Spatio-temporal Gaussian processes modeling of dynamical systems in systems biology

Mu Niu
University of Sheffield

Zhenwen Dai
University of Sheffield

Neil Lawrence
University of Sheffield

kolja becker
Institute of Molecular Biology, Mainz

Abstract

Quantitative modeling of post-transcriptional regulation process is a challenging problem in systems biology. A mechanical model of the regulatory process needs to be able to describe the available spatio-temporal protein concentration and mRNA expression data and recover the continuous spatio-temporal fields. Rigorous methods are required to identify model parameters. A promising approach to deal with these difficulties is proposed using Gaussian process as a prior distribution over the latent function of protein concentration and mRNA expression. In this study, we consider a partial differential equation mechanical model with differential operators and latent function. Since the operators at stake are linear, the information from the physical model can be encoded into the kernel function. Hybrid Monte Carlo methods are employed to carry out Bayesian inference of the partial differential equation parameters and Gaussian process kernel parameters. The spatio-temporal field of protein concentration and mRNA expression are reconstructed without explicitly solving the partial differential equation.

1 Introduction

Quantitative modeling of post-transcriptional regulation is highly topical in systems biology. Considering the vast possibilities of post-transcriptional gene regulation it is evident that protein expression patterns do not necessarily coincide with the location and timing of mRNA transcription. However, in many systems biology models the processes of transcription and translation have been considered together as a single step. This is partly due to the lack of relevant data but also owes to the higher complexity of the problem when allowing for multiple reactions. The exact timing and position of gene expression is crucial in a developmental context and therefore post-transcriptional regulation of gene expression need to be taken into account. Here we review a recently published model [Becker et al., 2013] of such regulation where mRNA and protein concentrations are explicitly considered separate state variables and data is available for both levels of description.

In [Becker et al., 2013] the dynamical model is fitted to spatio-temporal expression data using a weighted least square estimate and the inference is limited to the discrete spatio-temporal grid points. Here we present an efficient Gaussian process based method for Bayesian inference of latent function and associated model parameters. Gaussian processes have been effectively used in machine learning and statistical applications. [Graepel 2003, Lawrence et al. 2006 and Murray-Smith and Pearlmutter 2005] are closely related to this work. [Lawrence et al. 2006] have explored modeling a temporal dynamic system (ordinary differential equation) of transcriptional processes using Gaussian processes. An ordinary differential equation is represented by a integral operator operating on a latent function in [Lawrence et al. 2006]. The model parameters are estimated using maximum likelihood optimization. In this study, we expand the idea of using Gaussian processes for dynamical system modeling to the spatio-temporal case. The partial differential equation model is seen as a linear differential operator applied to a latent function. Considering such linear opera-
tors simplifies the derivation of the covariance kernel. Modeling dynamical systems in systems biology with Gaussian processes lead to several advantages. First, it allows for the inference of continuous quantities without discretization which account for the spatial and temporal structure of the data. Second, the measurement error is naturally inherent to such models and third, it does not require an additional interpolation step to estimate the protein production rate.

We describe in this paper a Gaussian process based approach for estimating the parameters of a model of post-transcriptional regulation described by a partial differential equation. In the second section, the model is rewritten in terms of linear operator and latent function. In the third section we show how linear operators can be included into Gaussian process regression to encode the biological information into the model. Since Becker et al. [2013] suggest the model parameters are correlated, Hybrid Monte Carlo (HMC) methods are used to carry out Bayesian inference of the partial differential equation parameters and Gaussian process hyper parameters in section four. The model is tested with simulated data and section five is dedicated to the application that originally motivated this work.

2 Discriminative Models

In the developing Drosophila embryo the early positioning of body segments is partially controlled by the so called gap genes [Jaeger, 2011; Nusslein-Volhard and Wieschaus, 1980]. These gap genes are expressed in broad overlapping domains along the embryos anterio-posterior axis with the precise positioning resulting from regulation by upstream maternal transcription factors but also from cross-regulation between the different gap genes. This gene regulatory network involved in early Drosophila development has been extensively studied (for example in Jaeger et al., 2004, 2007; Surkova et al., 2009). However, these studies consider either the expression domains of gap gene mRNA or the gap gene proteins separately due to the lack of relevant expression data and the increasing complexity of the problem. More recently, Becker et al. [2013] presented a model in which both layers of description (genes ans proteins) are resolved simultaneously and post-transcriptional regulation is taken into account. The expression of mRNA and protein along the embryos main axes can be quantified in detail using fluorescent microscopy [Becker et al., 2013; Pisarev et al., 2009; PoustelMKova et al., 2004]. In Becker et al. [2013], protein production of a single gap gene is considered to be linearly dependent on its input mRNA concentration at an earlier time point. The model also allows for diffusion of protein between cells and linear protein decay. These processes are dependent on the diffusion parameter and the degradation rate of protein respectively. In order to reformulate the model in the context of Gaussian processes a rescaling of model parameters was carried out:

\[ -D \frac{\partial^2 y(x,t)}{\partial x^2} + \alpha \frac{\partial y(x,t)}{\partial t} + \beta y(x,t) = f(x,t) \]  

(1)

where \( f(x,t) \) is mRNA concentration. \( y(x,t) \) is protein concentration. They are both function of time and space. \( D \) denotes to the diffusion rate of protein (corresponding to \( D/\alpha \) in Becker et al. [2013]). \( \alpha \) is the inverse of the rate of protein production and \( \beta \) is the scaled protein degradation rate. Considering Fick’s laws of diffusion, the diffusion rate \( D \) is fixed to be positive. In the process of translation protein molecules are produced from an mRNA template. The production rate can therefore at minimum be zero corresponding to no protein being produced. Therefore, \( \alpha \) and \( \beta \) need to be positive.

Equation 1 can be rewritten as

\[ f(x,t) = L_{x,t} y(x,t) \]  

(2)

where \( L_{x,t} \) is a linear operator that returns the function \( f(x,t) \) when applied to the function \( y(x,t) \). In general \( L_{x,t} \) can be any kind of linear operator such as integral operator \( L_{x,t} y(x) = \int_{-\infty}^{\infty} y(x_1) dx_1 \) or differential operator \( L_{x,y}(x) = \frac{\partial y(x) }{\partial x} \) [Sarkki, 2011]. In our case, \( L_x \) is a sum of partial differential operator defined as

\[ L_{x,t} = -D \frac{\partial^2}{\partial x^2} + \alpha \frac{\partial }{\partial t} + \beta \]  

(3)

3 Linear Operator and Gaussian Process

In this article, we will model the protein concentration \( y(x,t) \) as a latent function drawn from a zero mean Gaussian process prior distribution with covariance function \( E[y(x,t) y^T(x', t')] = k_{yy}(x,t,x', t') \), which is denoted as

\[ y(x,t) \sim \mathcal{GP}(0, k_{yy}(x,t,x', t')) \]  

(4)

\( k_{yy}(x,t,x', t') \) is defined as a 2 dimensional separable RBF (radial basis function, also known as exponentiated quadratic or squared exponential) spatio-temporal kernel, which is described as

\[ k_{yy}(x,t,x', t') = \sigma_y^2 \exp \left( -\frac{(t-t')^2}{\theta_t^2} - \frac{(x-x')^2}{\theta_x^2} \right) \]  

(5)
It is the tensor product of two separate RBF kernels with different lengthscales \( \theta^2_x \) and \( \theta^2_t \) for the temporal and spatial directions. Applying the rules of linear transformation of Gaussian processes [Papoulis and Pillai 2002, Rasmussen 2006] to \( f(x,t) = \mathcal{L}_{x,t} y(x,t) \) gives

\[
k_{ff}(x,t,x',t') = \text{cov}(f(x,t), f(x',t')) = \text{cov}(\mathcal{L}_{x,t} y(x,t), \mathcal{L}_{x',t'} y(x',t')) = \mathcal{L}_{x,t} \mathcal{L}_{x',t'} k_{yy}(x,t,x',t')
\] (6)

Here \( \mathcal{L}_{x',t'} \) only operates on the second pair of arguments of the kernel \( k_{yy}(x,t,x',t') \). Explicitly, \( k_{ff}(x,t,x',t') \) is given by the following formula

\[
k_{ff}(x,t,x',t') = \left( -D \frac{\partial^2}{\partial x^2} + \alpha \frac{\partial}{\partial t} + \beta \right) k_{yy}(x,t,x',t')
\] (7)

As it is shown in equation (7) the covariance kernel of \( f(x,t,x',t') \) is the result of applying the partial differential operator twice on the kernel of \( y(x,t,x',t') \). The analytical result writes:

\[
k_{ff}(x,t,x',t') = \sigma^2_y \left( \alpha^2 \left( \frac{1}{\theta^2_t} - \frac{(t-t')^2}{\theta^2_t} \right) - 2D\beta \left( \frac{(x-x')^2}{\theta^2_x} - \frac{1}{\theta^2_x} \right) + D^2 \left( \frac{3}{\theta^2_x} - \frac{6(x-x')^2}{\theta^2_x} + \frac{(x-x')^4}{\theta^2_x^3} + \beta^2 \right) \right) \exp \left( -\frac{(t-t')^2}{2\theta^2_t} - \frac{(x-x')^2}{2\theta^2_x} \right)
\] (8)

To infer the mRNA expression function \( f \) and \( y \) can be derived as

\[
k_{fy}(x,t,x',t') = \mathcal{L}_{x,t} k_{fy}(x,t,x',t')
\] (9)

\[
k_{yy}(x,t,x',t') = \mathcal{L}_{x',t'} k_{yy}(x,t,x',t')
\] (10)

Again, the cross covariance function \( k_{fy} \) can be obtained explicitly for the RBF prior on the latent function \( y \):

\[
k_{fy}(x,t,x',t') = \left( -D \left( \frac{(x-x')^2}{\theta^2_x} - \frac{1}{\theta^2_x} \right) + \alpha \left( \frac{t-t'}{\theta^2_t} + \beta \right) \right) \sigma^2_y \exp \left( -\frac{(t-t')^2}{2\theta^2_t} - \frac{(x-x')^2}{2\theta^2_x} \right)
\] (11)

By combining \( f \) and \( y \) as a state vector \( s \), a multi-output Gaussian Process model can be constructed as per equation (12). The variance function \( k_{yy} \) and \( k_{ff} \) can be calculated according to equation (8) and (9). The cross covariance function \( k_{fy} \) and \( k_{gy} \) are derived using equation (11)

\[
s(x,t) = \begin{bmatrix} y(x,t) \\ f(x,t) \end{bmatrix}, \quad K = \begin{bmatrix} k_{yy} & k_{fy} \\ k_{fy} & k_{ff} \end{bmatrix}
\] (12)

\[
s(x,t) \sim \mathcal{GP}(0, K(x,t,x',t'))
\] (13)

where \( K \) is the matrix of covariance function.

### 3.1 Gaussian Process regression

The latent regression function \( f \) and \( y \) have been modeled as zero mean multi-output Gaussian process. The multi-output Gaussian process regression model can be stated as

\[
s(x,t) = \mathcal{GP}(0, K(x,t,x',t'))
\]

\[
S(x_i,t_i) = s(x_i,t_i) + \epsilon_i
\]

\[
\mathcal{E}_{1:n,m} \sim \mathcal{N}(0, \Sigma)
\]

Let us assume that we have noisy observations \( Y \) and \( F \) on a grid of \( n \) in spatial direction and \( m \) in temporal direction. \( S \) is vector representation of the observation. The measurement error \( \epsilon \) is given by the covariance matrix \( \Sigma \). In our case, \( \Sigma \) is a diagonal matrix. Given the vector of measurements \( S = (Y_1, ..., Y_{n \times m}, F_1, ..., F_{n \times m}) \), the posterior mean and variance are given by the Gaussian process regression equations [O’Hagan and Kingman 1978, Rasmussen 2006]:

\[
E[s(x,t)|S] = \mathcal{K}(x,t,x',t') \left( \mathcal{K}(x,t,x',t') + \Sigma \right)^{-1} S
\]

\[
\text{Var}[s(x,t)|S] = \mathcal{K}(x,t,x',t') - \mathcal{K}(x,t,x',t') \left( \mathcal{K}(x,t,x',t') + \Sigma \right)^{-1} \mathcal{K}^T(x,t,x',t') \left( \mathcal{K}(x,t,x',t') + \Sigma \right)^{-1}
\]

Although the main objective of this article is to estimate the partial differential equation parameters, a similar method can also be applied to give the probabilistic solution of the partial differential equation given the known model parameters. For example if we only have measurements of \( f \) and we want to predict \( y \) given the model parameters. Without solving the partial differential equation, we can use Gaussian Process regression model to predict \( y \). The conditional mean and covariance of \( y \) become:

\[
E[y(x,t)|F] = k_{ff}(x,t,x',t') \mathcal{K}(x,t,x',t') \left( \mathcal{K}(x,t,x',t') + \Sigma \right)^{-1} F
\]

\[
[k_{ff}(x',t',t'',t') + \Sigma]^{-1} F
\]
4 Hybrid Monte Carlo

High correlations between model parameters have been suggested by Becker et al. 2013. The normal Markov Chain Monte Carlo sampling takes very long time to traverse the parameter space. Increasing the proposal variance to make bigger transitions may also result in low rates of acceptance and poor mixing of the chain. The consequence of the inefficient random walk proposal is a small effective sample size from the chain [Liu 2008, Robert and Casella 2004]. Hybrid Monte Carlo is believed to be more efficient by producing distant proposal, thereby avoiding the slow exploration of the parameter space [Neal 2011 1996].

In Hybrid Monte Carlo, the deterministic proposal based on Hamiltonian equation is applied along with stochastic proposal to provide ergodic Markov chain [Duane et al. 1987]. The intuition is that the density is treated as the kinetic energy. Each parameter \( \theta_i \) (we use \( \theta \) to represent all the model parameters and kernel hyper parameters) is paired with a momentum variable \( h_i \). The total energy is defined as the negative joint log probability:

\[
H(\theta, h) = -L(\theta) + \frac{1}{2} \log(2\pi)^d |M| + \frac{1}{2} h^T M^{-1} h
\]

(17)

where \( L(\theta) \) is the log of target density. \( h^T M^{-1} h \) is the kinetic energy term. The covariance matrix \( M \) denotes a mass matrix and \( d \) is the dimension of the parameter \( \theta \). The physical analogy of this negative joint log-probability is a Hamiltonian [Duane et al. 1987, Leimkuhler and Reich 2004]. The time evolution of the system is defined by the Hamilton equations [Neal 2011]

\[
\begin{align*}
\frac{d\theta}{d\tau} &= \frac{\partial H}{\partial h} = M^{-1} h \\
\frac{dh}{d\tau} &= -\frac{\partial H}{\partial \theta} = \nabla_{\theta} L(\theta(\tau))
\end{align*}
\]

(18)

where \( \nabla_{\theta} L(\theta(\tau)) \) is the derivative of \( L(\theta(\tau)) \) evaluated at \( \theta(\tau) \). If the parameters \( \theta \) move along the paths based on equation above, they will essentially move along the contours of the target distribution. In practice the Hamiltonian equation is solved numerically to propose movement for \( \theta \). A few steps of parameter updating, known as leapfrog steps, are carried on, in which the auxiliary variable \( h \) and the parameters \( \theta \) are updated alternately. A leapfrog step is defined as:

\[
\begin{align*}
\tau(\tau + \frac{\epsilon}{2}) &= h(\tau(\tau) + \frac{\epsilon}{2} \nabla_{\theta} L(\theta(\tau))) \\
\theta(\tau + \epsilon) &= \theta(\tau) + \epsilon M^{-1} h(\tau(\tau) + \frac{\epsilon}{2}) \\
\tau(\tau + \epsilon) &= h(\tau(\tau) + \frac{\epsilon}{2} + \frac{\epsilon}{2} \nabla_{\theta} L(\theta(\tau) + \epsilon))
\end{align*}
\]

(19)

where \( \epsilon \) is the step size of Hamiltonian move step. After a given number of leapfrog steps, a metropolis update is performed, in which the proposed \( \theta \) and \( h \) is accepted based on the previous \( \theta^* \) and \( h^* \) with probability

\[
\min \left( 1, \exp \left( -H(\theta, h) + H(\theta^*, h^*) \right) \right)
\]

(20)

If the proposed state is not accepted, the next state is kept the same as the previous one. The step size \( \epsilon \) and number of integration steps can be tuned based on the acceptance rate. Heuristics suggest that the choice of matrix \( M \) should rely on the knowledge of marginal variance of the target distribution [Girolami and Calderhead 2011, Liu 2008, Neal 2011].

As it can be seen in the Hamiltonian equation, the updating of \( h \) needs the gradient of the log target density \( \nabla_{\theta} L(\theta) \) which is the gradient of log posterior of \( \theta \). This requires calculating the derivative of the log likelihood respect to each parameters. The log posterior is calculated based on equation

\[
\log(L(\theta)) = \frac{1}{2} \log((2\pi)^n |K + \Sigma|) + \\
\frac{1}{2} S^T (K + \Sigma)^{-1} S - \log(p(\theta))
\]

(20)

where \( p(\theta) \) is the prior. The gradient of log likelihood is given by

\[
\frac{\partial \log(L(\theta))}{\partial \theta} = S^T K^{-1} \frac{\partial K}{\partial \theta} K^{-1} S = \sum \frac{K^{-1} S S^T K^{-1}}{\partial \theta} \frac{\partial K}{\partial \theta} = \sum \frac{\partial L}{\partial \theta} \frac{\partial K}{\partial \theta}
\]

(21)

If we replace \( K^{-1} S S^T K^{-1} \) as \( \partial L/\partial K \), the gradient of log likelihood become the element sum of hadamard product of \( \partial L/\partial K \) and \( \partial K/\partial \theta \). \( \partial K/\partial \theta \) is defined as

\[
\frac{\partial K}{\partial \theta} = \begin{bmatrix} \frac{\partial k_{i,i}}{\partial \theta_{1,1}} & \cdots & \frac{\partial k_{i,i}}{\partial \theta_{d,d}} \\
\vdots & \ddots & \vdots \\
\frac{\partial k_{i,i}}{\partial \theta_{1,1}} & \cdots & \frac{\partial k_{i,i}}{\partial \theta_{d,d}} \end{bmatrix}
\]

(22)

\( \partial K/\partial \theta \) is the gradient of kernel. It is derivative of equation (12) which is the matrix of gradient of \( k_{yy} , k_{ff} \) and cross covariance.
4.1 Implementation with simulation data

Before tackling real data in section 5, we look at inference for simulated data. We consider trigonometric functions for $y$ and $f$ with some observation noise. $D$, $\alpha$ and $\beta$ are chosen to be 1. We randomly pick some time points and spatial location. Noisy observations of $y$ and $f$ are generated using equation 23:

\[
Y(x,t) = \cos(x) + \sin(t) + e
\]
\[
F(x,t) = \sin(t) + \cos(t) + 2\cos(x) + e
\]
\[
e \sim N(0, \sigma_0)
\] (23)

The model parameters are estimated using HMC algorithm. We run it for 7000 iterations after burn-in. The posterior mean and standard deviation of model parameters are shown in Table 1, along with the actual values used in simulation. Posterior density plots for the parameters are given in Figure 2. All the posterior distributions are consistent with the true values. The significant correlation between $\alpha$ and $\beta$ (Figure 2 d) implies that HMC is more effective than MCMC for these settings.

| Parameter | Mean  | Standard deviation | True |
|-----------|-------|--------------------|------|
| D         | 0.993 | 0.02               | 1    |
| $\alpha$  | 1.002 | 0.028              | 1    |
| $\beta$   | 1.002 | 0.013              | 1    |

Figure 1: spatio-temporal field of $y$ and $f$ with optimized model parameters for simulation data.

Figure 2: Posterior densities for model parameters and joint posterior of $\alpha$ and $\beta$ with simulation data, based on the Hybrid Monte Carlo runs of 7000 iterations. (a) $D$ is the protein diffusion rate. (b) $\alpha$ is the inverse of protein production rate. (c) $\beta$ is the scaled protein degradation rate. (d) joint density of $\alpha$ and $\beta$.

The optimized spatio-temporal field is plotted in Figure 1 by setting model parameter as the posterior mean. The colored contour in the figure represent the predicted mean of $y$ and $f$. The colored circle represent the data points simulated from trigonometric functions. These results show that the multioutput Gaussian process model and Bayesian inference capture the key mechanism at work here.

5 Implementation with real data

In this section we have fitted the spatio-temporal multioutput Gaussian processes to existing data from Becker et al. [2013]. For this we have chosen the mRNA and protein expression data of the gap gene Knirps.
Considering the available information for the measurement error, the heteroscedastic Gaussian process regression model is applied by assigning fixed variance (measurement error) for each data point. The prior distribution of model parameters are given based on the results in Becker et al. 2013.

Table 2: Model Parameter estimate with real data

| Parameter | Mean | Sd  | Becker[2013] |
|-----------|------|-----|--------------|
| D         | 0.017| 0.023| 0.16         |
| α         | 13.04| 0.72 | 12.771       |
| β         | 0.65 | 0.022| 0.983        |

The density of model parameters are plotted in Figure 5. As can be seen in panel (d), the parameters α and β are correlated. The spatio-temporal field of y and f are presented in Figure 4. As we have noised observations, the colored circle represent the mean of data points.

The order of magnitude of the estimated parameters α and β are comparable to the estimated parameter values in Becker et al. 2013. According to the model, these parameters are confidently estimated since the standard deviation of their posterior is small. On the other hand, there is a ten fold difference between the estimated parameter value \( D = 0.017 \) in this article and the one in Becker et al. 2013 \((D = 0.16 \text{ after rescaling})\). The difference might be caused by the different modeling approach. The exclusion of the delay parameter from the original model could also potentially account for the difference. However, the estimated parameter value for D of the results presented here is well contained within the confidence interval given by Becker et al. 2013. Also the significant correlation between production and decay rates has been correctly captured using the Gaussian process approach (see Figure 4 (d)). Before concluding, we would like to stress that both the approach based on Gaussian processes presented in this paper and the parameter estimation and identifiability analysis method in Becker et al. 2013 reach qualitatively similar results. The biological interpretation regarding the time scale of protein decay and the practical non-existence of diffusion in the system therefore remain untouched.

Figure 3: Trace plot of model parameter for simulated data based on the Hybrid Monte Carlo runs of 7000 iterations. (a) \( D \) is the protein diffusion rate. (b) \( α \) is the inverse of protein production rate. (c) \( β \) is the scaled protein degradation rate. (d) joint posterior of \( α \) and \( β \).

Figure 4: spatio-temporal field of y and f with optimized model parameters for real data.
Figure 5: Posterior densities for model parameters and joint posterior of $\alpha$ and $\beta$ with real data, based on the Hybrid Monte Carlo runs of 10000 iterations. (a) $D$ is the protein diffusion rate (b) $\alpha$ is the inverse of protein production rate (c) $\beta$ is the scaled protein degradation rate (d) joint density of $\alpha$ and $\beta$.

6 Conclusion

In this article we discussed how Gaussian processes can be effectively used to model a spatio-temporal dynamical model of post-transcriptional regulation. Comparing with the standard parametric method, our non-parametric approach does not require restricting the inference to the observed spatio-temporal data points or any form of grid points. The continuity of the spatio-temporal field is naturally accounted for in our Gaussian process model. Hybrid Monte Carlo is employed to infer model parameters. The correlation between partial differential equation parameters are successfully captured from the posterior density.

Although the transcriptional delay has been dropped from the original partial differential equation, the delay operator is linear so it could be included in the analysis without requiring any additional theoretical results. However, Lawrence et al. (2006) states the data need to be sampled at a reasonably high frequency to identify the delays. A promising development for future research is to introduce nonlinearities into the dynamic model as in the Michaelis-Menten model (Rogers et al., 2006).

Figure 6: Trace plot of model parameter for real data based on the Hybrid Monte Carlo runs of 10000 iterations. (a) $D$ is the protein diffusion rate. (b) $\alpha$ is the inverse of protein production rate. (c) $\beta$ is the scaled protein degradation rate. (d) joint posterior of $\alpha$ and $\beta$.

Acknowledgements

References

Kolja Becker, Eva Balsa-Canto, Damjan Cicin-Sain, Astrid Hoermann, Hilde Janssens, Julio R Banga, and Johannes Jaeger. Reverse-engineering post-transcriptional regulation of gap genes in drosophila melanogaster. PLoS computational biology, 9(10): e1003281, 2013.

Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. Physics letters B, 195(2):216–222, 1987.

Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(2):123–214, 2011.

Thore Graepel. Solving noisy linear operator equations by gaussian processes: Application to ordinary and partial differential equations. In ICML, pages 234–241, 2003.

J. Jaeger. The gap gene network. Cell. Mol. Life Sci., 68(2):243–274, Jan 2011.

J. Jaeger, S Surkova, M Blagov, H Janssens, D Kosman, KN Kozlov, Manu, E Myasnikova, CE Vanario-Alonso, M Samsonova, DH Sharp, and J Reinitz. Dynamic control of positional information in the early Drosophila embryo. Nature, 430(6997): 368–371, 2004.
J. Jaeger, D. H. Sharp, and J. Reinitz. Known material gradients are not sufficient for the establishment of gap domains in *Drosophila melanogaster*. *Mech. Dev.*, 124(2):108–128, Feb 2007.

Neil Lawrence, Guido Sanguinetti, and Magnus Ratclay. Modelling transcriptional regulation using gaussian processes. 2006.

Benedict Leimkuhler and Sebastian Reich. *Simulating hamiltonian dynamics*, volume 14. Cambridge University Press, 2004.

Jun S Liu. *Monte Carlo strategies in scientific computing*. springer, 2008.

Roderick Murray-Smith and Barak A Pearlmutter. Transformations of gaussian process priors. In *Deterministic and Statistical Methods in Machine Learning*, pages 110–123. Springer, 2005.

Radford Neal. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2, 2011.

Radford M Neal. Sampling from multimodal distributions using tempered transitions. *Statistics and computing*, 6(4):353–366, 1996.

C. Nusslein-Volhard and E. Wieschaus. Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287(5785):795–801, Oct 1980.

Anthony O’Hagan and JFC Kingman. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–42, 1978.

Athanassios Papoulis and S Ummikrishna Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

A. Pisarev, E. Poustelnikova, M. Samsonova, and J. Reinitz. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic Acids Res.*, 37(Database issue):D560–566, Jan 2009.

E. Poustelnikova, A. Pisarev, M. Blagov, M. Samsonova, and J. Reinitz. A database for management of gene expression data in situ. *Bioinformatics*, 20 (14):2212–2221, Sep 2004.

Simo Särkkä. Linear operators and stochastic partial differential equations in gaussian process regression. In *Artificial Neural Networks and Machine Learning–ICANN 2011*, pages 151–158. Springer, 2011.

Svetlana Surkova, Alexander V Spirov, Vitaly V Gursky, Hilde Janssens, Ah-Ram Kim, Ovidiu Radulescu, Carlos E Vanario-Alonso, David H Sharp, Maria Samsonova, John Reinitz, et al. Canalization of gene expression and domain shifts in the drosophila blastoderm by dynamical attractors. *PLoS computational biology*, 5(3):e1000303, 2009.