2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Recursive Bayesian Estimation of Stochastic Rate Constants from Heterogeneous Cell Populations

C. Zechner, S. Pelet, M. Peter, and H. Koeppl

Abstract— **Robust estimation of kinetic parameters of intracellular processes requires large amounts of quantitative data. Due to the high uncertainty of such processes and the fact that recent single-cell measurement techniques have limited resolution and dimensionality, estimation should pool recordings of multiple cells of an isogenic cell population. However, experimental results have shown that several factors such as cell volume or cell-cycle stage can drastically affect signaling as well as protein expression, leading to inherent heterogeneities in the cell population measurements. Here we present a recursive Bayesian estimation procedure for stochastic kinetic model calibration using heterogeneous cell population data. While obtaining optimal estimates for the rate constants, this approach allows to reconstruct missing species as well as to quantitatively capture extrinsic variability. The proposed algorithm is applied to a model of the osmo-stress induced MAPK Hog1 activation in the cytoplasm and its translocation to the nucleus.**

I. Introduction

Although experimental techniques in molecular cell biology are advancing rapidly, quantitative data is still characterized by large uncertainties and low-dimensionality with respect to the complexity of the cellular process under study. On top of acquisition uncertainty one encounters fluctuations due to the inherent stochasticity of chemical kinetics and heterogeneity of the cell population from which the data is extracted. Performing two-color experiments, the latter was shown to dominate the former with respect the variability observed in single-cell measurements [1], [2]. Heterogeneity is present in a population of isogenic cells due to non-synchronized cell-cycle stage, difference in local growth conditions, difference in expression capacity and so forths. This cell-to-cell variability is commonly referred to as extrinsic noise [1]. The naming is unfortunate, as it indicates that these variation are inherently stochastic. Accordingly, mathematical accounts for cell-to-cell variability often choose to vary the kinetic rate constants by a stochastic process, such as the Ornstein-Uhlenbeck process [3], [4]. However, often single physiological states of the cell were shown to be good predictors of the variation of the above mentioned features [5] and cell-to-cell variability can to a large extent considered to be deterministic. For instance, cell volume increase was shown to align well with expression capacity with a correlation coefficient of 0.77 [2].

Several authors address the problem of calibrating a stochastic kinetic model to quantitative experimental data. Some neglect any contribution coming of cell-to-cell variability and thus attribute the variability solely to the stochasticity of chemical events. Approaches involve Markov-chain Monte-Carlo (MCMC) based Bayesian inference variants that either work on the jump process directly [6], [7], or on its diffusion approximation [8]. Moreover, probability metrics have been put forth as cost functions [9] to deploy general gradient-type algorithms for parameter estimation. Approaches that account for the heterogeneity of the data-generating population are limited. In a recent study, Rand and co-workers [4] outline an estimation scheme that accounts for extrinsic variability and also gives estimates for the strength and half-life of the stochastic process on the kinetic rate constants. In the context of differential equation models heterogeneity is considered in [10]. We recently proposed an novel approach to capture cell-to-cell variability and showed how to incorporate it into an estimation scheme [11]. The approach is based on the observation that the abundance of proteins or their concentrations vary from cell to cell, while the kinetics of elementary events, such as association and post-translational modification is determined by the biophysics of the interacting biomolecules and are thus invariant over a heterogeneous but isogenic population. Clearly, this model accounts just for one aspect of cell-to-cell variability and needs to be complemented with existing approaches in general.

In this article we alleviate many of the limitations of the estimation algorithm presented in [11]. In particular, the previous work assumes observation of all species as well as observation of the complete sample path – equivalent to resolving every single reaction. The novel method can cope with the important realistic scenario of having unobserved species and noisy subsampled paths as observations. It is based on a recursive Bayesian estimation scheme and is thus favorable in terms of complexity with respect to standard MCMC schemes. The scheme is novel as such, and is also applicable to the general situation of continous-time Markov chain (CTMC) inference.

The remaining part of the work is organized as follows. In Section II we first introduce a CTMC description for stochastic chemical reaction systems. In Section II-A, the basic notation is extended to statistically model the heterogeneity over cell population measurements. Section III is divided into two parts. In Section III-A, we introduce a general state space model which is compatible with real-world experimental data. A simple bootstrap filter is proposed, which allows sequential state and parameter estimation using a Metropolis-within-Gibbs scheme. Section III-B extends this approach for cell population data according to the Bayesian models

C. Zechner and H. Koeppl are with the Automatic Control Lab., Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland {zechner,koeppl}@control.ee.ethz.ch

S. Pelet and M. Peter are with the Institute for Biochemistry, Department of Biology, ETH Zurich, Switzerland {serge.pelet,matthias.peter}@bc.biol.ethz.ch

978-1-61284-799-3/11/$26.00 ©2011 IEEE

from sections II-A and II-B. Simulations are performed under realistic conditions in Section IV, where the proposed algorithms are applied to a model of the osmo-stress induced Hog1 activation in yeast.

II. Continuous-Time Markov Chains

In this work, we describe the temporal evolution of chemical reaction system with n species and v reactions as a CTMC $\mathbf{X}$ with state-space $\mathbb{Z}^n_{\geq 0}$. The *stoichiometry matrix* $\mathbf{S} \in \mathbb{Z}^{n \times v}$ is composed of v change vectors $\left(\boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-\right)$, $k \in \{1, \ldots, v\}$, where $\boldsymbol{\nu}_k^-$ and $\boldsymbol{\nu}_k^+$ count the number of molecules consumed and produced when the k-th reaction fires, respectively. We know from [12] that the state of the system can be written as

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_{k=1}^{v} \xi_k \left(\int_0^t a_k(\mathbf{X}(s), c_k)\ ds \right) (\boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-), \tag{1}$$

where $\mathcal{C} = \{c_1, \ldots, c_v\}$ is the set of stochastic rate constants, $\xi_k,\ k \in \{1, \ldots, v\}$ are unit Poisson processes and $a_k\left(\mathbf{x}, c_k\right) = c_k g_k(\mathbf{x}), \mathbf{x} \in \mathbb{Z}^n_{\geq 0}$, are the reaction propensities for reaction k with function g_k given by the law of mass-action. Furthermore, we define the total reaction propensity as $a\left(\mathbf{x}, \mathcal{C}\right) = \sum_{k=1}^{v} a_k\left(\mathbf{x}, c_k\right)$. The key quantity in estimating $\mathcal{C}$ from a particular sample path $\mathbf{X}_{[0,t_M]}$[1] is the conditional path density (or *likelihood* function)

$$\begin{aligned} p(\mathbf{X}_{[0,t_M]}|\mathcal{C}) = \pi_0(\mathbf{X}(0)) & \\ \times \prod_{j=1}^{M} \exp\left\{-a\left(\mathbf{X}(t_{j-1}), \mathcal{C}\right)\right)(t_j - t_{j-1})\} & \\ \times\, a_{r_j}\left(\mathbf{X}(t_{j-1}, c_{r_j})\right), & \end{aligned} \tag{2}$$

where $t_j, j \in \{1, \ldots, M\}$ are the times when the system state jumps (i.e., when a reaction occurs) and $\pi_0(\mathbf{X}(0))$ is the initial distribution over the system state [13], [11]. It is well known that given a complete sample path $\mathbf{X}_{[0,t_M]}$ and assuming priors $p(c_k) = \Gamma(\alpha_k, \beta_k)$, the posterior distribution over the k-th rate constants is given by

$$p\left(c_k|\mathbf{X}_{[0,t_M]}\right) = \Gamma\left(r_k + \alpha_k, \int_0^{t_M} g_k(\mathbf{X}(s))ds + \beta_k\right), \tag{3}$$

with r_k as the number of occurrences of reaction k in the interval $[0, t_M]$ and $\Gamma(\cdot, \cdot)$ as the Gamma distribution [13].

A. Modeling Extrinsic Variablity

To model the heterogeneity over a cell population (i.e., ensemble of CTMCs), we use the same approach as in [11], where extrinsic variability is expressed via certain conservation laws. First, we introduce the parameterized system state $\mathbf{Z}(t, \mathbf{b})$, as a solution to

$$\begin{aligned} \mathbf{Z}(t, \mathbf{b}) = \mathbf{Z}(0, \mathbf{b}) + & \\ \sum_{k=1}^{v} \xi_k \left(\int_0^t \tilde{a}_k(\mathbf{Z}(s, \mathbf{b}), \mathbf{b}, c_k)\ ds \right) (\boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-). & \end{aligned} \tag{4}$$

where $\mathbf{b} \in \mathbb{Z}^u_{\geq 0}$ is a vector specifying the total number of molecules for each set of conserved species. We furthermore assume this quantity to be *extrinsic* and hence, to vary from cell to cell. It follows that the new propensities $\tilde{a}_k(\mathbf{Z}(t, \mathbf{b}), \mathbf{b}, c_k)$, which we define as

$$\begin{aligned} \tilde{a}_k(\mathbf{x}, \mathbf{b}, c_k) &= a_k(\mathbf{x}, c_k) 1_{\{\mathbf{Nx}=\mathbf{b}\}} \\ &\text{with } \mathbf{x} \in \mathbb{Z}^n_{\geq 0} \end{aligned} \tag{5}$$

will also depend on $\mathbf{b}$. Note that $\{\mathbf{x} \in \mathbb{Z}^n_{\geq 0} : \mathbf{Nx} = \mathbf{b}\}$ defines the state-space of the (equivalent) Markov chains $\mathbf{Z}(t, \mathbf{b})$ and $\mathbf{X}(t)$ under presence of mass-conservation laws. For relation

$$\mathbf{Nx} = \mathbf{b} \tag{6}$$

we have defined $\mathbf{N} \in \mathbb{Z}^{u \times n}_{\geq 0}$ as the smallest positive integer base of the u-dimensional left null space Null $\{\mathbf{S}\}$. Rearranging equation (6), we obtain

$$\mathbf{Nx} = \begin{pmatrix} \tilde{\mathbf{N}} & \bar{\mathbf{N}} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{x}} \\ \bar{\mathbf{x}} \end{pmatrix} = \mathbf{b}. \tag{7}$$

Now, we can express a set of species $\bar{\mathbf{x}}$ as a function of $\tilde{\mathbf{x}}$ and $\mathbf{b}$, i.e.,

$$\bar{\mathbf{x}} = \bar{\mathbf{N}}^{-1}\left(\mathbf{b} - \tilde{\mathbf{N}}\tilde{\mathbf{x}}\right) \equiv \mathbf{F}\left(\tilde{\mathbf{x}}, \mathbf{b}\right). \tag{8}$$

Furthermore, when defining $\mathbf{X}(t) = (\,\tilde{\mathbf{X}}(t); \bar{\mathbf{X}}(t)\,)$ and reordering the dimensions of $\mathbf{Z}(t, \mathbf{b})$ accordingly, we can write $\mathbf{Z}(t, \mathbf{b}) = (\,\tilde{\mathbf{X}}(t); \mathbf{F}(\tilde{\mathbf{X}}(t), \mathbf{b})\,)$. Note that $\mathbf{X}(t)$ and $\mathbf{Z}(t, \mathbf{b})$ are still equivalent, but $\mathbf{Z}(t, \mathbf{b})$ gives rise to a data model governed by the conservation constant $\mathbf{b}$, which allows us to perform inference with respect to this quantity [11].

Rewriting the likelihood from (2) in terms of $\mathbf{Z}(t, \mathbf{b})$ is straight-forward and the full conditional distribution over a single rate constant again takes the form of a Gamma distribution.

B. A Hierarchical Bayesian Model

Using the previous considerations, we will now set up our extrinsic noise model. First of all, let us assume that we have complete measurement data observed from cells $m \in \{1, \ldots, L\}$, each of them giving rise to one particular Markov process $\mathbf{Z}(t, \mathbf{b}_m)$. We denote the set of conservation constants as $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_L\}$, where each of the components $\mathbf{b}_m$ is independently drawn from one common distribution, i.e., $\mathbf{b}_m \sim p(\mathbf{b}|\alpha)$, where α denotes a set of *hyperparameters* with *hyperpriors* $p(\alpha)$. The corresponding hierarchical Bayesian model can be seen in Figure 1. During model calibration, we are particularly interested in computing the posterior distribution

$$\begin{aligned} & p(\mathcal{C}, \alpha|\mathbf{Z}(\mathbf{b}_1)_{[0,t_M]}, \ldots, \mathbf{Z}(\mathbf{b}_L)_{[0,t_M]}) = \\ & \int p(\mathcal{C}, \mathcal{B}, \alpha|\mathbf{Z}(\mathbf{b}_1)_{[0,t_M]}, \ldots, \mathbf{Z}(\mathbf{b}_L)_{[0,t_M]}) d\mathcal{B}, \end{aligned} \tag{9}$$

which is analytically intractable. Anyway, it was shown in [11] that MCMC methods can be used to efficiently draw samples from (9), to calibrate the model, while accounting for extrinsic noise.

[1] Quantities with subscript $[a, b]$ denote piecewise constant functions, evaluated on the interval $[a, b]$.

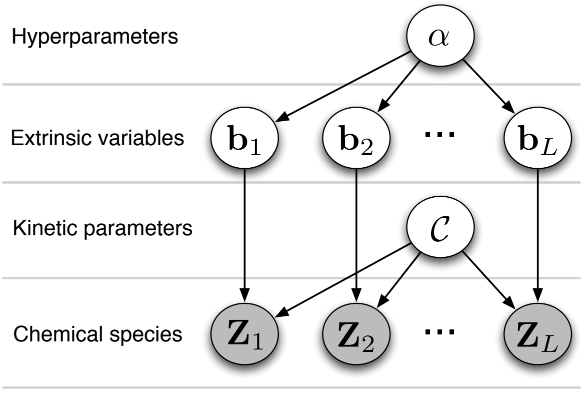

Fig. 1. Graphical model for a heterogeneous cell population. Shaded circles denote observed quantities. Furthermore, note that the L different CTMCs $\mathbf{Z}(t, \mathbf{b}_m), m \in \{1, \ldots, L\}$ are simply denoted as $\mathbf{Z}_m, m \in \{1, \ldots, L\}$.

III. Recursive Bayesian Estimation

Unfortunately, assuming complete experimental data is far from reality. Usually, we can only obtain noisy measurements of a small fraction of the involved species, captured at discrete time points. In the following we will describe how the probabilistic model from Section II can be extended to fit real-world single-cell as well as cell-population data. Furthermore, we propose a recursive Bayesian estimation procedure, being capable of solving the corresponding inference tasks.

A. Model Calibration on a Single Cell Level

For the moment let us go back to a single cell system, with $\mathbf{X}$ as the CTMC and $\mathcal{C}$ as a set of unknown parameters. Assuming noisy and incomplete measurements at consecutive time points t_l, $\forall l \in \{1, \ldots, N\}$, such a system can be written as a general state space model, governed by

1) a *state transition kernel*

$$p(\mathbf{X}(t_l)|\mathbf{X}(t_{l-1}), \mathcal{C}) \tag{10}$$

2) and a *measurement likelihood function*

$$p(\mathbf{Y}(t_l)|\mathbf{X}(t_l)). \tag{11}$$

If $\mathbf{X}$ describes a well-stirred, chemically reacting system, the transition kernel (10) is given by the *chemical master equation* (CME). The measurement likelihood function (11) should reflect the uncertainty, introduced by the experimental data acquisition. For instance, the measurement noise is often described as an additive and/or multiplicative component [14]. Here, we assume that the observation $\mathbf{Y}(t_l) \in \mathbb{R}^d$ with $d \leq n$ of a single cell at time point t_l is given as

$$\mathbf{Y}(t_l) = \mathbf{W}\mathbf{X}(t_l) + \boldsymbol{\epsilon}, \tag{12}$$

with $\mathbf{W} \in \mathbb{R}^{d \times n}$ and additive measurement noise $\boldsymbol{\epsilon}$, that we assume to be i.i.d. random variables with a certain distribution $p(\boldsymbol{\epsilon})$. The matrix $\mathbf{W}$ is usually a known quantity, given by the underlying experimental setup. For instance, we often cannot directly access certain chemical species, but only linear combinations of them, which would be reflected by $\mathbf{W}$. Also note that most measurement techniques are only capable of retrieving observations proportional to the quantity of interest. In this case, additional (unknown) scaling factors enter $\mathbf{W}$, which have to be tuned during model calibration. However, in this work we assume $\mathbf{W}$ to be entirely known.

In order to calibrate the model to the N discrete-time observations, we are interested in finding the posterior distribution $p(\mathcal{C}|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_N))$. Since $\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_N)$ depend on the parameters $\mathcal{C}$ over the latent states $\mathbf{X}_{[t_1,t_N]}$, this turns out to be a challenging task. Practically, one has to compute the joint posterior $p(\mathcal{C}, \mathbf{X}_{[t_1,t_N]}|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_N))$ and then marginalize with respect to $\mathbf{X}_{[t_1,t_N]}$ to infer $\mathcal{C}$ or vice versa. However, exact analytical analysis is still impossible for all but the simplest models and hence, several approximate solutions for directly sampling from $p(\mathcal{C}, \mathbf{X}_{[t_1,t_N]}|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_N))$ have been proposed in literature, e.g., [15] or [6]. As this is computationally demanding, we prefer using a comparably fast sequential approach, referred to as *recursive Bayesian estimation*. The latter is carried out by iteratively computing the following two quantities:

1) *Prediction* for time t_l:

$$\begin{aligned} p\left(\mathbf{X}(t_l)|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_{l-1}), \mathcal{C}\right) = \\ \sum_{\mathbf{X}(t_{l-1}) \in \mathbb{Z}^n_{\geq 0}} p\left(\mathbf{X}(t_l)|\mathbf{X}(t_{l-1}), \mathcal{C}\right) \times \\ p\left(\mathbf{X}(t_{l-1})|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_{l-1}), \mathcal{C}\right) \end{aligned} \tag{13}$$

2) *Correction* for time t_l:

$$\begin{aligned} p\left(\mathbf{X}(t_l)|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_l), \mathcal{C}\right) \propto p\left(\mathbf{Y}(t_l)|\mathbf{X}(t_l)\right) \times \\ p\left(\mathbf{X}(t_l)|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_{l-1}), \mathcal{C}\right) \end{aligned} \tag{14}$$

Note that except for linear and fully Gaussian models, equations (13) and (14) cannot be evaluated analytically. However, particle filters can be used to recursively draw samples from (13) and (14) based on importance sampling techniques. In this work, we utilize one simple but powerful variant of the particle filter, known as the *bootstrap* filter [16]. The only (remarkable) thing required by the algorithm is sampling from (10). However, we know that we can draw sample paths from a CTMC between two arbitrary, consecutive time points t_{l-1} and t_l using Gillespie's direct simulation algorithm [17], i.e., $\mathbf{x}_{[t_{l-1},t_l]} \sim p(\mathbf{X}_{[t_{l-1},t_l]}|\mathbf{X}(t_{l-1}), \mathcal{C})$. Hence, we obtain a sample $\mathbf{x}(t_l) \sim p(\mathbf{X}(t_l)|\mathbf{X}(t_{l-1}), \mathcal{C})$ by evaluating $\mathbf{x}_{[t_{l-1},t_l]}$ at t_l.

According to the concept of importance sampling [16], (13) can then be approximated by the mixture distribution

$$\begin{aligned} p\left(\mathbf{X}(t_l)|\mathbf{Y}(t_1), \ldots, \mathbf{Y}(t_{l-1}), \mathcal{C}\right) \approx \\ \sum_{i=1}^{P} w^{(i)} p\left(\mathbf{X}(t_l)|\mathbf{x}^{(i)}(t_{l-1}), \mathcal{C}\right), \end{aligned} \tag{15}$$

where $\mathbf{x}^{(i)}(t_{l-1})$ are samples drawn from the posterior distribution of the previous time step, i.e.,

$\mathbf{x}^{(i)}(t_{l-1}) \sim p\left(\mathbf{X}(t_{l-1})|\mathbf{Y}(t_1),\ldots,\mathbf{Y}(t_{l-2}),\mathcal{C}\right)$, with corresponding weights

$$w^{(i)} \propto p\left(\mathbf{Y}(t_{l-1})|\mathbf{x}^{(i)}(t_{l-1})\right). \tag{16}$$

Note that the weights $w^{(i)}$ have to sum up to one and thus, require to be normalized after computation. While the simple bootstrap filter already works for estimating the states of our continuous-time model, simultaneously estimating fixed model parameters is not covered by the standard algorithm, but can be easily integrated within the sequential Monte-Carlo framework. In this work we adopt a technique, which performs an additional Metropolis-within-Gibbs step with respect to the parameters after each resampling step as described in [18] or [19]. Simply speaking, at time t_l we predict the latent states on the next measurement interval $[t_l, t_{l+1}]$ given the most recent parameter values. Afterwards, at time t_{l+1} we resample the parameter vector for each particle i by incorporating the newly sampled states associated with that particle and so on. In order to do so, we need to sample from the full conditional distributions, i.e., $p(\mathbf{X}(t_l)|\mathbf{X}(t_{l-1}),\mathcal{C})$ and $p\left(c_k|\mathbf{X}_{[t_1,t_l]}\right)$. As already mentioned, we can easily draw from $p(\mathbf{X}(t_l)|\mathbf{X}(t_{l-1}),\mathcal{C})$ and the same holds for $p\left(c_k|\mathbf{X}_{[t_1,t_l]}\right)$ as it takes the form of a standard Gamma distribution. It is important to note that the conditional distribution $p\left(c_k|\mathbf{X}_{[t_1,t_l]}\right)$ naturally depends on the entire sample path $\mathbf{X}_{[t_1,t_l]}$, which means that each particle i is represented at time t_l by a set $\{\mathcal{C}^{(i)}, \mathbf{x}^{(i)}_{[t_1,t_l]}\}$. Consequently, the computational sampling effort increases with time. This, however, can be avoided if the parameter densities can be represented by low-dimensional sufficient statistics, which can be recursively updated as time increases [18]. This is the case for the conditional density $p\left(c_k|\mathbf{X}_{[t_1,t_l]}\right)$, as it only depends on the number of reactions of type k happened within $[t_1, t_l]$ and on $\int_{t_1}^{t_l} g_k(\mathbf{X}(s))ds$[2]. A possible realization of the sequential state and parameter estimation algorithm is summarized in Algorithm 1.

Algorithm 1 Bootstrap filter for joint state and parameter estimation given single cell measurements.

```
for $t_l \in \{t_2,\ldots,t_N\}$ do
    Resample $P$ particles according to their weights $w^{(i)}$
    for $i \in \{1,\ldots,P\}$ do
        {Resample parameters}
        for each rate constant $c_k^{(i)} \in \mathcal{C}^i$ do
            $c_k^{(i)} \sim$ eq. (3)
        end for
        {Simulate CTMC between $t_{l-1}$ and $t_l$}
        $\mathbf{x}^{(i)}(t_l) \sim$ eq. (10)
        {Compute importance weights}
        $w^{(i)} =$ eq. (16)
    end for
    Normalize weights such that $\sum_{i=1}^{P} w^{(i)} = 1$
end for
```

B. Model Calibration on a Cell Population Level

In the following we will pick up the concept from Section III-A and extend it for model calibration tasks based on measurements captured from heterogeneous cell populations. In order to do so, we first construct an ensemble state space model, collecting the individual system descriptions of each cell. Afterwards, it is straight-forward to design a bootstrap filter, which acts on this unified model. Similar approaches can be found in for instance [20].

Assume now that we are able to obtain discrete-time measurements of L independent cells. As we expect a certain heterogeneity over the cell population, we parameterize each cell's state by a quantity which is assumed to vary from cell to cell, as it was done in Section II-A. Then, each cell $m \in \{1,\ldots,L\}$ gives rise to one particular system state $\mathbf{Z}(t,\mathbf{b}_m)$ and output $\mathbf{Y}_m(t_l) = \mathbf{W}\mathbf{Z}(t_l,\mathbf{b}_m) + \boldsymbol{\epsilon}_m$, whereas we assume that the sampling points t_l are equivalent for each cell and that $\boldsymbol{\epsilon}_m$ are independent across cells. We denote the set of all states as

$$\mathcal{Z}(t) \equiv \left\{\mathbf{Z}(t,\mathbf{b}_m)\,|\, m \in \{1,\ldots,L\}\right\}.$$

Given $\mathcal{C}$ and $\mathcal{B}$, the transition kernel factorizes such that

$$\begin{aligned} &p\left(\mathcal{Z}(t_l)|\mathcal{Z}(t_{l-1}),\mathcal{C},\mathcal{B}\right) = \\ &\qquad \prod_{m=1}^{L} p\left(\mathbf{Z}(t_l,\mathbf{b}_m)|\mathbf{Z}(t_{l-1},\mathbf{b}_m),\mathcal{C},\mathbf{b}_m\right). \end{aligned} \tag{17}$$

Similarly, the set of observations can be written as

$$\mathcal{Y}(t_l) \equiv \left\{\mathbf{W}\mathbf{Z}(t_l,\mathbf{b}_m) + \boldsymbol{\epsilon}_m \,|\, m \in \{1,\ldots,L\}\right\}. \tag{18}$$

Consequently, the conditional density over $\mathcal{Y}(t_l)$ splits up into the product

$$p\left(\mathcal{Y}(t_l)|\mathcal{Z}(t_l)\right) = \prod_{m=1}^{L} p\left(\mathbf{Y}_m(t_l)|\mathbf{Z}(t_l,\mathbf{b}_m)\right). \tag{19}$$

According to the model from Section II-A, we construct a bootstrap filter to first sample from the posterior $p(\mathcal{C},\mathcal{B},\alpha,\mathcal{Z}_{[t_1,t_N]}|\mathcal{Y}(t_1),\ldots,\mathcal{Y}(t_N))$ and then marginalize with respect to $\mathcal{B}$. Again, this requires sampling from the full parameter conditional distributions at each timepoint t_l, whereas we demand that we can find sufficient statistics which can be recursively updated as time increases. In case of the rate constants c_k, we obtain $p\left(c_k|\mathcal{Z}_{[t_1,t_l]},\mathcal{C}\backslash\{c_k\},\mathcal{B}\right) = p\left(c_k|\mathcal{Z}_{[t_1,t_l]},\mathcal{B}\right)$. As the individual system states $\mathbf{Z}(t,\mathbf{b}_m)$ are conditionally independent from each other, the likelihood function over all sample paths factorizes and as a consequence, the full conditional density over c_k is again a Gamma distribution

$$\begin{aligned} &p\left(c_k|\mathcal{Z}_{[t_1,t_l]},\mathcal{B}\right) = \\ &\qquad \Gamma\left(r_k^\star + \alpha_k, \sum_{m=1}^{L}\int_0^{t_l} g_k(\mathbf{Z}(s,\mathbf{b}_m))ds + \beta_k\right), \end{aligned} \tag{20}$$

where $r_k^\star$ is the total number of occurrences of reaction k over all L sample paths. At this point, we just state that the corresponding sufficient statistics can be recursively updated,

[2] Note that as $\mathbf{X}(t)$ is piecewise constant, the integral can be evaluated straight-forward using a finite sum.

even if they depend on $\mathbf{b}_m$, but a detailed discussion is skipped due to the space limitations.

In contrast, the conservation constants $\mathbf{b}_m$ enter the data likelihood in a complicated manner, such that the full conditional distributions with respect to $\mathbf{b}_m$, i.e., $p\left(\mathbf{b}_m|\mathcal{Z}_{[t_1,t_l]},\mathcal{C},\mathcal{B}\backslash\mathbf{b}_m,\alpha\right) = p\left(\mathbf{b}_m|\mathbf{Z}(\mathbf{b}_m)_{[t_1,t_l]},\mathcal{C},\alpha\right)$, are not of standard form. At this point, we could use a Metropolis-Hastings (M-H) update instead [16], to sample from $p\left(\mathbf{b}_m|\mathbf{Z}(\mathbf{b}_m)_{[t_1,t_l]},\mathcal{C},\alpha\right)$. This, however, requires evaluating the data likelihood function which would again lead to an increasing computational complexity over time. To our current knowledge, it is not possible in this case to find sufficient statistics, which can be updated over time, independently from all other relevant parameters (i.e. only $\mathcal{C}$, because α only shows up in the prior). We remember that the null space matrix $\mathbf{N}$ was defined to be a minimum positive integer base and thus, $\mathbf{b}_m$ can only take integer values as well. Thus, the simplest solution to that problem is to recursively update a one-dimensional discrete distribution with finite support, for each component of $\mathbf{b}_m$ separately. As before, we skip a detailed discussion on the construction of such a distribution.

Finally, we need to find a strategy to sample from $p\left(\alpha|\mathcal{Z}_{[t_1,t_l]},\mathcal{C},\mathcal{B}\right) = p\left(\alpha|\mathcal{B}\right)$. As we prefer allowing arbitrary distributions for $p(\mathcal{B}|\alpha)$ and $p(\alpha)$, we make us of a M-H acceptance criterion, to sample from $p\left(\alpha|\mathcal{B}\right)$, i.e., accept a proposed sample α^{new} with probability

$$a_{MH} = \min\left\{1, \frac{p(\mathcal{B}|\alpha^{new})p(\alpha^{new})q(\alpha^{old}|\alpha^{new})}{p(\mathcal{B}|\alpha^{old})p(\alpha^{old})q(\alpha^{new}|\alpha^{old})}\right\}, \quad (21)$$

where $q(\cdot|\cdot)$ denotes an arbitrary proposal distribution. The final algorithm is outlined in Algorithm 2.

Algorithm 2 Bootstrap filter for joint state and parameter estimation given cell population measurements.

```
for $t_l \in \{t_2, \ldots, t_N\}$ do
    Resample $P$ particles according to their weights $w^{(i)}$
    for each particle $i \in \{1, \ldots, P\}$ do
        {Resample rate constants}
        for each rate constant $c_k^{(i)} \in \mathcal{C}^{(i)}$ do
            $c_k^{(i)} \sim$ eq. (20)
        end for
        {Resample hyperparameters using M-H}
        $\alpha^{(i),new} \sim q\left(\alpha|\alpha^{(i),old}\right)$
        Accept $\alpha^{(i),new}$ with probability eq. (21)
        for each cell $m \in \{1, \ldots, L\}$ do
            {Resample conservation constants $\mathbf{b}_m^{(i)}$ as discussed}
            {Simulate CTMC between $t_{l-1}$ and $t_l$}
            $\mathbf{z}^{(i)}\left(t_l, \mathbf{b}_m^{(i)}\right) \sim$ eq. (10)
        end for
        {Compute importance weights}
        $w^{(i)} =$ eq. (19)
    end for
    Normalize weights such that $\sum_{i=1}^{P} w^{(i)} = 1$
end for
```

IV. A Case Study

In this section, the proposed algorithms from Section III are applied to a reaction system, modeling the Hog1 driven transcriptional process in yeast cells. The MAPK (Mitogen Activated Protein Kinase) Hog1 is the most downstream kinase of a signaling cascade which is activated by osmotic stress [21]. Upon activation of this pathway, a large fraction of the activated Hog1 is relocated from the cytoplasm to the nucleus of the cell to initiate a transcriptional program resulting in the up-regulation of roughly 300 genes [22]. This was recently shown to happen with different efficiency, leading to bimodal expression profiles [23]. The kinase activity of Hog1 in the cytoplasm leads to the production of glycerol, which allows the cells to equilibrate the interior and exterior osmotic pressures of the cell. Several levels of feedback adaptation are present. Experimental evidence from [24] suggest that activated Hog1 exerts a negative feedback loop on itself by activating phosphatases (e.g. Ptp3) targeting activated Hog1. Once adaptation has been achieved, Hog1 activity returns to basal levels and the active MAPK leaves the nucleus thereby offering only a short temporal window for the transcription of downstream target genes. Co-authors are measuring this transient relocation of Hog1 by fluorescent microscopy.

A. The Model

The model focuses on the activation of a single gene upon entry of the active MAPK Hog1 in the nucleus. The signaling part is thus reduced to a minimal model, being consistent with the high-basal level [24] and feedback adaptation [21]. The reaction network of the model is explained in Fig. 2.

We remark that the proposed model comprises events that are clearly non-elementary and thus, might in reality be subject to extrinsic variability. The reader should note that the proposed framework could as well account for a variability in kinetic parameters. For the sake of simplicity however, we decided not to cover such a scenario in this work.

The stochastic kinetic model comprises 14 species and 15 reactions, whereas we skip a detailed description due to the space limitations. Furthermore, to obtain a fair assessment of the proposed algorithm, we decided to use synthetic rather than experimental data. Thus, we apply Gillespie's direct simulation algorithm to our model in order to generate reference time course data for $L = 10$ cells. In this work we assume extrinsic variability to affect the chromatin remodeling process, which is reflected by a variability of RSC molecules. Assuming enzymatic reactions as indicated in Fig. 2, the corresponding conservation law for cell m is given as

$$b_{m,RSC} = RSC + Hog1^{P,N} : GPD1 : RSC,$$

where we assume that $b_{m,RSC}$ varies from cell to cell according to a log-normal distribution, i.e., $b_{m,RSC} \sim \mathcal{LN}\left(\ln 50, \sigma_b^2\right)$[3]. Here, σ_b denotes the hyperparemeter of interest which we estimate in order to capture the heterogeneity

[3] As $b_{m,RSC}$ denotes a discrete number of molecules, it has to be rounded after drawing it from a continuous distribution. The resulting artifacts are assumed to be negligibly small.

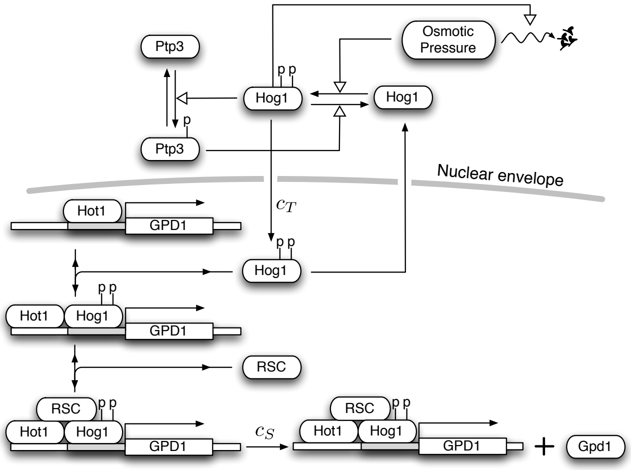

Fig. 2. Schematic model of MAPK Hog1 activation, shuttling and Hog1-induced synthesis of enzyme $Gpd1$. All reactions are modeled according to mass-action kinetics. Filled arrowheads denote elementary mass-action reactions, whereas non-filled arrowheads denote catalytic reactions, modeled as three mass-action events in case of $Ptp3$ activation, or as a single event otherwise. Once cytoplasmic Hog1 ($Hog1^C$) has been activated due to osmotic pressure ($Hog1^{P,C}$), it translocates to the nucleus ($Hog1^{P,N}$) and binds via transcription factors (such as Hot1) to the gene coding $GPD1$ of protein $Gpd1$. Efficient transcription of the gene product requires interaction of chromatin remodeling factors such as RSC with the $Hog1^{P,N}$: $GPD1$ complex to open the chromatin structure [23]. Once $Hog1^{P,N}$ is dephosphorylated via a nuclear phosphatase, it is again exported to the cytoplasm. In the model, those two steps are fused to a single reaction (i.e., a first-order conversion event). $Hog1^{P,C}$ exerts a negative feedback loop on itself by activating $Ptp3$ and additionally, equilibrates osmotic pressure (via production of glycerol).

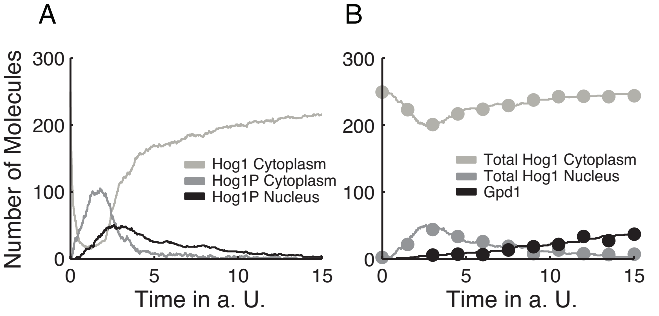

Fig. 3. Sample traces generated by the reference model. Figure 3A shows the temporal evolution of active Hog1 in the cytoplasm as well as in the nucleus, where it is dephosphorylated to Hog1 and again exported. The quantities available from fluorescent microscopy are shown in 3B, i.e., total Hog1 in the cytoplasm and in the nucleus as well as $Gpd1$. The observations (shown as circles) were acquired with equidistant time intervals $\Delta t = 1.5$ and corrupted by uncorrelated Gaussian noise with $\sigma = 5$. Note that the x-axes have arbitrary time scale.

over the cell population. During generation of the reference data we set $\sigma_b = 0.3$. For simplicity we assume that all of the 15 rate constants of the stochastic model are known except c_T the transportation rate of $Hog1^{P,C}$ to the nucleus (i.e., conversion to $Hog1^{P,N}$) and c_S, the rate at which new proteins $Gpd1$ are synthesized.

As in real-world fluorescent microscopy, we assume that we can measure only total Hog1 in the cytoplasm and the nucleus[4] as well as $Gpd1$. The obtained paths are sampled at time points $t_l = (l-1)\Delta t$, $\forall l \in \{1,\ldots,N\}$ with $\Delta t = 1.5$ up to a maximum time of $t = 40$. Additionally, we assume additive uncorrelated Gaussian observation noise, i.e., $\boldsymbol{\epsilon}_m \sim \mathcal{N}\left(\mathbf{0}, 5^2\mathbf{I}\right)$, $\forall i$ and $\forall m \in \{1,\ldots,L\}$. Fig. 3 shows exemplary cytoplasmic and nuclear Hog1 responses (Figure 3A) as well as the corresponding observations, later used for model calibration (Figure 3B).

B. Model Calibration

Using the algorithms described in Section III and the data generated according to the previous section, we can now perform model calibration to estimate parameters c_T and c_S as well as hyperparameter σ_b. In the following we provide a detailed description of the underlying algorithm configuration.

For sampling the rate constants according to (20) we assume that c_T and c_S are Gamma distributed a-priori, i.e., $p(c_T) = \Gamma(1,1)$ and $p(c_S) = \Gamma(4,1)$. As described in in Section III-B, we compute the discrete conditional distribution over $b_{m,RSC}$ for integer values between 1 and 3000 and using the log-normal prior distribution $p(b_{m,RSC}) = \mathcal{LN}\left(\ln 50, \sigma_b^2\right)$. Furthermore, we sample from $p(\alpha|\mathcal{B}) = p(\sigma_b|b_1,\ldots,b_L)$ by using a M-H update with proposal density $q(\sigma_b^{new}|\sigma_b^{old}) = \mathcal{N}\left(\sigma_b^{old}, 0.1^2\right)$. For σ_b we do not incorporate prior knowledge.

For cell population based model calibration, we apply Algorithm 2 to the generated measurement data using $P = 500$ particles, where initial values of the parameters were drawn according to their prior distributions, if available. After the burn-in of the bootstrap filter, one can accumulate the obtained samples over time to improve the subsequent posterior estimation. Even if the described algorithm returns estimates of the joint parameter posteriors, for clarity we rather present the marginal posterior distributions of each parameter separately as shown in Figure 4.

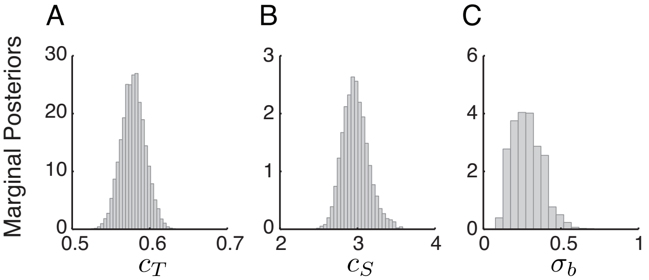

Fig. 4. Results for estimating the marginal parameter posterior distributions. Maximum a-posteriori (MAP) estimates were computed from the particles as $\bar{c}_T = 0.5830$ $\bar{c}_S = 2.9396$ and $\bar{\sigma}_b = 0.2492$. Figure 4A and 4B show the estimated histograms for the rate constants c_T and c_S. Results show small deviations to the reference values $c_T = 0.5$ and $c_S = 3$ but correspond well to the MAP estimates $\hat{c}_T = 0.5331$ and $\hat{c}_S = 2.9609$, obtained from noise-free complete data. The estimated posterior density over the hyperparameter σ_b can be seen in Figure 4C . Even if we have used only $L = 10$ cells, the results fit well the true value of $\sigma_b = 0.3$. We have also computed the empirical standard deviation $\hat{\sigma}_b$ over the (true) values of $b_{m,RSC}, m \in \{1,\ldots,L\}$, which was found as $\hat{\sigma}_b = 0.2557$.

Clearly, it would be of great interest to obtain estimates

[4] This means that we also consider Hog1 molecules that are bound to a complex.

for the species that have not been directly measured. For demonstration, we estimated the amounts of $Hog1^{P,C}$ and $Hog1^{C}$. The corresponding results are shown in Figure 5.

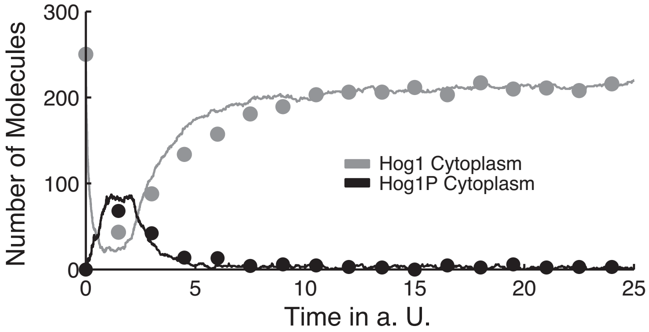

Fig. 5. Reconstruction of missing species of a single cell. Circles denote the MAP state at the measurement time points, calculated over the particle distribution. As expected, estimation accuracy increases with time.

V. Conclusions

We presented an efficient recursive Bayesian estimation procedure, which allows model calibration on single cell - as well as heterogeneous cell population data under realistic conditions. For the latter case, we have used the hierarchical Bayesian model from [11], which allowed us to additionally estimate hyperparameters, representing low-dimensional statistics of the extrinsic variability. As the formulated prediction-correction procedure is analytically intractable, we have combined a bootstrap filter with a Metropolis-within-Gibbs iteration, to jointly carry out the sequential state and parameter estimation. Our algorithms were applied to synthetic data, generated from a model of the osmo-stress regulation in budding yeast (MAPK Hog1). We have shown that state and parameter estimation works comparably well, even if we can measure only total nuclear and cytoplasmic Hog1 and Gpd1 at discrete time points. The fact that this configuration corresponds to a realistic scenario suggests that the proposed estimation scheme can deal well with real-world experimental data.

A. Future Work

The main draw-back of the algorithm is it's initialization, especially if little prior knowledge is available. In fact, convergence can be very slow for improper starting values, or the particle filter might even produce degenerate solutions. Thus, an efficient algorithm for finding a suitable initialization would be of great interest. Furthermore, it seems natural to use *smoothing* variants of the proposed particle filter [19], [25] when dealing with off-line data. This would allow to incorporate knowledge also from future time points to improve the estimation accuracy.

References

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.

[2] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, G. Pesce, and R. Brent, "Regulated cell-to-cell variation in a cell-fate decision system," *Nature*, vol. 437, no. 29, pp. 699–502, 2005.

[3] V. Shahrezaei, J. F. Ollivier, and P. S. Swain, "Colored extrinsic fluctuations and stochastic gene expression," *Molecular Systems Biology*, vol. 4, May 2008.

[4] M. Komorowski, B. Finkenstädt, and D. Rand, "Using a Single Fluorescent Reporter Gene to Infer Half-Life of Extrinsic Noise and Other Parameters of Gene Expression," *Biophysical Journal*, vol. 98, no. 12, pp. 2759–2769, June 2010.

[5] B. Snijder and L. Pelkmans, "Origins of regulated cell-to-cell variability," *Nature Reviews Molecular Cell Biology*, vol. 12, no. 2, pp. 119–125, Jan. 2011.

[6] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood, "Bayesian inference for a discretely observed stochastic kinetic model," *Stat. Comput.*, vol. 18, no. 2, pp. 125–135, 2008.

[7] A. Ruttor, G. Sanguinetti, and M. Opper, "Approximate inference for stochastic reaction processes," in *Learning and Inference in Computational Systems Biology*, pp. 189–205. The MIT Press, 2009.

[8] A. Golightly and D. J. Wilkinson, "Bayesian inference for nonlinear multivariate diffusion models observed with error," *Comput. Stat. Data An.*, vol. 83, pp. 1891–1901, 2008.

[9] D. Thorsley and E. Klavins, "Model reduction of stochastic processes using Wasserstein pseudometrics," in *Amer Contr Conf*, June 11–13 2008, pp. 1374–1381.

[10] J. Hasenauer, S. Waldherr, N. Radde, M. Doszczak, and F. Allgoewer, "A maximum likelihood estimator for parameter distributions in heterogeneous cell populations," in *Procedia Computer Science*, 2010, vol. 1, pp. 1655–1663.

[11] H. Koeppl, C. Zechner, A. Ganguly, S. Pelet, and M. Peter, "Accounting for Extrinsic Variability in the Estimation of Stochastic Rate Constants," Submitted to Journal of Robust and Nonlinear Control, accepted with minor revisions.

[12] D. F. Anderson and T. G. Kurtz, "Continuous time Markov chain models for chemical reaction networks," *in press*, 2011.

[13] D. J. Wilkinson, *Stochastic Modelling for Systems Biology (Chapman & Hall/CRC Mathematical & Computational Biology)*, Chapman and Hall/CRC, 1 edition, April 2006.

[14] C. Kreutz, M. M. Bartolome Rodriguez, T. Maiwald, M. Seidl, H. E. Blum, L. Mohr, and J. Timmer, "An error model for protein quantification.," *Bioinformatics (Oxford, England)*, vol. 23, no. 20, pp. 2747–53, Oct. 2007.

[15] M. Amrein and H. Künsch, "Rate estimation in partially observed markov jump processes with measurement errors," *Statistics and Computing*, pp. 1–14, 10.1007/s11222-011-9244-1.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 1st ed. 2006. corr. 2nd printing edition, October 2007.

[17] D. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of computational physics*, vol. 22, no. 4, pp. 403–434, 1976.

[18] P. Fearnhead, "Markov chain Monte Carlo, sufficient statistics, and particle filters," *Journal of Computational and Graphical Statistics*, vol. 11, pp. 848–862, 2002.

[19] O. Cappe, S. J. Godsill, and E. Moulines, "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, May 2007.

[20] Z. Khan, T. Balch, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 11, pp. 1805–19, Nov. 2005.

[21] S. Hohmann, M. Krantz, and B. Nordlander, "Yeast osmoregulation," *Meth Enzymol*, vol. 428, pp. 29–45, Jan 2007.

[22] A. P. Capaldi, T. Kaplan, Y. Liu, N. Habib, A. Regev, N. Friedman, and E. K. O'Shea, "Structure and function of a transcriptional network activated by the MAPK Hog1," *Nat Genet*, vol. 40, no. 11, pp. 1300–6, Nov 2008.

[23] S. Pelet, F. Rudolf, M. Nadal-Ribelles, E. de Nadal, F. Posas, and M. Peter, "Transient activation of the HOG MAPK pathway regulates bimodal gene expression.," *Science (New York, N.Y.)*, vol. 332, no. 6030, pp. 732–5, May 2011.

[24] J. Macia, S. Regot, T. Peeters, N. Conde, R. Sole, and F. Posas, "Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction.," *Science signaling*, vol. 2, no. 63, pp. ra13, Jan. 2009.

[25] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky, Eds. Oxford University Press, 2009.