

---

# Approximate parameter inference in a stochastic reaction-diffusion model

---

**Andreas Ruttor**  
TU Berlin, Germany

**Manfred Opper**  
TU Berlin, Germany

## Abstract

We present an approximate inference approach to parameter estimation in a spatio-temporal stochastic process of the reaction-diffusion type. The continuous space limit of an inference method for Markov jump processes leads to an approximation which is related to a spatial Gaussian process. An efficient solution in feature space using a Fourier basis is applied to inference on simulational data.

## 1 Introduction

Reaction-diffusion models are used to describe spatio-temporal stochastic processes where molecules diffuse through space, are created and destroyed, and can participate in chemical reactions when they are close. This type of process finds a variety of applications in chemistry, physics (Gardiner, 1996), and also in the field of computational biology. Recently the propagation of the *Bicoid* protein, which is relevant for morphogenesis in the *Drosophila* embryo, has been modelled by such a process (Wu et al., 2007).

In many applications, one can assume a qualitative knowledge of basic processes in the model. However, the numerical values of the parameters corresponding to these processes, such as the microscopic rates for diffusion, decay, and reactions, are often unknown. Hence, it is important to be able to infer such quantities from concrete experimental data.

This is a problem of non-parametric statistical inference, because the natural description of the state of such systems is not in terms of the dynamics of individual molecules, but of the *molecule density as a function of continuous space and time*. Experiments typically provide (noisy) measurements of the (smoothed) molecule densities over a range of spatial locations at discrete times.

---

Appearing in Proceedings of the 13<sup>th</sup> International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9. Copyright 2010 by the authors.

Statistical inference for such processes is highly nontrivial because

- observations are not independent but are generated from a **Markov process with continuous time**.
- **Space is continuous**. It is not clear if an inference method for a spatially discretized model remains computationally tractable when discretization becomes finer and finer.
- **Fluctuations may not be irrelevant**: The transition to a spatial continuum limit is often enabled in the limit where the number of molecules is large enough to safely neglect density fluctuations. The evolution of the macroscopic density is then typically governed by a partial differential equation (PDE). In some cases it is possible to solve the PDE for the stationary (time independent) density analytically which could be used as a fit to experiments. However, neglecting the temporal dynamics usually does not allow us to estimate all rate parameters, but only certain ratios. Furthermore, measurements are often indirect and allow to estimate molecule densities only up to an unknown scale factor. Then even a time dependent solution of the macroscopic PDEs would not be sufficient to infer those parameters which directly determine molecule numbers (such as creation rates) and the stochasticity of the spatio-temporal process must be discussed. For the relevance of fluctuations in the biological context, see (Tostevin et al., 2007).

The problem of inference in continuous time Markov processes has recently attracted attention in the machine learning/AI community and a variety of approximate inference methods for these models have been developed (Archambeau et al., 2008; Cohn et al., 2009; Nodelman et al., 2005; Opper and Sanguinetti, 2008). However, it is not clear if these methods will scale up to be able to deal with the complexity of fluctuating spatial densities.

In this paper we present an approach for approximate inference in a reaction-diffusion model which takes both Markov dynamics and spatial fluctuations into account. The method is based on a recently developed approximate inference method for Markov jump processes (Ruttor and

Opper, 2009) which easily allows for the transition to a continuous space.

We apply the method to a model which contains the basic processes relevant for the *Bicoid* protein evolution in *Drosophila* (i.e. creation and decay of molecules, but no chemical reactions between molecules), but uses a simplified *one dimensional* geometry. We have used this model as a proof of concept for our inference approach. The data sets were obtained from simulations of the model and the measurement process using realistic true system parameters (Wu et al., 2007). This allows us to see how well these parameters can be inferred from data.

We have not yet applied the model to *real measurements*, because available data at different time points are taken from *different embryos* and are thus not based on the statistical assumptions of our model. Obtaining the more interesting *in vivo* measurements on a single embryo is non-trivial, but should hopefully be possible in the near future (Wu et al., 2007).

## 2 Reaction-diffusion models as Markov jump processes

A sound description of the stochastics of reaction-diffusion models can be based on a *compartment* model (Gardiner, 1996), where space is discretized into small cells of size  $\Delta x$  (assuming a single space dimension for simplicity). The state of the system is described by a vector  $\mathbf{n} = (n_0, \dots, n_M)$ , where  $n_i$  is the number of molecules in a cell. The stochastic dynamics is assumed to be a Markov jump process (MJP) defined by a rate function  $f(\mathbf{n}'|\mathbf{n})$  which determines the temporal change of transition probabilities via

$$P(\mathbf{n}', t + \Delta t | \mathbf{n}, t) \simeq \delta_{\mathbf{n}', \mathbf{n}} + \Delta t f(\mathbf{n}' | \mathbf{n}) \quad (1)$$

for  $\Delta t \rightarrow 0$ . The rate  $f(\mathbf{n}'|\mathbf{n})$  is the sum of the rates for all individual processes which lead from  $\mathbf{n}$  to  $\mathbf{n}'$ .

In the *Bicoid* system we consider three types of elementary processes:

- **Particle diffusion**, say from cell  $i$  into the neighboring cell  $i + 1$ , is modelled by a transition  $n'_i = n_i - 1$  and  $n'_{i+1} = n_{i+1} + 1$  with a rate  $d n_i / \Delta x^2$ , where  $d$  is the diffusion constant of the system. The proportionality of the rate to the number  $n_i$  of molecules accounts for the fact that each of the molecules can perform the jump to the neighboring cell. As we will show later, the factor  $1/\Delta x^2$  will allow for a proper limit  $\Delta x \rightarrow 0$  of the compartment model.
- In a similar way we can treat **Molecule degradation**  $n'_i = n_i - 1$  which occurs with a rate  $\lambda n_i$  in all cells.
- **Creation of molecules** in the system occurs by injection with a fixed rate  $c/\Delta x$  in a single cell only.

## 3 Approximate Inference for MJPs

While it is not possible to perform exact inference on this model, we will resort to an approximate inference approach for Markov jump processes following the work of Rutter and Opper (2009). The main contribution of the present paper is to extend this method to the *non-parametric continuum* limit required for reaction-diffusion models. Here the cell size  $\Delta x$  shrinks to zero and molecule numbers  $n_i$  are replaced by densities  $\rho(x)$  via

$$n_i \rightarrow \rho(x_i) \Delta x. \quad (2)$$

We will next briefly review the main results of Rutter and Opper (2009) needed for our application.

Parameter inference of Markov jump processes can be based on the recursive computation of

$$\psi(\mathbf{n}, t) \equiv P(D_{\geq t} | \theta, \mathbf{n}(t) = \mathbf{n}), \quad (3)$$

the likelihood of *future observations*  $D_{\geq t} = \{\mathbf{y}_i\}_{t_i \geq t}$  conditioned on the state  $\mathbf{n}(t) = \mathbf{n}$  at the present time  $t$ . The likelihood of all data is used for maximum likelihood or Bayesian estimation of rate constants  $c$ ,  $d$ , and  $\lambda$  is consequently

$$p(D|c, d, \lambda) = \sum_{\mathbf{n}} \psi(\mathbf{n}, 0) p_0(\mathbf{n}), \quad (4)$$

where  $p_0(\mathbf{n})$  is the distribution of the initial state.

For times *between two observations*  $\psi(\mathbf{n}, t)$  obeys the *Kolmogorov backward* equation

$$\frac{d}{dt} \psi(\mathbf{n}, t) = \sum_{\mathbf{n}' \neq \mathbf{n}} f(\mathbf{n}' | \mathbf{n}) [\psi(\mathbf{n}', t) - \psi(\mathbf{n}, t)], \quad (5)$$

which has to be solved backward in time with the end condition  $\psi(\mathbf{n}, t_N) = p(\mathbf{y}_N | \mathbf{n}(t_N))$ . The observations enter consecutively from the latest to the first one through their conditional distributions (assuming independent noise)  $p(\mathbf{y} | \mathbf{n})$ , in the jump conditions

$$\lim_{t \rightarrow t_l^-} \psi(\mathbf{n}, t) = p(\mathbf{y}_l | \mathbf{n}(t_l)) \lim_{t \rightarrow t_l^+} \psi(\mathbf{x}, t), \quad (6)$$

where  $t_l^-$  and  $t_l^+$  denote the left and right side limits. Similar to the backward-forward algorithm in Hidden Markov Models state inference at arbitrary times  $t$  can be performed using  $\psi(\mathbf{n}, t)$  together with a solution of the (forward) master equation.

The approximation method of Rutter and Opper (2009) is an extension of van Kampen's system size expansion approach (Gardiner, 1996; van Kampen, 1981) originally developed for solving the forward Master equation to the problem of statistical inference. Assuming systems with sufficiently large numbers of molecules, the state  $\mathbf{n}$  of the system is essentially approximated by a vector with continuous components which have Gaussian fluctuations around a *macroscopic* state.

Formally, the system size expansion can be derived by a combination of a diffusion approximation to the Master equation together with a subsequent Gaussian ‘weak noise’ approximation, where small relative fluctuations are assumed.<sup>1</sup>

Applying the same ideas to the *Kolmogorov backward equation* Ruttur and Opper (2009) derive an approximation to (5) by the partial differential equation

$$\frac{\partial \psi(\mathbf{n}, t)}{\partial t} + (\mathbf{n} - \mathbf{b})^\top \mathbf{A}^\top(\mathbf{b}(t)) \nabla \psi(\mathbf{n}, t) + \frac{1}{2} \text{Tr}(\mathbf{D}(\mathbf{b}(t)) \nabla \nabla^\top) \psi(\mathbf{n}, t) = 0. \quad (7)$$

Here  $\mathbf{b}$  denotes the *macroscopic* state of the system for which fluctuations are ignored and  $\nabla$  is the vector of partial derivatives with respect to  $\mathbf{n}$ . This fulfils the classical rate equation

$$\frac{d\mathbf{b}}{dt} = \mathbf{f}(\mathbf{b}), \quad (8)$$

where

$$\mathbf{f}(\mathbf{n}) = \sum_{\mathbf{n}' \neq \mathbf{n}} f(\mathbf{n}' | \mathbf{n}) (\mathbf{n}' - \mathbf{n}) \quad (9)$$

is the first jump moment or *drift vector* of the transition rates. The matrix  $\mathbf{A} = \nabla \mathbf{f}^\top$  is the Jacobian of the first jump moment. The *diffusion matrix* (again not to be confused with the diffusion constant  $d$  for the molecules)

$$\mathbf{D}(\mathbf{n}) = \sum_{\mathbf{n}' \neq \mathbf{n}} (\mathbf{n}' - \mathbf{n}) f(\mathbf{n}' | \mathbf{n}) (\mathbf{n}' - \mathbf{n})^\top \quad (10)$$

is defined by the second jump moments. The dependency of  $D(\mathbf{b})$  on  $\mathbf{b}$  accounts for the fact that fluctuations depend on the number of molecules.

Eq. (7) is the backward equation for a Gaussian diffusion process of the *Ornstein-Uhlenbeck* type (Gardiner, 1996). For observations with Gaussian noise, it can be shown that the solution  $\psi(\mathbf{n}, t)$  of (7) is of the form

$$\psi \approx \frac{z(t)}{|\mathbf{S}|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{n} - \mathbf{b}(t))^\top \mathbf{S}^{-1}(t) (\mathbf{n} - \mathbf{b}(t)) \right]. \quad (11)$$

Although  $\psi(\mathbf{n}, t)$  (as a solution to a backward equation) is not a normalized probability, we will refer to (11) as a *Gaussian*. This analogy is helpful when we later apply linear transformations such as Fourier transforms to the variables.

Between observations the dynamics of  $\mathbf{b}(t)$  is given by (8) and the matrix  $\mathbf{S}(t)$  and normalizer  $z(t)$  evolve according to the ODEs

$$\frac{d\mathbf{S}}{dt} = \mathbf{A}\mathbf{S} + \mathbf{S}\mathbf{A}^\top - \mathbf{D}(\mathbf{b}), \quad (12)$$

$$\frac{dz}{dt} = z(t) \text{Tr}(\mathbf{A}). \quad (13)$$

<sup>1</sup>Note, that the term diffusion here refers to the stochastic dynamics of the vector  $\mathbf{n}$  and should not be confused with the diffusion of the molecules.

The end condition for  $\mathbf{S}$  is given by  $\mathbf{S}(t_N) = \sigma^2 \mathbf{I}$  in the case of independent Gaussian measurement noise with standard deviation  $\sigma$ . Here  $\mathbf{I}$  denotes the unit matrix.

Using the rates of the MJPs for the compartment model, assuming that molecules are created at site  $i = 0$  only, one can derive the following expressions from (9) and (10) for drift vector and diffusion matrix:

$$f_i(\mathbf{n}) = \frac{d}{\Delta x^2} (n_{i+1} + n_{i-1} - 2n_i) - \lambda n_i + c \delta_{i,0} \quad (14)$$

$$D_{i,j}(\mathbf{n}) = \frac{d}{\Delta x^2} (n_i + n_{i+1}) (\delta_{i,j} - \delta_{i+1,j}) + \frac{d}{\Delta x^2} (n_i + n_{i-1}) (\delta_{i,j} - \delta_{i-1,j}) + \delta_{i,j} \lambda n_i + \delta_{i,0} \delta_{j,0} c. \quad (15)$$

Here  $\delta_{i,j}$  denotes the *Kronecker* symbol which equals 1 for  $i = j$  and 0 else.

In spatially homogeneous reaction systems both drift vector  $\mathbf{f}$  and diffusion matrix  $\mathbf{D}$  only depend on the state of the system and the reaction constants. But here, in the case of a compartment model for a reaction-diffusion system, there is an additional parameter to choose: the cell size  $\Delta x$ . Of course, all information contained in the observations should be used by the inference algorithm. For that purpose  $\Delta x$  has to be smaller than or at least equal to the spatial distance between two adjacent data points  $y_{i,j}$ . However, shrinking the size of the compartments increases their number and the dimension of the matrices used in the calculation of the likelihood. Consequently, fulfilling this condition on  $\Delta x$  is often not possible because of limited computational resources.

## 4 Continuum limit

A better way to solve this problem is calculating drift vector  $\mathbf{f}$  and diffusion matrix  $\mathbf{D}$  in the limit  $\Delta x \rightarrow 0$  analytically. By doing so, we obtain a representation of the *Bicoid* model which decouples the effective dimension of the system state used in numerical calculations from the number of spatial components found in each observation  $\mathbf{y}_i$ .

To perform the continuum limit we introduce the spatial positions  $x_i = i\Delta x$  and particle densities via  $\rho(x_i) = n_i \Delta x$ . Denoting the macroscopic density corresponding to  $\mathbf{n}$  by  $\bar{\rho}$  it is straightforward to perform this limit for the macroscopic rate equation (8). Using a Taylor series expansion to 2nd order around  $x_i$  one obtains the well known classical diffusion equation

$$\frac{\partial}{\partial t} \bar{\rho}(x, t) = d \cdot \frac{\partial^2}{\partial x^2} \bar{\rho}(x, t) - \lambda \bar{\rho}(x, t) + c \delta(x) \quad (16)$$

for molecules, which include the decay and a source term at  $x = 0$ , where  $\delta(x)$  is Dirac’s  $\delta$  distribution.

To include *density fluctuations* in the continuum limit a useful strategy would seem to replace the multivariate Gaussian (11) for  $n_i(t)$  by a corresponding spatial *Gaussian process* (Rasmussen and Williams, 2006) for the density  $\rho(x, t)$ . The matrix  $\mathbf{S}$  would then become a type of covariance operator. An explicit expression in terms of differential operators for this operator is indeed possible. Surprisingly, it is also possible to solve the operator differential equation corresponding to (12) analytically for our model. However, the subsequent computations of explicit operator inverses which are needed at the observations together with the normalizing determinant turn out to be analytically intractable in *position space*. Hence, we have not pursued this route any further and give details elsewhere.

It turns out that an equivalent approach to defining a Gaussian process in *position space* is to do that in a *feature space* (Rasmussen and Williams, 2006) where the fluctuating densities are expanded as an infinite linear combination of functions. The natural choice in our case is a *Fourier basis* for which a truncation with a relatively small number of features usually gives good results.

The particle density of the molecules is expanded into a Fourier series,

$$\rho(x, t) = \tilde{\rho}_0(t) + 2 \sum_{k=1}^{\infty} \tilde{\rho}_k(t) \cos(k\pi x/L), \quad (17)$$

as well as the macroscopic density corresponding to  $\mathbf{n}$ ,

$$\bar{\rho}(x, t) = \hat{\rho}_0(t) + 2 \sum_{k=1}^{\infty} \hat{\rho}_k(t) \cos(k\pi x/L). \quad (18)$$

Assuming reflecting boundary conditions at  $x=0$  and  $x=L$  the series can only contain cosine waves with frequencies  $k\pi/L$ . In order to simplify further calculations we continue our model periodically in space with period  $2L$  and symmetrically to  $x=0$ , so that  $\rho(-x, t) = \rho(x, t)$ . This is already a property of the Fourier series (17) and (18).

Applying the Fourier transform to the classical diffusion equation (16) and taking the boundary conditions into account leads to

$$\tilde{f}_k = - \left[ d \left( \frac{k\pi}{L} \right)^2 + \lambda \right] \tilde{\rho}_k + \frac{c}{L}, \quad (19)$$

which describes the drift  $\mathbf{f}$  of the continuous model in Fourier space. It is possible to do a similar transformation for the diffusion operator  $\mathbf{D}$ . But, in this case, a more elegant approach is to apply the discrete Fourier transform defined by

$$\tilde{\rho}_k = \frac{1}{L} \sum_{j=0}^M n_j \cos \left( \frac{k\pi}{L} j \Delta x \right) \quad (20)$$

to the diffusion matrix of the discrete compartment model,

which results in

$$\tilde{D}_{r,s} = \frac{1}{L^2} \sum_{i=0}^M \sum_{j=0}^M D_{i,j} \cos \left( \frac{r\pi}{L} i \Delta x \right) \cos \left( \frac{s\pi}{L} j \Delta x \right). \quad (21)$$

Afterwards we take the limit  $\Delta x \rightarrow 0$ ,  $M \rightarrow \infty$  and arrive at

$$\tilde{D}_{r,s} = \frac{1}{2L} \left[ 2d \left( \frac{r\pi}{L} \right) \left( \frac{s\pi}{L} \right) + \lambda \right] \tilde{\rho}_{|r-s|} + \frac{c}{2L^2} \quad (22)$$

for the diffusion  $\mathbf{D}$  in the continuous model. In contrast to the corresponding equations (14) and (15) for the discrete compartment model, (19) and (22) do not contain the cell size  $\Delta x$ .

Then the solution  $\psi(\tilde{\rho}, t)$  of the backward equation in the Fourier representation is found to be

$$\psi \approx \frac{z(t)}{|\tilde{\mathbf{S}}|^{1/2}} \exp \left[ -\frac{1}{2} (\tilde{\rho} - \hat{\rho}(t))^T \tilde{\mathbf{S}}^{-1}(t) (\tilde{\rho} - \hat{\rho}(t)) \right]. \quad (23)$$

The Fourier representation of the densities has a remarkable property: Each component  $f_k$  only depends linearly on the series coefficient  $\tilde{\rho}_k(t)$  for the same spatial frequency. Consequently, operator  $\mathbf{A}$  is diagonal in Fourier space and the system of coupled matrix differential equations equations (8) and (12) for the compartment model becomes a set of uncoupled linear differential equations after applying Fourier transform and continuum limit:

$$\begin{aligned} \frac{d}{dt} \hat{\rho}_k &= - \left[ d \left( \frac{k\pi}{L} \right)^2 + \lambda \right] \hat{\rho}_k + \frac{c}{L}, \quad (24) \\ \frac{d}{dt} \tilde{S}_{r,s} &= - \left[ d \left( \frac{r\pi}{L} \right)^2 + d \left( \frac{s\pi}{L} \right)^2 + 2\lambda \right] \tilde{S}_{r,s} \\ &\quad - \frac{1}{2L} \left[ 2d \left( \frac{r\pi}{L} \right) \left( \frac{s\pi}{L} \right) + \lambda \right] \hat{\rho}_{|r-s|} \\ &\quad - \frac{c}{2L^2}. \quad (25) \end{aligned}$$

All equations can be solved analytically in terms of exponential functions. Additionally, covariance components  $\tilde{S}_{r,s}$  are only influenced by mean values  $\hat{\rho}_k$  with  $k \leq r$  and  $k \leq s$ . Therefore it is possible to truncate all Fourier series at order  $N$  without affecting these calculations for lower frequencies at all. The only exception is the normalization factor  $z$ , because the likelihood of a *Gaussian process* depends on which features are actually observed:

$$\frac{d}{dt} \ln z = - \sum_{k=0}^N \left[ d \left( \frac{k\pi}{L} \right)^2 + \lambda \right]. \quad (26)$$

However, as the right-hand side is independent of the system state, this dependency on  $N$  does not influence parameter estimation.

In summary,  $\psi(\tilde{\rho}, 0)$  is calculated by integrating the ODEs (24), (25), (26) backwards in time and applying the jump

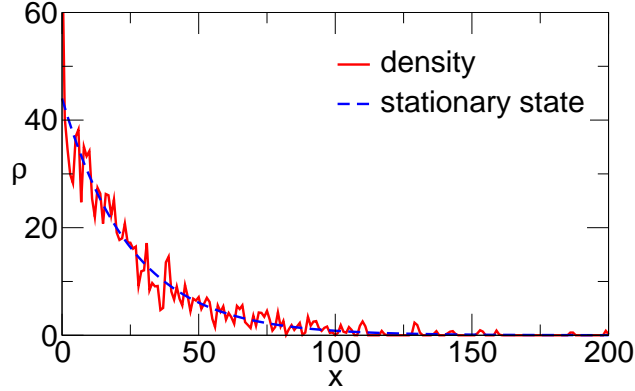


Figure 1: Typical sample of the system after reaching the stationary state for parameters  $c = 30$ ,  $d = 17.2$ , and  $\lambda = 0.027$ . The solid line shows the actual density of molecules, while the stationary solution of (16), the expectation value of the prior process, is given by the dashed line.

condition (6) at each observation. Then the total likelihood is given by

$$p(D|c, d, \lambda) = \int \psi(\tilde{\rho}, 0) p(\tilde{\rho}_0) d\tilde{\rho}, \quad (27)$$

where  $p(\tilde{\rho}_0)$  denotes the prior for the initial conditions. For the results shown here, we have used an uninformative prior, which is flat at observed lower frequencies, while it suppresses fluctuations at unobserved higher frequencies. In this special case we find  $p(D|c, d, \lambda) \propto z$ .

## 5 Observations

We also have to consider how observations are generated from the continuous density  $\rho(x, t)$ . We assume that this measurement process for position  $x_r$  and time  $t_s$  can be described by a convolution of  $\rho(x, t)$  with a weight function  $w(x)$  in position space,

$$y_{r,s} = \int_{-L}^L w(x_r - x) \rho(x, t_s) dx + \xi_{r,s} \quad (28)$$

where  $\xi_{r,s} \sim \mathcal{N}(0, \sigma^2)$  denotes additive Gaussian white noise. This convolution corresponds to a product in Fourier space, so that each observed feature is just given as

$$\tilde{y}_{k,s} = \tilde{w}_k \tilde{\rho}_k(t_s) + \tilde{\xi}_{k,s}. \quad (29)$$

We obtain  $M$  data points at positions  $x_r = rL/M$  and use a triangle window of width  $l = 2L/M$ , so that the weights  $\tilde{w}_k$  in Fourier space are given by

$$\tilde{w}_k = \left( \frac{2\pi}{L} kl \right)^{-2} \sin^2 \left( \frac{2\pi}{L} kl \right). \quad (30)$$

Here high frequencies are just averaged out in the measurement process. Therefore all information contained in

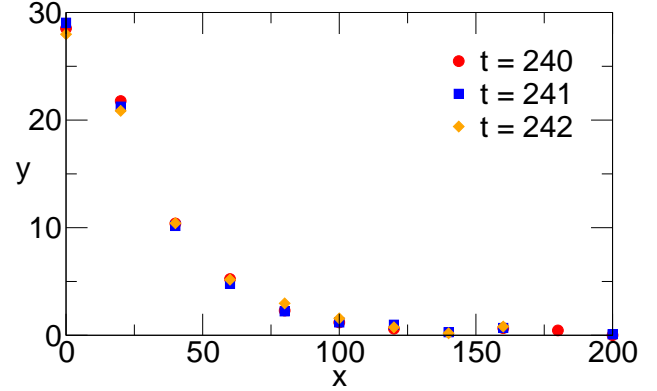


Figure 2: The first three observations of a data set as a function of position  $x$ . The data points were obtained from a simulation with parameters  $c = 30$ ,  $d = 17.2$ ,  $\lambda = 0.027$ , and noise level  $\sigma = 0.4$  by convoluting the particle density with a triangle window of width  $l = 40$ .

the observations can be found in the low-order terms of the series expansion, while higher-order terms mostly contain noise and can be omitted in practise by truncating the Fourier series at order  $N$ . Note, that  $N$  can be smaller than  $M$ , which allows for a reduction of the computational complexity. We found that parameter estimation works well even if  $N$  is as small as  $M/2$ , which has been used to obtain the results shown in this paper. In contrast, the number of cells in the compartment model has to be greater than or equal to the number of spatial data points.

Our inference method can use other weight functions, too. This is especially useful if the characteristics of the real measurement process are known. As long as the measurement window is large compared to the spatial distance between adjacent observations, our algorithm works without modifications. But if a small window does not suppress high frequencies sufficiently, these are folded back into the low frequency components of  $\tilde{y}_{k,s}$  by the discrete Fourier transform. In this case one has to replace  $\psi(\tilde{\rho}, t)$  with the corresponding likelihood  $\psi(\tilde{w}\tilde{\rho}, t)$  for a weighted density.

## 6 Simulations

Although real data from biochemical experiments is not yet available, we can generate observations using simulations of the model. For that purpose, we could use the Gillespie algorithm (Gillespie, 1992) applied to the compartment model, which is quite a standard approach for Markov jump processes in a discrete state space. But choosing a well suited compartment size  $\Delta x$  is difficult. If it is large, we only have to deal with a small number of state variables. Transfer reactions between the compartments are rare events, as the diffusion rate is inverse proportional to  $\Delta x^2$ . In this case the Gillespie algorithm works well, but the spatial resolution of the simulation is rather low. Of

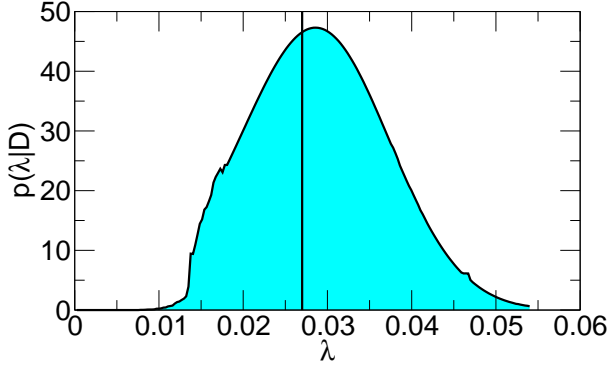


Figure 3: Marginal posterior averaged over 200 data sets for the decay constant. The vertical line shows the true parameter value  $\lambda = 0.027$ .

course, improving it is possible by just using a large number of compartments. But a small compartment size  $\Delta x$  also leads to a very high frequency of transfer reactions. Therefore the average waiting time between reaction events is small and the Gillespie algorithm, which processes single reaction events, becomes very slow.

Simulating the molecular dynamics of a reaction-diffusion model directly instead of using the compartment approximation can be an alternative. By doing so, we avoid the discretization error caused by the finite compartment size  $\Delta x$ . As the Bicoid model only contains first-order reactions, diffusion processes of single molecules do not influence each other and can be solved analytically. Therefore the new position  $x(t + \Delta t)$  of a molecule starting at  $x(t)$  after a time span  $\Delta t$  is given by

$$x(t + \Delta t) = x(t) + \varepsilon, \quad (31)$$

where  $\varepsilon$  is a Gaussian distributed random number,

$$\varepsilon \sim \mathcal{N}\left(0, \sqrt{2d\Delta t}\right), \quad (32)$$

with zero mean and standard deviation  $\sqrt{2d\Delta t}$ . Exponentially distributed life times  $\tau$  for each molecule,

$$p(\tau) = \lambda \exp(-\lambda \tau), \quad (33)$$

take the degradation into account. And production is simply simulated by adding new molecules to the system. The waiting time  $t_w$  between two production events for single molecules is also exponentially distributed according to

$$p(t_w) = c \exp(-c t_w). \quad (34)$$

As this approach works well, we have used it to generate observations as input for our inference algorithm. Figure 1 shows a typical sample obtained from such a simulation. It is clearly visible that spatial fluctuations cannot be neglected in this case.

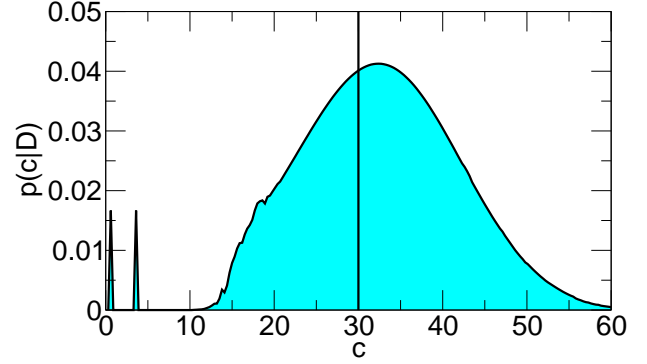


Figure 4: Marginal posterior averaged over 200 data sets for the production rate. The vertical line shows the true parameter value  $c = 30$ .

## 7 Results

In order to test our method we have generated 200 different data sets from simulations of the Bicoid reaction-diffusion system with the same parameters. The values of the parameters are biologically plausible and were taken from Wu et al. (2007). Each sample contained  $11 \times 11$  observations with spatial distance  $\delta x = 20$ , temporal distance  $\delta t = 1$  and noise level  $\sigma = 0.4$ . An example is shown in figure 2.

For each data set we have calculated marginal posteriors for the vector of parameters  $\theta = (c, d, \lambda)$  from a Laplace approximation to the posterior density

$$p(\theta|D) \propto p(D|\theta) p(\theta). \quad (35)$$

Setting

$$F(\theta) \equiv -\ln(p(D|\theta) p(\theta)), \quad (36)$$

Laplace's approximation is given by

$$-\log p(\theta_i|D) \approx F(\theta_i, \theta_i^*) + C + \frac{1}{2} \log \left| \frac{\partial^2 F(\theta_i, \theta_{\setminus i})}{\partial \theta^2} \right|_{\theta=\theta^*}, \quad (37)$$

where  $\theta^*$  denotes the most likely parameters, i.e.  $\theta^* = \arg \min_{\theta} F(\theta)$ ,  $\theta_{\setminus i}$  all parameters *without*  $\theta_i$ , and  $p(D|\theta)$  is computed from (27) using our approximate inference approach. For the prior distribution  $p(\theta)$  we have chosen a product of Gaussian densities cut off at negative values, with standard deviations equal to *half* of and means at the true parameter values. This choice was mainly done for simplicity and may have caused the peaks, which appear in the marginal posteriors for very low parameter values.

The results of averaging the posteriors over the data sets are shown in figures 3, 4 and 5. It is clearly visible, that the modes of the averaged marginal posteriors are good estimates of the rate constants.

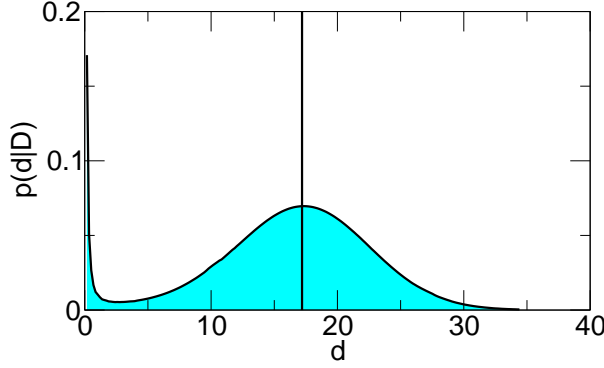


Figure 5: Marginal posterior averaged over 200 data sets for the diffusion constant. The vertical line shows the true parameter value  $d = 17.2$ .

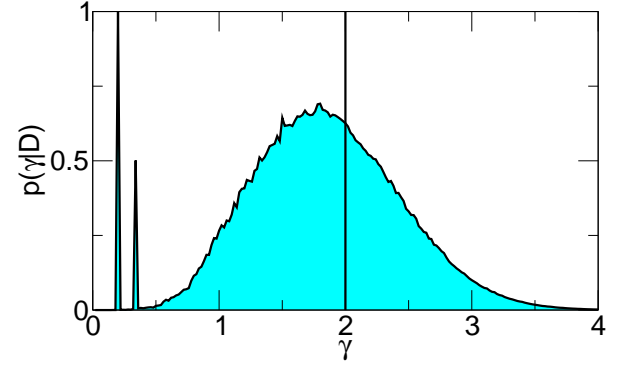


Figure 6: Marginal posterior for  $\gamma$  averaged over 100 rescaled data sets with reaction constants as before. The vertical line shows the true parameter value  $\gamma = 2$ .

## 8 Indirect measurements and the role of fluctuations

While our model uses the density  $\rho(x, t)$  of the molecules in order to describe the state of the reaction-diffusion system, this quantity is usually not directly observable in biochemical experiments. There the molecules are marked with fluorescent particles and one observes the intensity  $I(x, t)$  of the emitted light, which is proportional to the density of the molecules (Wu et al., 2007):

$$I(x, t) = \rho(x, t) / \gamma. \quad (38)$$

However, the constant of proportionality  $\gamma$ , which depends on the preparation of the experiment and other factors, is usually unknown.

We can take these indirect measurements into account by substituting  $\rho(x, t)$  with  $\gamma I(x, t)$  in our model. In this case the production rate  $c$  for molecules is replaced by a production rate  $c_I = c / \gamma$  for intensity. The other rate constants  $d$  and  $\lambda$ , which describe diffusion and degradation respectively, are not changed by rescaling the state variables, as these are first order reactions. Therefore it is not possible to estimate  $\gamma$  using only the deterministic part (16) of the dynamics.

But the covariances of the state variables contained in  $\mathbf{S}$  and likewise the diffusion matrix  $\mathbf{D}(\tilde{\rho})$  are rescaled by  $1/\gamma^2$ , which results in

$$\begin{aligned} (\tilde{\mathbf{D}}_I(\tilde{\mathbf{I}}))_{r,s} &= \frac{1}{2\gamma L} \left[ 2d \left( \frac{r\pi}{L} \right) \left( \frac{s\pi}{L} \right) + \lambda \right] \tilde{I}_{|r-s|} \\ &+ \frac{c_I}{2\gamma L^2}. \end{aligned} \quad (39)$$

As  $\tilde{\mathbf{D}}$  is the inhomogeneity in the linear differential equation (25), it determines the size of fluctuations around the stationary state, which scale proportional to  $1/\gamma$ . Consequently, we can estimate  $\gamma$ , because our model takes the

internal noise of the system into account. For that purpose, we just have to treat  $\gamma$  as additional parameter.

Figure 6 shows the preliminary result of estimating  $\gamma$  averaged over 100 rescaled data sets. Each set contained  $21 \times 6$  observations with spatial distance  $\delta x = 10$ , temporal distance  $\delta t = 2$  and noise level  $\sigma = 0.4$  for the intensities. Here we have kept the parameter  $\lambda$  at its true value. It is clearly visible, that estimation of the number of molecules from intensity measurements is possible using our algorithm. However, the uncertainty for  $\gamma$  is large, which indicates that this is a more difficult inference task.

## 9 Discussion and Outlook

We have developed an approximate inference approach to parameter estimation for a simple class of reaction-diffusion models. Simulations suggest that our method is capable of dealing efficiently with the limit of continuous space and with fluctuations in the density.

While the type of elementary processes considered so far are relevant to the biological *Bicoid* system, the one dimensional spatial geometry might be a strong simplification. Working with a 3-dimensional box geometry would be still possible using our Fourier basis without conceptual changes. More realistic geometries could be treated by using other types of specialized feature functions that are adapted to the boundaries of the problem. Bessel functions, for example, are a suitable basis for circular boundaries.

A more challenging problem is the inclusion of chemical reactions between molecules. Again this can be modelled within the compartment approach and the continuum limit can be taken. However, this will usually lead to *nonlinear* macroscopic rate equations. Fluctuations may be still approximated by Gaussians analogous to (23) and would also be tractable by a suitable feature space representation. However in this case, analytical solutions to the temporal dynamics are not to be expected and a numerical integra-

tions of ODEs is required. But we expect to find only a weak coupling between features of different order. Then we can still work with a reasonably small number of features and the nonlinearity is not a serious problem.

Finally, it will be important to assess the validity of the assumption of Gaussian fluctuations in our approximation. One might think that the success of such an approximation in the limit of infinitely small cell sizes in the compartment model is totally counterintuitive. Small cells contain only few molecules (or none at all) and fluctuations within a cell would be far from being Gaussian. One should note however, that what we really use in the inference method are the fluctuations of the leading Fourier coefficients of the density. Those are obtained from spatial averages of densities modulated with trigonometric functions of long wavelengths. This averaging takes advantage of the smoothness of the density function  $\rho(x, t)$  and might make fluctuations indeed *more* Gaussian. We expect that a formal analysis of our assumptions together with corrections to the approximation could be obtained from a functional integral approach to inference in reaction-diffusion models using ideas similar to Täuber et al. (2005).

## References

- C. Archambeau, M. Opper, Y. Shen, D. Cornford, and J. Shawe-Taylor. Variational inference for diffusion processes. In J. C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, pages 17–24. MIT Press, Cambridge, MA, 2008.
- I. Cohn, T. El-hay, R. Kupferman, and N. Friedman. Mean field variational approximation for continuous-time bayesian networks. In *Proc. Twenty Fifth Conf. on Uncertainty in Artificial Intelligence (UAI 09)*, 2009.
- C. W. Gardiner. *Handbook of Stochastic Methods*. Springer, Berlin, second edition, 1996.
- D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992.
- U. Nodelman, D. Koller, and C. R. Shelton. Expectation propagation for continuous time bayesian networks. In *Proceedings of the Twenty-first Conference on Uncertainty in AI (UAI)*, pages 431–440, 2005.
- M. Opper and G. Sanguinetti. Variational inference for markov jump processes. In J. C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, pages 1105–1112. MIT Press, Cambridge, MA, 2008.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- A. Rutter and M. Opper. Efficient statistical inference for stochastic reaction processes. *Phys. Rev. Lett.*, 103(23): 230601, 2009.
- U. C. Täuber, M. Howard, and B. P. Vollmayr-Lee. Applications of field-theoretic renormalization group methods to reaction-diffusion problems. *J. Phys. A: Math. Gen.*, 38(17):R79–R131, 2005.
- F. Tostevin, P. Rein ten Wolde, and M. Howard. Fundamental limits to position determination by concentration gradients. *PLoS Computational Biology*, 3(4):e78, 2007.
- N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, 1981.
- Y. F. Wu, E. Myasnikova, and J. Reinitz. Master equation simulation analysis of immunostained bicoid morphogen gradient. *BMC Systems Biology*, 1:52, 2007.