int_0^1 \Lambda_j(b_3)db_3 = 1$. The center points μ_j are equidistantly distributed on the interval 0[1]. The prior probability $p(\Theta_\varphi)$ is chosen to be

$$p(\Theta_\varphi) = \begin{cases} \prod_{j=1}^{n_\varphi} p_\beta(\varphi_j|\alpha_j, \beta_j) & \text{for } \mathbf{1}^T\varphi = 1 \\ 0 & \text{otherwise,} \end{cases} \quad (46)$$

in which p_β is the probability density of the beta-distribution. The parameter α_j and β_j are selected such that p_β $(\varphi_j|\alpha_j|\beta_j)$ has its extremum at $\varphi_{j,\text{ext}} = \dfrac{1}{s_j} \Big/ \left(\sum_{i=1}^{n_\varphi} \dfrac{1}{s_i}\right)$ and convariance σ^2. The distribution of a sample $\{\varphi^k\}_k$ drawn from this prior is shown in Figures 7 and 8. Note that the prior does not enforce a
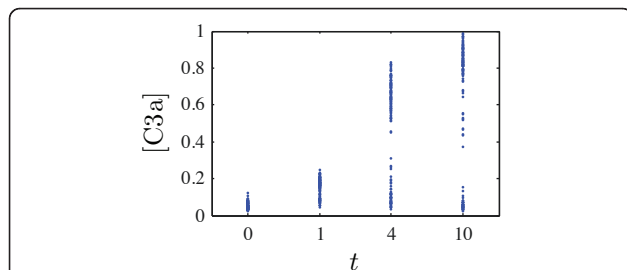

Figure 6 Artificial population snapshot data of C3a used to infer the parameter density within the cell population. Each *blue dot* represents a single measured cell.

trend to smaller or larger parameter values of b_3. Furthermore, it does not enforce a trend to unimodal or bimodal distributions Θ_φ (b_3). Such distribution properties shall be inferred from the data.

Given the ansatz functions Λ_j (45) the conditional probabilities $c_j^{(i)}$ of observing $\mathcal{D}_i = (\gamma^{(i)}, t^{(i)})$ are determined using importance sampling, according to (25). This computation takes about three minutes, on a standard personal computer using a single CPU. Thereby, 32% of the computation time are required for the simulation of the individual cells $y(t, \theta^{(j)})$ for individual parameter values $\theta^{(j)}$, and 59% for the evaluation of the conditional probability $p(\bar{\gamma}^{(i)}|\gamma(t^{(i)}, \theta^{(j)}))$. The rest is spent on pre- and post-processing. Subsequently, MCMC sampling is performed to obtain a sample $\{\varphi^k\}_{k=1}^{S_\varphi}$ of the prior (with $\sigma^2 = 0.05$), of the conditional, and of the posterior probability distribution. The sample has $S_\varphi = 10^6$ members and the generation takes only four minutes. The computation is very fast, as the proposed approach simplified the evaluation of the conditional probability to a matrix vector multiplication. Note, that all steps during the computation of the conditional probabilities and the MCMC sampling can be parallelized, yielding a tremendous speed-up for more complex models.

The results of the sampling are illustrated in Figure 7 using parallel coordinates [29]. From this representation of $\{\varphi^k\}_{k=1}^{S_\varphi}$ it can be seen that after the learning processes most of the density parameters still show large uncertainties. The uncertainty in the posterior distribution is a lot smaller than the uncertainty in the likelihood function, due to the stabilization via the prior. Note that the visualization also uncovers pronounced correlations between some parameters, e.g. φ_{10} and φ_{11} are negatively correlated for $\varphi^k \sim p(\mathcal{D}|\Theta_\varphi)$. This indicates that the model of the density of b_3 is over-parameterized with respect to the data. Thus, the number of ansatz functions could be reduced while still achieving a good fit.

To analyze the uncertainty of Θ_φ in more detail the sample $\{\Theta_{\varphi^k}\}_{k=1}^{S_\varphi}$ is employed to determine the 80%, 90%, 95%, and 99% Bayesian confidence intervals. The results are depicted in Figure 8. It can be seen that the confidence intervals for some values of b_3 are rather small, indicating that the data contain many information about these regions. Unfortunately, in particular for $b_3 > 0.6$ the confidence intervals are very wide showing that the parameter density in this area cannot be inferred precisely. But, although the amount of data is limited and the uncertainty with single φ_i's may be large, the posterior distribution of Θ_φ already shows key properties of the true parameter density, e.g. the bimodal shape,
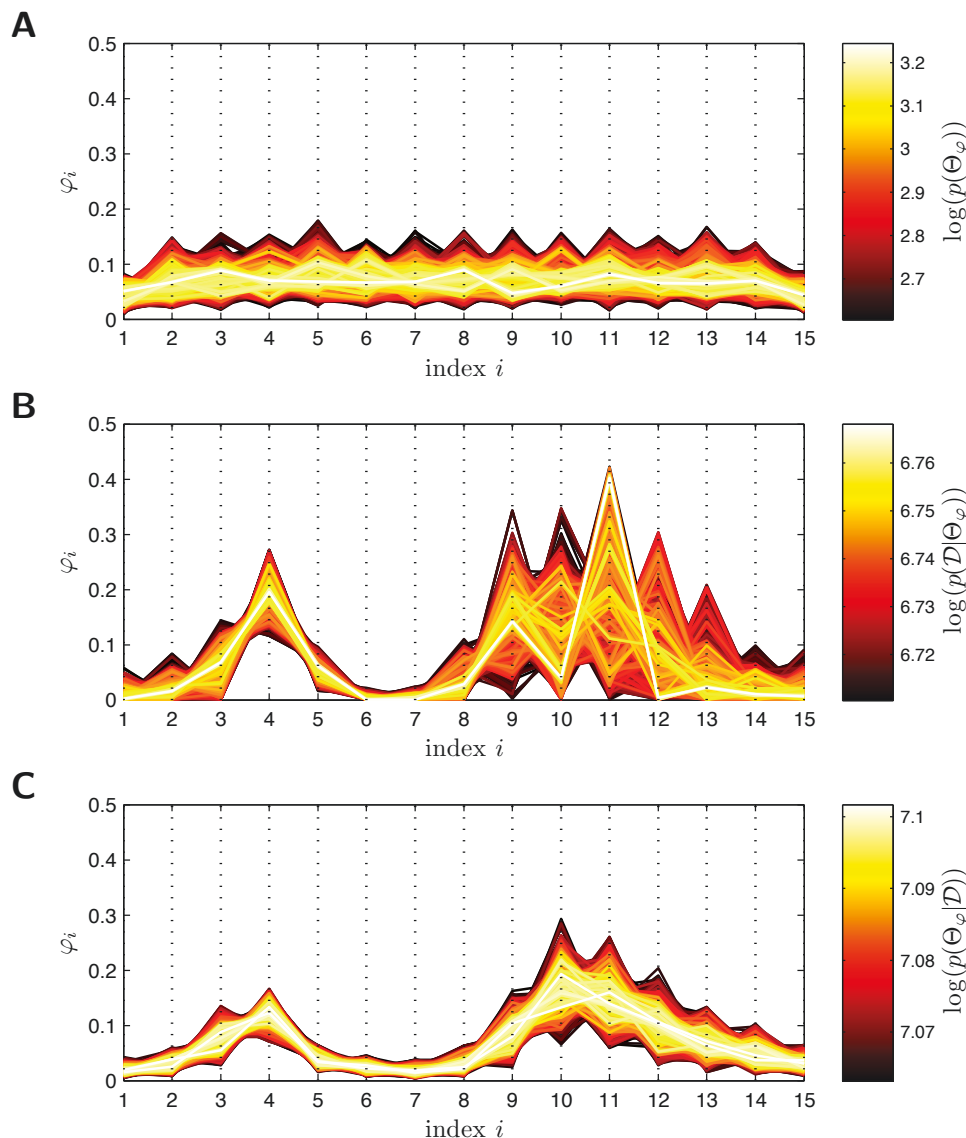
Figure 7 Visualization of 15-dimensional MCMC sample $\{\varphi^k\}_{k=1}^{S_\varphi}$ from prior, conditional and posterior probability density. **A**. Plot of sample $\{\varphi^k\}_{k=1}^{S_\varphi}$ drawn from prior probability density of $\Theta_\varphi(b_3)$, $p(\Theta_\varphi)$. **B**. Plot of sample $\{\varphi^k\}_{k=1}^{S_\varphi}$ drawn from conditional probability density of $\Theta_\varphi(b_3)$, $p(\mathcal{D}|\Theta_\varphi)$. **C**. Plot of sample $\{\varphi^k\}_{k=1}^{S_\varphi}$ drawn from posterior probability density of $\Theta_\varphi(b_3)$, $p(\Theta_\varphi|\mathcal{D})$. Each polyline represents hereby one point φ^k in the 15-dimensional density parameter space. The position of the vertex on the *i*-th dotted vertical line gives the value of the *i*-th density parameter. The color of the points indicates the logarithm of the unnormalized probability density of the data. Bright polylines (points φ^k) have a high posterior probability whereas dark polylines have a low posterior probability.

which has not been provided as prior information. This bimodal shape is also seen in the likelihood function, but there the uncertainties are larger than in the posterior probability distribution.

Besides the uncertainty of Θ_φ also the predictive power of the model can be evaluated. This is exemplified by studying the confidence interval of $\bar{\Upsilon}([\text{C3a}]|t, \Theta_\varphi)$ and $\bar{\Upsilon}([\text{NF} - \kappa\text{B}]|t, \Theta_\varphi)$ for the previously considered experimental setup. The bar indicates that the distribution of

the noise corrupted output $\bar{y}$ instead of the true output y is considered. This allows the direct comparison of the prediction with the data. The predictions are shown in Figure 9.

It is obvious that, although the parameter distributions show large uncertainties, the predictions are rather certain. This is indicated by tight confidence intervals. Furthermore, the mean predictions $\text{E}[\bar{\Upsilon}]$ and the predictions with highest posterior probability $\bar{\Upsilon}^*$ agree well
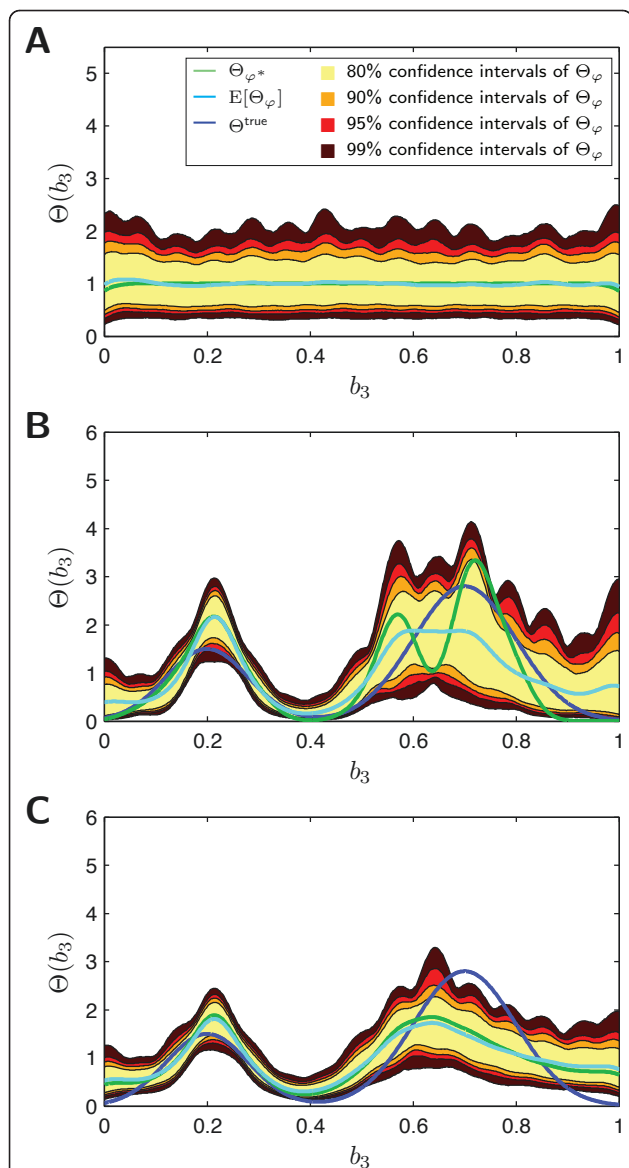
Figure 8 Prior, conditional and posterior probability density of Θ_φ (b_3) in b_3 - $\Theta(b_3)$ - plane. **A**. Plot of prior probability density of $\Theta_\varphi(b_3)$, $p(\Theta_\varphi)$. **B**. Plot of conditional probability density of $\Theta_\varphi(b_3)$, $p(\mathcal{D}|\Theta_\varphi)$. **C**. Plot of posterior probability density of $\Theta_\varphi(b_3)$, $p(\Theta_\varphi|\mathcal{D})$. The colored lines indicate the distribution with the highest posterior probability $\Theta_{\varphi*}$, and the mean distribution $E[\Theta_\varphi]$, for the individual probability densities, as well as the true parameter distribution Θ^{true}. The colored regions indicate the 80%, 90%, 95%, and 99% Bayesian confidence intervals of the parameter distribution Θ_φ.

Figure 9 Predicted measured output densities $\tilde{\Upsilon}([C3a]|t, \Theta_\varphi)$ and $\tilde{\Upsilon}([NF\text{-}\kappa B]|t, \Theta_\varphi)$. The colored lines indicate the distribution with the highest posterior probability $\tilde{\Upsilon}^*(\cdot|t, \Theta_{\varphi*})$, and the mean distribution $E[\hat{\Upsilon}^*(\cdot|t, \Theta_\varphi)]$, as well as the true measured output distribution $\tilde{\Upsilon}^{\text{true}}(\cdot|t, \Theta^{\text{true}})$. The colored regions indicate the 80%, 90%, 95%, and 99% Bayesian confidence intervals of the predicted distribution $\bar{\Upsilon}$.

with the true output distribution $\tilde{\Upsilon}^{\text{true}}$, for measured output C3a and predicted output NF-κB. The small prediction uncertainties can be explained to be sloppiness [30] of the density parameters φ_i parametrizing the distribution of b_3. A detailed analysis indicates (not shown here) that the number of ansatz function can be decreased, still ensuring a good approximation of the distribution of b_3.

Multivariate parameter density

In many biological systems several cellular parameters are heterogeneous and different cellular concentrations can be measured [7]. Therefore, we show in this section that the proposed method can also be employed to estimated multivariate parameter densities from multidimensional outputs. Furthermore, the influence of the choice of the prior on the estimation result is analyzed.

To perform this study we considered the same experimental setup as above. The only difference is that two concentrations are measured, C3a and NF-κB, $y = [x_2, x_3]^T$. The considered artificial experimental data of 10^4 cells are depicted in Figure 10. The ansatz function for Θ_φ are $n_\varphi = 100$ truncated multivariate Gaussians equivalently to (45). The covariance matrix is $0.06^2 \cdot I_2$ and the extrema are equidistantly distributed on a regular 2-dimensional grid which is aligned with the axes.
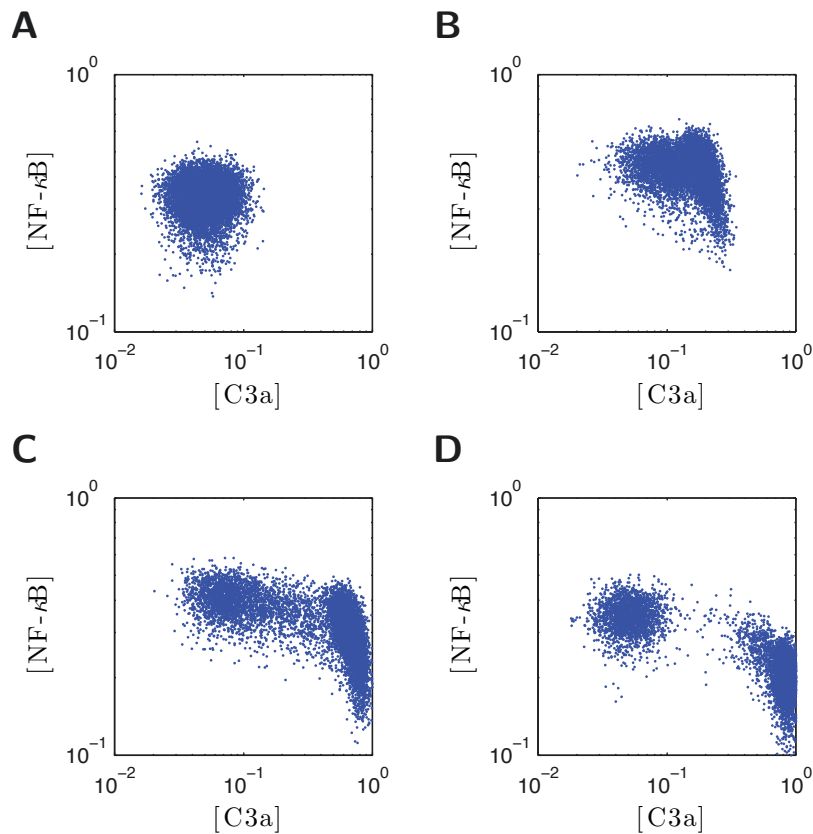
Figure 10 Artificial population snapshot data of C3a and NF-κB used to infer the parameter density within the cell population. **A**. Plot of snapshot data for $t = 0$. **B**. Plot of snapshot data for $t = 1$. **C**. Plot of snapshot data for $t = 4$. **D**. Plot of snapshot data for $t = 10$. Each *blue dot* represents a single measured cell.

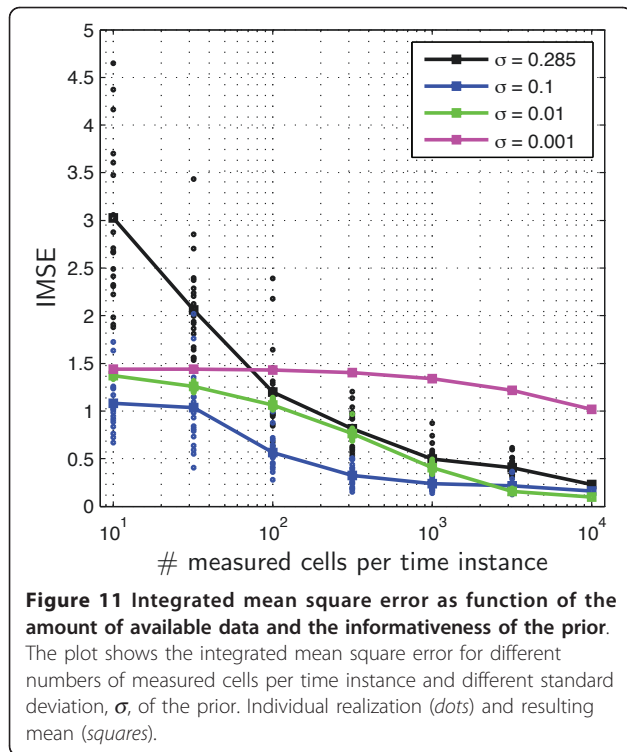Given this setup, the convergence rate is studied in terms of the integrated mean square error,

$$\text{IMSE} = \int_{[0,1]^2} \left| \Theta^{\text{ture}}(a_4, b_3) - \Theta_{\varphi^*}(a_4, b_3) \right|^2 da_4 db_3 \quad (47)$$

of true distribution and distribution with highest posterior probability Θ_{φ^*}. The IMSE is computed for amounts of measured cells per time instance and different priors. The priors are thereby again beta-distributions (46). The extrema φ_{ext} are chosen as in the last section such that the prior is flat. The standard deviation on the other hand is reduced step-wise from $\sigma = 0.285$ (completely uninformative as almost uniform on the feasible interval $\varphi \in [0,1]^{n_\varphi}$ to $\sigma = 0.001$ (very informative). Given this requirements, the values α_i and β_i of the prior (46) are determined. The result for different numbers of measured cells sampled from the available data set is shown in Figure 11. Note that the IMSE is a stochastic quantity as the selection of measured cells is a stochastic processes and hence also the estimated density Θ_{φ^*} is stochastic. To account for this stochasticity, several realizations are performed and the mean is computed.

From Figure 11 it becomes clear that the IMSE strongly depends on both, amount of data and informativeness of prior. For uninformative priors, the outcome for little data is highly uncertain and the IMSE is large and shows large variations. On the contrary, if the prior is very informative but wrong, the number of measurement data required to obtain a good estimate is tremendous. For the right choice of σ, one observes a fast convergence of the Θ_{φ^*} to Θ^{true}, as shown in Figure 12, and only little variation for small amounts of data. Hence, the usage of prior knowledge, even if it is only partially correct, yields for more stable estimates and faster convergence. Furthermore, this study suggests that the typical number of cells measured by flow cytometry (10^4) is informative enough to infer key features of population heterogeneity.

Conclusions

In this paper a Bayesian approach for inferring the parameter density for heterogenous cell populations with single cell resolution from population data is presented. We show that the proposed model can deal with limited and noisy measurement data as well as realistic noise

Figure 11 Integrated mean square error as function of the amount of available data and the informativeness of the prior. The plot shows the integrated mean square error for different numbers of measured cells per time instance and different standard deviation, *σ*, of the prior. Individual realization (*dots*) and resulting mean (*squares*).

models. The method utilizes a parameterization of the parameter density which, in combination with a reformulation of the conditional probability, allows a computationally efficient evaluation of the posterior probability. Compared to other methods for cell populations this approach does not rely on the approximation of the measured population response using density estimators.

For sampling from the posterior probability the Metropolis-Hastings algorithm is used. Here it has been adapted to be applicable to the considered constraint problem. Using this sampling strategy a sample from the posterior probability density is determined. This sample is employed to compute Bayesian confidence intervals for the parameter distribution, as well as for the model predictions. Also summary statistics like mean parameter density and mean predicted output density can easily be determined. For concave prior distributions we could even show that the calculation of the parameter density for the highest posterior probability is a convex problem.

The properties of the proposed scheme are evaluated using artificial data of a TNF signal transduction model. It could be shown that the proposed method yields good estimation results for a realistic experimental setup. Furthermore, although the remaining uncertainties are large, the predictions showg only small uncertainty indicating sloppiness of parameters.
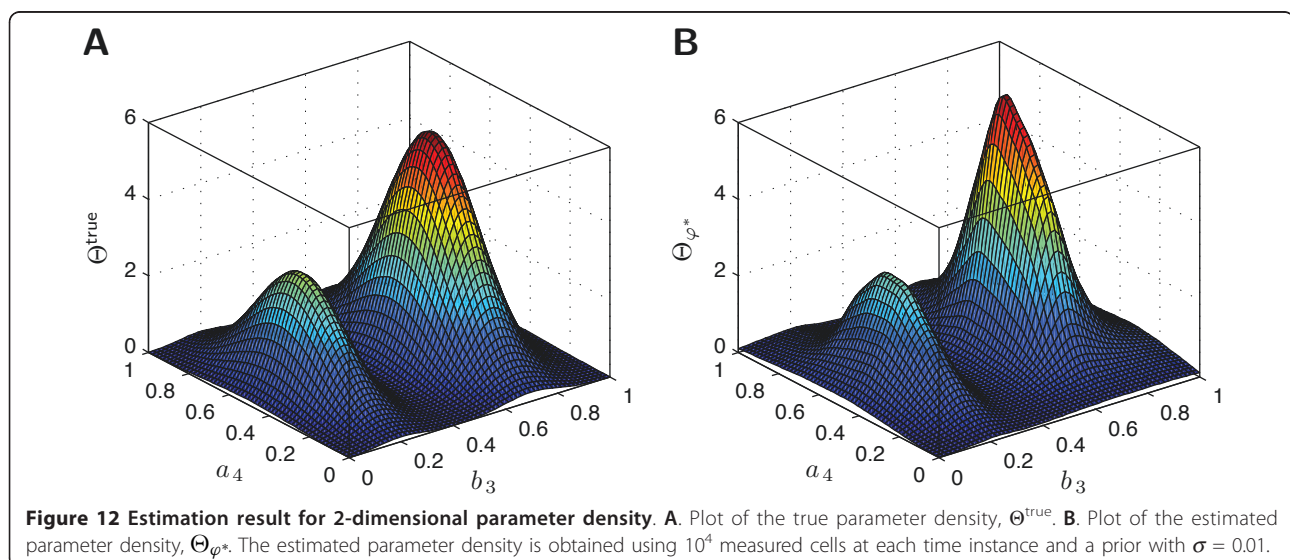
Concerning the choice of the prior distribution it could be shown that the regularizing effect is beneficial if only little data is available. On the other hand, if the amount of available data increases, informative but not carefully chosen priors slow down the convergence.

Author details

[1]Institute for Systems Theory and Automatic Control, University of Stuttgart, Germany. [2]Institute of Cell Biology and Immunology, University of Stuttgart, Germany.

Figure 12 Estimation result for 2-dimensional parameter density. **A**. Plot of the true parameter density, Θ^{true}. **B**. Plot of the estimated parameter density, Θ_{φ^*}. The estimated parameter density is obtained using 10^4 measured cells at each time instance and a prior with $\sigma = 0.01$.

Authors' contributions

References

1. Henson MA: Dynamic modeling of microbial cell populations. *Curr Opin Biotechnol* 2003, **14(5)**:460-467.
2. Mantzaris NV: From single-cell genetic architecture to cell population dynamics: Quantitatively decomposing the effects of different population heterogeneity sources for a genetic network with positive feedback architecture. *Biophys J* 2007, **92(12)**:4271-4288.
3. Munsky B, Trinh B, Khammash M: Listening to the noise: random fluctuations reveal gene network parameters. *Mol Syst Biol* 2009, **5(318)**:1-7.
4. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK: Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nat* 2009, **459(7245)**:428-433.
5. Stamatakis M, Zygourakis K: A mathematical and computational approach for integrating the major sources of cell population heterogeneity. *J Theor Biol* 2010, **266(1)**:41-61.
6. Avery SV: Microbial cell individuality and the underlying sources of heterogeneity. *Nat Rev Microbiol* 2006, **4**:577-587.
7. Albeck JG, Burke JM, Aldridge BB, Zhang M, Lauffenburger DA, Sorger PK: Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. *Mol Cell* 2008, **30(1)**:11-25.
8. Waldherr S, Hasenauer J, Allgöwer F: Estimation of biochemical network parameter distributions in cell populations. *Proc. of the 15th IFAC Symp. on Syst. Ident* 2009, **15**:1265-1270.
9. Al-Banna MK, Kelman AW, Whiting B: Experimental design and efficient parameter estimation in population pharmacokinetics. *J Pharmacokin Biopharm* 1990, **18(4)**:347-360.
10. Banks HT, Suttona KL, Thompson WC, Bocharov G, Roose D, Schenkel T, Meyerhans A: Estimation of cell proliferation dynamics using CFSE data. *Bull Math Biol* 2010, **73(1)**:116-150.
11. Luzyanina T, Roose D, Bocharov G: Distributed parameter identification for label-structured cell population dynamics model using CFSE histogram time-series data. *J Math Biol* 2009, **59(5)**:581-603.
12. Luzyanina T, Roose D, Schenkel T, Sester M, Ehl S, Meyerhans A, Bocharov G: Numerical modelling of label-structured cell population growth using CFSE distribution data. *Theor Biol Med Model* 2007, **4(26)**:1-14.
13. Hasenauer J, Waldherr S, Doszczak M, Scheurich P, Allgöwer F: Density-based modeling and identification of biochemical networks in cell populations. In *Proc. of 9th Int. Symp. on Dynamics and Control of Process Syst. (DYCOPS 2010), Leuven, Belgium, July 5-7* Edited by: Kothare M, Tade M, Wouwer AV, Smets I 2010, 306-311.
14. Hasenauer J, Waldherr S, Radde N, Doszczak M, Scheurich P, Allgöwer F: A maximum likelihood estimator for parameter distributions in heterogeneous cell populations. *Procedia Computer Science* 2010, **1(1)**:1649-1657.
15. Klinke DJ: An empirical Bayesian approach for model-based inference of cellular signaling networks. *BMC Bioinf* 2009, **10(371)**:1-18.
16. Wilkinson DJ: Bayesian methods in bioinformatics and computational systems biology. *Briefings in Bioinf* 2007, **8(2)**:109-116.
17. Joshi M, Seidel-Morgenstern A, Kremling A: Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Eng* 2006, **8**:447-455.
18. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinf* 2009, **25(25)**:1923-1929.
19. MacKay DJC: *Information Theory, Inference, and Learning Algorithms* Cambridge University Press; 2005.
20. Chaves M, Eissing T, Allgöwer F: Bistable biological systems: A characterization through local compact input-to-state stability. *IEEE Trans Autom Control* 2008, **53**:87-100.
21. Kreutz C, Bartolome Rodriguez MM, Maiwald T, Seidl M, Blum HE, Mohr L, Timmer J: An error model for protein quantification. *Bioinf* 2007, **23(20)**:2747-2753.
22. Gander W, Gautschi W: Adaptive quadrature-revisited. *Bit Numerical Mathematics* 2000, **40(18)**:84-101.
23. Silverman BW: *Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability* London: Chapman and Hall; 1986.
24. Surulescu C, Surulescu N: A nonparametric approach to cells dispersal. *Int J Biomath Biostat* 2010, **1**:109-128.
25. Boyd S, Vandenberghe L: *Convex Optimisation* Cambridge University Press, UK; 2004.
26. DiCiccio TJ, Efron B: Bootstrap confidence intervals. *Statist Sci* 1996, **11(3)**:189-228.
27. Wajant H, Pfizenmaier K, Scheurich P: Tumor necrosis factor signaling. *Cell Death Differ* 2003, **10**:45-65.
28. Paszek P, Ryan S, Ashall L, Sillitoe K, Harper CV, Spiller DG, Rand DA, White MRH: Population robustness arising from cellular heterogeneity. *PNAS* 2010, **107(25)**:1-6.
29. Inselberg A, Dimsdale B: Parallel coordinates: a tool for visualizing multi-dimensional geometry. *Proc of IEEE Visualization* 1990, 361-378.
30. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 2007, **3(10)**:1871-1878.