UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL AND APPLIED SCIENCES

Electronics and Computer Science

Machine Learning Approaches to Modelling Bicoid Morphogen in *Drosophila melanogaster*

by

Wei Liu

Thesis for the degree of Doctor of Philosophy

March 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
Electronics and Computer Science

Doctor of Philosophy

MACHINE LEARNING APPROACHES TO MODELLING BICOID MORPHOGEN IN *DROSOPHILA MELANOGASTER*

by Wei Liu

Bicoid morphogen is among the earliest triggers of differential spatial pattern of gene expression and subsequent cell fate determination in the embryonic development of *Drosophila melanogaster*. This maternally deposited morphogen, diffusing along the anterior-posterior axis of the embryo, establishes a concentration gradient which is sensed by target genes. In most computational model based analyses of this process, the translation of the *bicoid* mRNA is thought to take place at a fixed rate in the anterior pole of the embryo. Is this process of morphogen generation a passive one as assumed in the modelling literature so far, or would available data support an alternate hypothesis that the stability of the mRNA is regulated by active processes?

This thesis demonstrates a Bicoid spatio-temporal model in which the stability of the maternal mRNA is regulated by being held constant for a length of time, followed by rapid exponential degradation. With the mRNA regulation, three computational models of spatial morphogen propagation along the anterior-posterior axis are analysed: (a) passive diffusion with a deterministic differential equation, (b) diffusion enhanced by a cytoplasmic flow term and (c) stochastic diffusion modelled by Gillespie simulation. Comparison of the parameter estimation in these models by matching to the publicly available data, `FlyEx`, suggests strong support for mRNA regulated stability.

With a non-parametric Bayesian setting, we have applied Gaussian process regression to infer the mRNA regulation function as a posterior density. With synthetic data obtained from a linear spatio-temporal dynamical system and the experimental measurements (`FlyEx`), this approach is capable of inferring the driving input. Apart from confirming the validity of a regulated mRNA source, this work also demonstrates the applicability of a powerful non-parametric model of Gaussian processes in a spatio-temporal inference problem. In line with recent experimental works, we have also analysed this model with a spatial gradient of maternal mRNA, rather than being fixed at the anterior pole.

Our final work is to analyse the dynamical topology of the gap gene network, which is the major developmental activity, taking place after the establishment and interpretation

of maternal morphogen Bicoid. This network consists of six transcription factors cross-regulating each other. The main focus in this work is the external input – Bicoid, as a trigger to initialise all these gap genes. The establishment of the gap gene network happens about two hours following fertilisation, precisely the time duration in which translation of the *bicoid* maternal mRNA is switched off. Hence, we are interested in asking what effect the rapidly decaying Bicoid concentration might have on the behaviour of the gap gene network. On addressing this concern, we have refined the existing framework with time-varying Bicoid input from `FlyEx`. For this, our results suggest potential changes within the network structure.

Contents

List of Figures

List of Tables

Nomenclature

Mathematical notations

$m(\cdot)$	Morphogen concentration	Page 26
$J(\cdot)$	Flux	Page 26
h	Length of each compartment on A-P axis	Page 26
D	Diffusion coefficient	Page 27
Ω	Volume of each compartment	Page 28
$\partial\Omega$	Boundary of Ω	Page 28
τ_p	Protein degradation time	Page 29
λ	Morphogen gradient decaying length scale	Page 100
L	Embryo length	Page 100
S_0	Source concentration	Page 30
$\delta(\cdot)$	Kronecker delta function	Page 30
$\Theta(\cdot)$	Heaviside step function	Page 30
τ_m	*bicoid* mRNA degradation time	Page 31
t_0	mRNA decaying onset time	Page 31

V	Cytoplasm flow velocity	Page 33
k	Molecule degradation rate	Page 34
a	Propensity function	Page 34
r	Uniformly distributed number	Page 35
τ	Stochastic time step	Page 35
g	Production rate	Page 36
d	Diffusion rate	Page 41
m_d	Measurements from `FlyEx`	Page 46
$\mathcal{N}$	Gaussian distribution	Page 60
μ	Mean	Page 60
σ	Standard deviation	Page 60
$\mathbf{\Sigma}$	Covariance matrix	Page 60
l	Time scale of covariance matrix	Page 23
$\mathbf{K}$	Covariance matrix	Page 62
$y(\cdot)$	Observations	Page 63
ε	Gaussian noise	Page 63
v	Gap gene concentration	Page 86
Φ	Sigmoid function	Page 86
R	Synthesis rate	Page 86
$\mathbf{W}$	Genetic interconnectivity matrix	Page 86

Abbreviations

FISH	Fluorescent in Situ Hybridization
A-P	Anterior-Posterior
EL	Embryo Length
SDD	Synthesis, Diffusion and Degradation
D-V	Dorsal-Ventral
mRNA	messenger Ribonucleic Acid
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
eGFP	enhanced Green Fluorescent Protein
ARTS	Active RNA Transport and Synthesis
BIE	*bicoid* Instability Element
3′UTR	3′ Untranslated Region
PLSA	Parallel Lam Simulated Annealing
GP	Gaussian Process
RBF	Radial Basis Function
MLE	Maximum Likelihood Estimation
MCMC	Markov chain Monte Carlo

DECLARATION OF AUTHORSHIP

I, Wei Liu, declare that the thesis entitled `Machine learning approaches to modelling Bicoid morphogen in Drosophila melanogaster` and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:
  - W. Liu and M. Niranjan. Gaussian process modelling for *bicoid* mRNA regulation in spatio-temporal Bicoid profile. *Bioinformatics*, 28(3): 366–372, 2012.
  - W. Liu and M. Niranjan. The role of regulated mRNA stability in establishing Bicoid morphogen gradient in *Drosophila* embryonic development. *PLoS ONE*, 6(9): e24896, 2011.
  - W. Liu and M. Niranjan. Matching models to data in modelling morphogen diffusion. In S. Džeroski *et al.* (eds) Proceedings of The Third International Workshop on Machine Learning in Systems Biology. Helsinki University Printing House, Finland, pp. 55–64

Signed.....Date.....

Acknowledgements

This thesis is the most significant academic challenge I have faced since I started my research work in Communications, Signal Processing and Control Group. It would not have been possible without the support, patience and guidance of this group. It is my pleasure to thank those who help me during this period.

In the first place, I would like to record my gratitude to Professor Mahesan Niranjan for his supervision, advice and guidance during this early stage of my research journey. The encouragement and support he has provided always inspired me and made me understand how to be a researcher rather than a student. I am indebted to him more than he knows.

I would like to thank Dr. Srinandan Dasmahapatra and Dr. Ivan Markovsky for the valuable guidance and discussion of my research. I am also grateful to Professor Lajos Hanzo and Professor Lie-Liang Yang who supervised me during the MSc and inspired me to start PhD.

Many thanks to our research group – my friends and colleagues (in alphabetical order) – Abdullah Alrajeh, Ali Hassan, Chathurika Dharmagunawardhana, Daisy Tong, Mustansar Ali Ghazanfar, Shaobai Li, Sung Uk Jung, Tayyaba Azim, Wanmu Liu and Xin Liu. They made the office convivial to work. In particular, I would like to thank Dr. Ramanan, who left this group two years ago, for his advice on my research work and the mementos of the long time we spent working together in the office.

My special thanks go to Ke Yuan, my love, without whom my effort would have been worth nothing. His love and support always inspires me not only for research but also for life.

This is also a great opportunity to express my love to my family – Mum, Dad and Uncle. They deserve this special mention for their ongoing support and gentle love.

To my love . . .

Chapter 1

Introduction

1.1 Background

Life is as beautiful as an electrifying concert. From a few notes on the piano to a wonderful musical composition, what we can hear from a single cell to a complicated multicellular organism is the music of life. Differing from the non-living instruments, this special concern is composited with the uncountable biochemical interactions based on genetic keys. One can imagine how enormous the biological data are produced from these ongoing interactions in our bodies. In spite of the complexity, scientists are still searching for the generalised laws of the biological systems. However, the traditional experimental technologies are no longer appropriate to help us understand the principles of development. A more powerful approach needs to be considered, such as systems biology, which demonstrates biology at the system level and integrates the dispersive elements as an ensemble. By studying the biological principles based on mathematical frameworks, systems biology provides a system-level understanding of the connections, the dynamics and the structures between the elements. It is notable that the basic mathematical equations could reveal some biological principles, such as transcription networks, robustness of the protein circuits which are detailed by Alon (2006).

The particular interest in this thesis is bringing machine learning approaches to model morphogen propagation in the *Drosophila* embryo. In biological development, the most interesting and challenging problem is pattern formation, in which the differential cell fate determination is mainly thought to be caused by spatial concentration gradients of morphogens. This view, put forward by Turing (1952) over six decades ago, is a computational model which predicted the mechanism long before an example of it was discovered in the real world.

The idea that morphogens diffuse from a localised source and provide positional information by different concentration thresholds is formalised by Wolpert (1969) with a

French flag model. The most definitive example, *bicoid*[1], maternally deposited as mRNA at the anterior pole of *Drosophila melanogaster* embryos, is translated into protein and propagates along the anterior-posterior (A-P) axis, setting up a concentration gradient (Driever and Nüsslein-Volhard, 1988a,b; St Johnston et al., 1989). This, in conjunction with other similar transcription factors, regulates the establishment of the segmental structure by precise activation of gap genes (Driever and Nüsslein-Volhard, 1989; Struhl et al., 1989; Ephrussi and Johnston, 2004). Several computational models of Bicoid gradient formation have been published over the last three decades (reviewed by Grimm et al. (2010)). The most widely used model is the formulation of Wolpert (1969) and Crick (1970) published before the discovery of the role of Bicoid, in which a combination of protein synthesis, diffusion and degradation (SDD) is the underlying mechanism that derives a steady state concentration gradient. Decoding differential concentrations from such a gradient, which is spatially exponential in the steady state, and its robustness properties are discussed in (Driever and Nüsslein-Volhard, 1988b; Houchmandzadeh et al., 2002). The topic of the Bicoid gradient formation has become increasingly popular in recent years because the traditional steady state decoding is hard to explain some experimental findings (Gregor et al., 2007b,a). Therefore, the question of how Bicoid gradient forms still remains open.

1.2 Motivation

To the best of our knowledge, all the computational models in the literature assume that the translation of *bicoid* maternal mRNA takes place at a constant rate at the anterior end, resulting in a constant supply of morphogen to diffuse in the system. Although mathematically convenient, in that it leads to easy closed-form solutions, this is an unrealistic assumption, for there is no need for the embryo to continue to maintain a constant supply of morphogen beyond what is needed for genetic decoding.

A particular view on this subject, supported by experimental findings, is advanced by Surdej and Jacobs-Lorena (1998) who argued that the stability of the *bicoid* mRNA is regulated during development; the mRNA being held stable during the first two hours of development and rapidly killed off thereafter. Spirov et al. (2009)'s work, proposing an mRNA spatial distribution for *bicoid*, also contains further experimental evidence pointing in this direction. By the fluorescence in situ hybridization (FISH) method and confocal microscopy, these authors confirm that *bicoid* mRNA disappears below detectable levels around 16 min after the onset of nuclear cycle 14 with complete mRNA degradation taking place over a time interval of 15 to 20 min.

Following these two experiments, we have established the *bicoid* mRNA regulation model with the degradation in the later developmental stage. Such a regulation framework

[1]The italic lower-case *bicoid* represents the gene and Bicoid refers to protein. It is similar for the other genes.

can provide a better fitting between protein formation and real measurements. More importantly, the establishment of zygotic gene network occurs in the same time duration that mRNA degrades. It is interesting to ask how regulated Bicoid affects the gap gene expressions during cleavage cycle 14.

1.3 Contribution

1.3.1 *bicoid* mRNA stability regulation – *in-silico* analysis

In the first part of this thesis, we have pursued these observations of the regulation of stability, leading to a model of morphogen propagation in which the source supply is assumed to consist of a constant part during early development, followed by an exponential decay. In Chapter 3, we integrate such a source model into two different models of morphogen propagation and match the resulting spatio-temporal profiles to measurements published by `FlyEx` database (Myasnikova et al., 2001; Poustelnikova et al., 2004; Pisarev et al., 2009). By matching the model output to `FlyEx` measurements, using a least squares fitting method, we infer optimal parameters of each of the models, including the time at which mRNA stability is destroyed. We also quantify the uncertainties in these estimates by constructing bootstrapped sample paths through different individual fly measurements, taken at different developmental stages. Our results show that the estimated parameters all lie within sensible ranges of values, and the decay onset time inferred from the data coincides well with the experimental observations in (Surdej and Jacobs-Lorena, 1998; Spirov et al., 2009).

As such, ours is the first *in-silico* study that incorporates a novel mechanism of developmental regulation by which a morphogen gradient is established when needed, and killed off by some active processes once its task is accomplished. This is something one would naturally expect, but which has been ignored in three decades of modelling work on the subject.

1.3.2 Inference of *bicoid mRNA* – Gaussian process approach

Secondly, we build on the algorithmic foundations and apply the non-parametric probabilistic approach of Gaussian process (GP) regression (Rasmussen and Williams, 2006) to address the problem of modelling Bicoid morphogen propagation. The GP model is the tool of choice for regression problems characterised by the need to model uncertainties and deal with unobserved data in a systematic way, both of which are aspects of our problem. The approach has been demonstrated in a wide range of applications in the machine learning literature. Specifically, Lawrence et al. (2007); Gao et al. (2008)

Table 1.1: Maternal and gap genes in the early embryonic development.

	Gene Name	Maternal/Zygotic	Protein Distribution
Maternal genes	*bicoid*	M	Anterior
	hunchback (hb)	M/Z	Anterior & Posterior
	nanos (nos)	M	Posterior
	caudal (cad)	M/Z	Posterior
Gap genes	*hunchback (hb)*	M/Z	Anterior & Posterior
	Krüppel (Kr)[2]	Z	Middle
	giant (gt)	Z	Anterior & Posterior
	knirps (kni)	Z	middle
	tailless (tll)	Z	Anterior &Posterior
	huckebein (hkb)	Z	Anterior & Posterior

[2] The initial letter of *Krüppel (Kr)* is capital since the mutation of this gene is dominant.

demonstrate pioneering work in this area, in which GPs are shown to be very effective in inferring target genes regulated by the tumour suppression transcription factor $p53$.

In Chapter 4, we derive the computational strategy needed to extend Lawrence et al. (2007)'s work to deal with spatio-temporal problems, and demonstrate its application to Bicoid regulation modelling using both synthetic data and real data from `FlyEx` database (Pisarev et al., 2009). The results demonstrate the power of the method on synthetic data and its limitations on real data. As such, this work is the first contribution that adapts the powerful algorithmic setting of non-parametric regression to tackle an important spatio-temporal inference problem in developmental biology.

1.3.3 *bicoid* mRNA regulation in segmentation – gap gene network

The complex expression of gap genes that drive segmentation along the A-P axis is studied in a number of works (Reinitz and Sharp, 1995; Jaeger et al., 2004b,a; Manu. et al., 2009a,b; Ashyraliyev et al., 2009) by the construction of a gap gene regulatory model. This body of work shows how dynamical properties of a non-linear network of interacting transcription factors achieves segmentation along the A-P axis by differential expressions. Remarkably, the models are able to exhibit dynamical shifts of gap gene expression peaks from posterior towards anterior. These authors mostly assume Bicoid to have a sustained exponential steady state profile throughout the analysis intervals they consider, a questionable assumption since it is precisely during this time interval that the morphogen degrades rapidly. In Chapter 5, we refine the gap gene model proposed by Jaeger et al. (2004b,a) with the time-varying external input – Bicoid concentration gradient from `FlyEx`. This work suggests that the genetic topology is potentially effected.

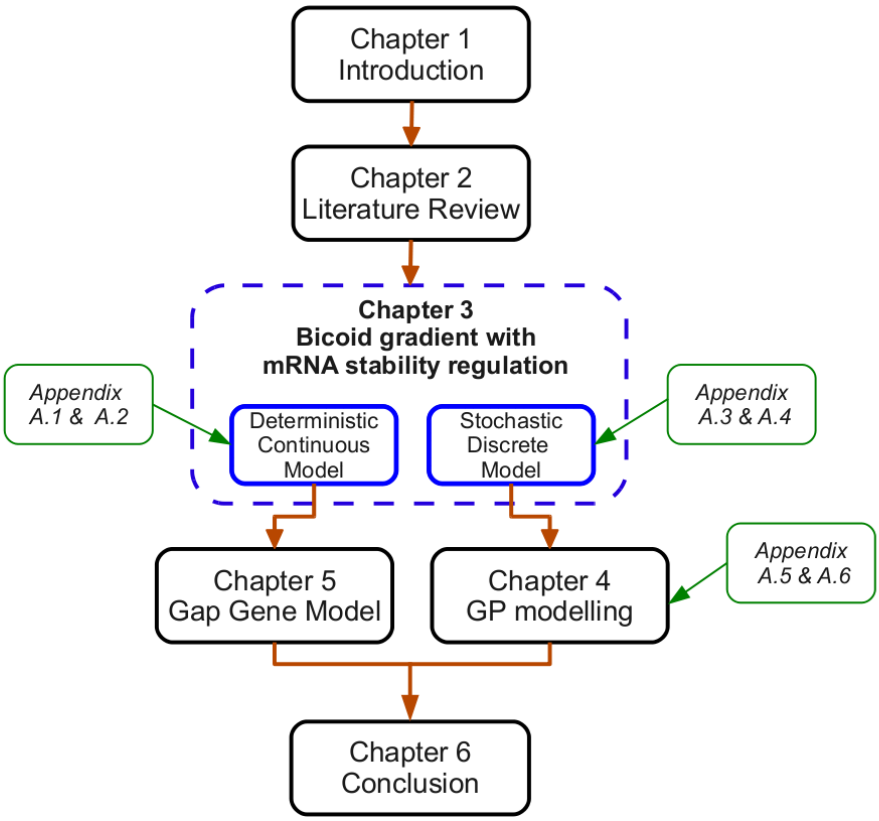


Figure 1.1: Thesis structure

1.4 Thesis organisation

This thesis is organised as in Figure 1.1. After this introductory chapter, Chapter 2 reviews models of Bicoid protein concentration gradient establishment in *Drosophila* early embryonic development, cooperating with constant Bicoid translation rate, which has been a general assumption over the decades. Further, the more complicated development, segment determination involving gap gene connections, described by the gap gene network has been reviewed. At the end, the review turns to a powerful Bayesian inference framework, GPs, of which the particular interest is to infer posterior function of the latent chemical species in biological system. Chapters 3, 4 and 5 comprise the main works of the thesis. In Chapter 3, we introduce two mathematical Bicoid reaction-diffusion systems – deterministic and stochastic models – in which the stability of *bicoid* mRNA has been regulated. In Chapter 4, by GP modelling, we infer such source regulation as a driving function from noisy observations of Bicoid spatio-temporal protein profile. In Chapter 5, we discuss the potential change in gap gene regulatory network when the constant Bicoid intensity input is replaced with the regulated intensity from `FlyEx` database. Finally, the conclusions of this research and future works will be described in Chapter 6. Appendix A (A.1-A.6) includes several derivation details and is attached after Chapter 6.

1.5 Publications and presentations

The following publications and presentations are based on the contributions made during this research.

Publications

- W. Liu and M. Niranjan. Gaussian process modelling for *bicoid* mRNA regulation in spatio-temporal Bicoid profile. *Bioinformatics*, 28(3): 366–372, 2012.
- W. Liu and M. Niranjan. The role of regulated mRNA stability in establishing Bicoid morphogen gradient in *Drosophila* embryonic development. *PLoS ONE*, 6(9): e24896, 2011.
- W. Liu and M. Niranjan. Matching models to data in modelling morphogen diffusion. In S. Džeroski *et al.* (eds) Proceedings of The Third International Workshop on Machine Learning in Systems Biology. Helsinki University Printing House, Finland, pp. 55–64

Presentations

- Learning and Inference in Computational Systems Biology, Warwick, United Kingdom, March 2010.
- Pascal2 Workshop on Spatio-temporal Modelling, Edinburgh, United Kingdom, October 2009.
- Third International Workshop on Machine Learning in Systems Biology, Ljubljana, Slovenia, September 2009.

Chapter 2

Literature review

This chapter provides a review of the literature on *Drosophila melanogaster* early embryonic development during the first three hours after fertilisation. We start from the positional information in Section 2.1 which explains how gene expresses on the proper location. This information is provided by a class of signalling molecules known as morphogens. The first experimentally discovered maternal provided morphogen Bicoid is introduced and computational models of Bicoid propagation are reviewed in Section 2.2. The quantitative experimental database `FlyEx` is introduced in Section 2.3. In Section 2.4, we review the gap gene regulatory network which is initialised by Bicoid. As the Bayesian approach to study the biological system, Gaussian process (GP) modelling is reviewed in Section 2.5.

2.1 Positional information and morphogen

How does a simple embryo develop as a well organised body with head, eyes and other organs? How can we determine the fate of millions or billions of cells in an organism? The answers to these questions remained elusive concerning the development of complicated organisms, until the end of the nineteenth century. By separating the cells of the sea urchin embryo after the first cell division, in 1891, Driesch discovered that each of the cells developed into a small larva instead of the corresponding half of the sea urchin (Driesch, 1908). This experiment implies that there is a coordinate system that can specify the cells' position information. Based on these findings, Morgan (1901) later suggested that this regeneration is controlled by some substances which can determine the pattern formation by gradients (Wolpert, 1996; Wartlick et al., 2009). The gradients became a popular topic of study in the 1930s (Wolpert, 1996). Spemann and Mangold (1924) suggested that such a gradient is released from a group of localised cells and a particular gradient had been found in metabolism by Child (1941). Although these

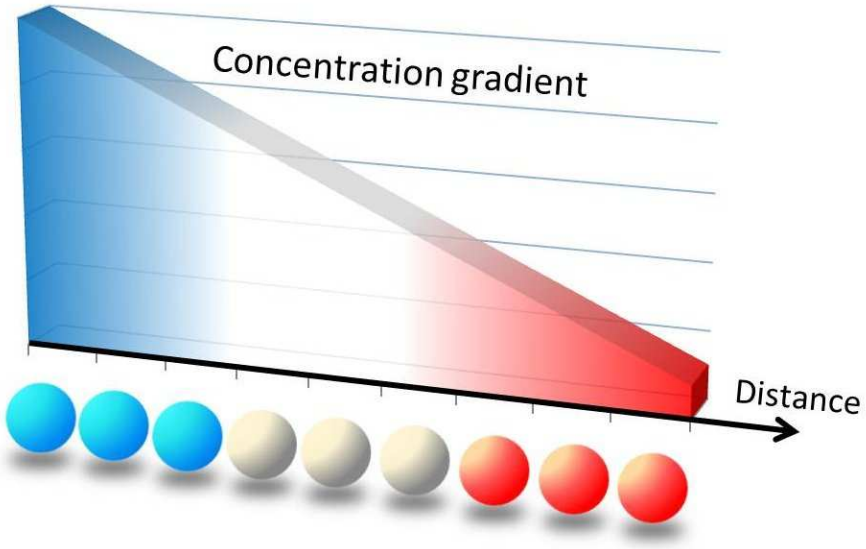


Figure 2.1: French flag model. Morphogen concentration is shown with flag and the corresponding cells are shown with spheres.

works could explain Driesch's discovery to a certain extent, the generation of the discrete regions of the pattern from a continuous gradient still remained a question. Dalcq (1938) proposed the idea of thresholds of the morphogenesis and Sander in 1960 proved the possibility that the determination of the patterns can be controlled by the gradients within the particular boundaries (Sander, 1971; Rogers and Schier, 2011).

Morphogen, which represents the chemical substance above, was first named by Turing (1952). He introduced a possible mechanism of the progress in self-organised patterns, the morphogen reaction-diffusion system, which establishes the concentrations through tissues by the interactions between the diffusing molecules and generates the resulting organism. With Turing's reaction-diffusion model, as shown below, the two dimensional spatial patterns are formed by considering the chemical reactions of two or more morphogens in a field of cells. Morphogens x and y reaction-diffusion system in the cell r is given by:

$$\begin{aligned} \frac{dx_r}{dt} &= \mu(x_{r+1} - 2x_r + x_{r-1}) + f(x_r, y_r), \quad r = 1, ..., N \\ \frac{dy_r}{dt} &= \nu(y_{r+1} - 2y_r + y_{r-1}) + g(x_r, y_r), \quad r = 1, ..., N \end{aligned} \tag{2.1}$$

where μ and ν are diffusion constant and $f(\cdot)$ and $g(\cdot)$ are morphogen reaction functions. This is the first theoretical model formalising the idea of a concentration-dependent manner in biological pattern formation.

In the biological field, on the studies of *Galleria mellonella* pupa, Stumpf (1966) showed that the concentration gradient of inductive molecules have contributes to developmental patterning. Later, the positional information mechanism that explains how the genetic information can be translated into the spatial patterns was formalised by Wolpert (1969). The basic idea of the position information, specifying and interpreting of the cells' positional values, is illustrated by a classic model, the French flag, as shown in Figure 2.1. Two key factors that establish positional information are boundary and scalar which define the specified position and the distance from the boundary, respectively (Wolpert, 1989). Before the genetic programme starts, the cells have the potentials to develop as one of three colours of the French flag. After the morphogen concentration gradient is established, these cells are given the positional values specified by the concentration at each point and the flag pattern, shown in Figure 2.1, will be generated after the positional values are interpreted by cells. The three colours are differentiated according to different pre-determined genetic information in cells (Wolpert, 1996, 2007). Importantly, the position information only controls the decisions of relating to the fate of the cells instead of specifying the genetic instructions. Therefore, the same morphogen could generate different tissues by providing position information only (Wolpert, 1969).

An early simple morphogen diffusible model (Source-Sink model) in embryogenesis is proposed by Crick (1970). In this model, morphogen molecules are produced from the localised source cells and destroyed in the other end by sink cells to form morphogen distribution by the property of diffusion over a field of cells (Figure 2.2). This linear gradient can induce different gene expressions in the concentration-dependent manner, which can be explained well by Wolpert's French flag model above. This simple Source-Sink model had a significant influence on the later morphogen works although we know the sink component is not necessary nowadays because the morphogen concentration can be formed even without degradation. At the same time, other morphogen studies were continued. For example, the digits development of the chick limb bud, explained by morphogen gradient diffusion, evidences that the pattern formation can be formed by interpretation of position information (Summerbell et al., 1973; Tickle et al., 1975).

2.2 Bicoid morphogen in *Drosophila* early embryonic development

Morphogen, as a popular genetic control topic, continues to appeal to scientists because it offers an easy way to learn the positional information and pattern formation by measuring the physical parameters, such as diffusion constant or degradation time (Grimm et al., 2010). In this thesis, we focus on Bicoid morphogen, which is among the earliest triggers of differential spatial pattern of gene expression and subsequent cell fate determination in the embryonic development of *Drosophila melanogaster*.

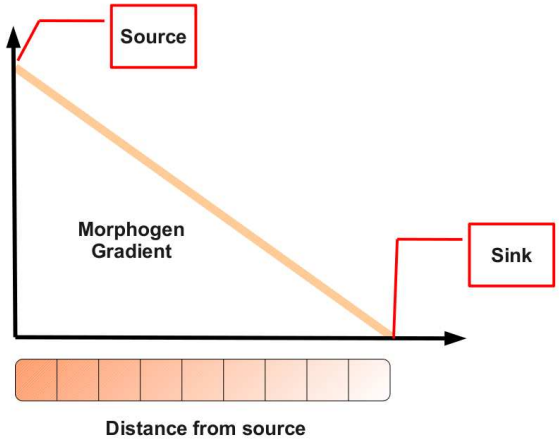


Figure 2.2: Source-Sink model proposed by Crick (1970). It is assuming that a linear morphogen gradient is formed by source and sink at the opposite ends.

Drosophila is a two-winged insect which is about 3 mm long as an adult and 0.5 mm long as an embryo. Embryonic development takes place in an egg follow by the larva hatching from the egg and going through various stages as it grows larger. The *Drosophila* was first introduced to genetic experiments by Morgan in 1909 and a Nobel Prize was given to C. Nüsslein-Volhard, E. B. Lewis and E. Wieschaus almost a century later for their contribution to the field of genetic control of early embryonic development. In order to understand the role of morphogen in developmental biology, we study the early development in *Drosophila* in this thesis because the genetic basis of *Drosophila* is well understood and similar to that of many animals.

In the *Drosophila* embryo, the most important process after fertilisation is cleavage. There are 14 cleavage cycles that occur during the first three hours in early development. The first ten cycles take place within almost 80 min after fertilisation lasting 8 min each. Being different from other animal embryos, there is no cytoplasm cleavage in this early stage of *Drosophila* and the mitotic divisions take place in the zygote nucleus without cell walls forming. The proteins diffusion becomes much easier in this shared cytoplasm due to the lack of cell membranes. After 10 nuclear divisions, the embryo develops as a multinuclear syncytium which has 5000 nuclei surrounded by the common cytoplasm. The duration for each cycle from 10 to 13 is around 10 min and Cycle 14*A* lasts 50 min (Foe et al., 1993; Poustelnikova et al., 2004). After 3 hours, the cellular blastoderm develops. Although several genes are involved in *Drosophila* early embryonic development, we only focus on a few genes which have significant effects such as *bicoid*. Bicoid protein, as the first morphogen known molecularly, is a maternally supplied transcription factor which plays an important role in establishing the embryonic anterior-posterior (A-P)

axis during early development of the *Drosophila* embryo (Driever and Nüsslein-Volhard, 1988a,b; Ephrussi and Johnston, 2004). Bicoid gradient is required for both head and thorax development of the embryo, which will develop as duplicated tail structures if the gradient is lacking (Frohnhöfer and Nüsslein-Volhard, 1986). The anterior structures of the embryos, whose mothers lack the *bicoid* gene, can not be determined; however, with the injection of *bicoid* mRNA, some anterior regions of these embryos are developed at the site of injection (Driever et al., 1990).

bicoid maternally provided mRNA, is synthesised in the nurse cells and transported to the oocyte. It is strictly localised at the anterior pole of oocytes (Berleth et al., 1988; St Johnston et al., 1989). After fertilisation, *bicoid* mRNA is translated to Bicoid proteins which will diffuse from the localised source and form a concentration gradient along the A-P axis (Driever and Nüsslein-Volhard, 1988a). After establishing such a gradient, Bicoid, as a transcription factor, binds to the enhancers of target genes and induces following gene expression.

2.2.1 Deterministic Bicoid reaction-diffusion computational model

Driever and Nüsslein-Volhard (1988a) showed that Bicoid proteins set up an exponential decay concentration gradient with maximum in the anterior part which spans the anterior two-thirds of the egg's length. This distribution model involves three parts: synthesis, diffusion and degradation.

As previously stated, the Bicoid gradient is established during the syncytium stage where the nuclei replicate in a common cytoplasm without cell division. Therefore, the morphogen gradient establishment becomes much easier. This diffusive concentration has the property that it can determine cell fate along the A-P axis in the concentration-dependent manner (Driever and Nüsslein-Volhard, 1988b). Many computational and experimental works have been published towards an understanding of the precision with which spatial boundaries are established and the scaling behaviour of the concentration gradients have been analysed (Houchmandzadeh et al., 2002; Gregor et al., 2005, 2007a,b; Holloway et al., 2006; Bergmann et al., 2007; Porcher and Dostatni, 2010; Löhr et al., 2010; Okabe-Oho et al., 2009; He et al., 2010).

To a first approximation, a one-dimensional model is regarded as adequate, as several authors have considered. The Bicoid spatio-temporal reaction-diffusion system is given by:

$$\frac{\partial m(x,t)}{\partial t} = D\frac{\partial^2}{\partial x^2}m(x,t) - \tau_p^{-1}m(x,t) + S(x,t) \tag{2.2}$$

Unlike Crick (1970)'s Source-Sink model, in which the 'sink' is at the other end of the embryo and destroys molecules, Equation 2.2 shows the Bicoid reaction-diffusion model with proteins diffusing along the A-P axis with diffusion coefficient D and decaying

everywhere with degradation time τ_p. $S(x,t)$ is a source function which describes the production of morphogen molecules. There is an implicit assumption of a one to one mapping between mRNA regulation and the corresponding protein production, which we believe is justified as there is no evidence available of non-linear feedback in this developmental system (Gregor et al., 2007a). $m(x,t)$, with an exponential decaying profile, is formed by localised synthesis, diffusion and spatially uniform degradation and refereed to as the widely used SDD model (Driever and Nüsslein-Volhard, 1988a; Houchmandzadeh et al., 2002; Gregor et al., 2005, 2007b; Bergmann et al., 2007).

When the protein degradation time is short, there is a possibility that the protein synthesis rate is balanced by degradation and the reaction-diffusion system reaches steady state. In this stage, the Bicoid concentration profile remains stable and the rate of protein changing over time becomes zero: $\partial m(x,t)/\partial t = 0$. As the localised constant source being considered as boundary condition, we rewrite Equation 2.2 as

$$D\frac{d^2}{dx^2}m_s(x) - \tau_p^{-1}m_s(x) = 0. \tag{2.3}$$

The solution to this homogeneous second-order linear equation is given by:

$$m_s(x) = C_1 \exp\left(-\frac{x}{\lambda}\right), \tag{2.4}$$

with $\lambda = \sqrt{D\tau_p}$ and $C_1 = \frac{J\lambda}{D}$ (derivation is shown in Appendix A.2). C_1 is the morphogen concentration at the anterior pole where $x = 0$. λ is the decay length and the concentration will decrease to $C_1 \exp(-1)$ when $x = \lambda$. This solution shows that the morphogen profile, given a position x, is only dependent on diffusion constant and protein life time at steady state.

The concept of steady state emerged from the field of mathematics and has been applied in many other fields. In the biological reaction-diffusion system, the general idea of the steady state is that the protein production compensates degradation equally and the concentration remains stable with developmental time increasing. Most Bicoid studies are based on steady state by assuming that downstream genes will express after the morphogen concentration becomes stable. (Reinitz et al., 1995; Aegerter-Wilmsen et al., 2005; Houchmandzadeh et al., 2005; Jaeger et al., 2004b). However, in the quantitative measurements, by changing the Bicoid dosage, Houchmandzadeh et al. (2002) revealed that downstream gene *hunchback* (*hb*) undergoes a shift which is much more smaller than the expected one, which is evaluated using morphogen steady state model. This simple morphogen model needs to be questioned concerning whether it is sufficient to explain the fate map of downstream gap genes. Later, this findings was extended by Gregor et al. (2007a), who found that Bicoid concentration gradient drops 10% between neighbouring nuclei, independent of the location along the A-P axis and such a gradient is readout precisely by the target downstream gap gene *hunchback* with a sigmoidal dependence. It is worth asking how the nuclei detect the drop reliably. To address this

question, a possible strategy proposed by Gregor et al. (2007a) is the communication between the nearby nuclei to estimate Bicoid concentration jointly. This is different from (Houchmandzadeh et al., 2002), in which the authors assumed other maternal genes may be involved during interpretation. Although both findings imply the probable mechanisms to control the variation of gene expression, however, the capability of the traditional Bicoid steady state system to provide precise downstream gene expression is still being explored.

Additionally, in biological experiments, a real steady state may take too long time to be observed if the protein lifetime τ_p is long and the tissue patterning may occur earlier before the steady state (Grimm et al., 2010). By simulating the Sonic hedgehog homolog (Shh) morphogen formation numerically, Saha and Schaffer (2006) provided a different perspective that the cell fate can be determined when the morphogen increases over the threshold, instead of reaching steady state.

Based on the query relating to the Bicoid steady state, Bergmann et al. (2007) proposed a pre-steady state decoding hypothesis that gap gene decoding occurs earlier than morphogen gradient steady state establishing. The small shifts in gap and pair-rule gene expression domains are observed when the Bicoid dosage changes and the authors found that these shifts of gap genes are dependent on their positions to the source. Along the A-P axis, the genes in the posterior part are less sensitive to Bicoid dosage. This is inconsistent with the traditional assumption that the decoding takes place after the Bicoid gradient reached steady state. This hypothesis is also supported by the experimental evidence in their work. Figure 2.3 shows the Bicoid gradient shifting comparison between steady state and pre-steady state models with the Bicoid dosage varying. Such a shift becomes insensitive in the posterior part of the pre-steady state model (Figure 2.3D), which is consistent with the measurements. This work also implies that gap gene expressions defined by this transient pre-steady-state could reduce patterning errors caused by fluctuations in the rate of morphogen production.

At the same time, the Bicoid concentration gradient formation is also questioned by Gregor et al. (2007b), who characterised Bicoid gradient dynamics in live imaging by replacing the endogenous Bicoid with enhanced green fluorescent protein (eGFP) in living transgenic *Drosophila* embryos. The authors found that Bicoid gradient forms rapidly within about 90 min after fertilisation, which is inconsistent with the measurement of diffusion coefficient D, where $D = 0.3\,\mu m^2/s$. For example, a gradual morphogen gradient has been found during the experiment, with length scale λ more than $60\,\mu m$ at 90 min after fertilisation; however, with the relatively slow diffusion obtained from measurement, the expected length scale is around $40\,\mu m$. In Gregor and colleagues' finding, the traditional Bicoid reaction-diffusion model has been questioned because the low mobility of the Bicoid protein is insufficient to establish a stable gradient within 90 min. A better understanding of how Bicoid gradient forms is still being exploited. Besides the pre-steady state decoding framework proposed by Bergmann et al. (2007, 2008), an

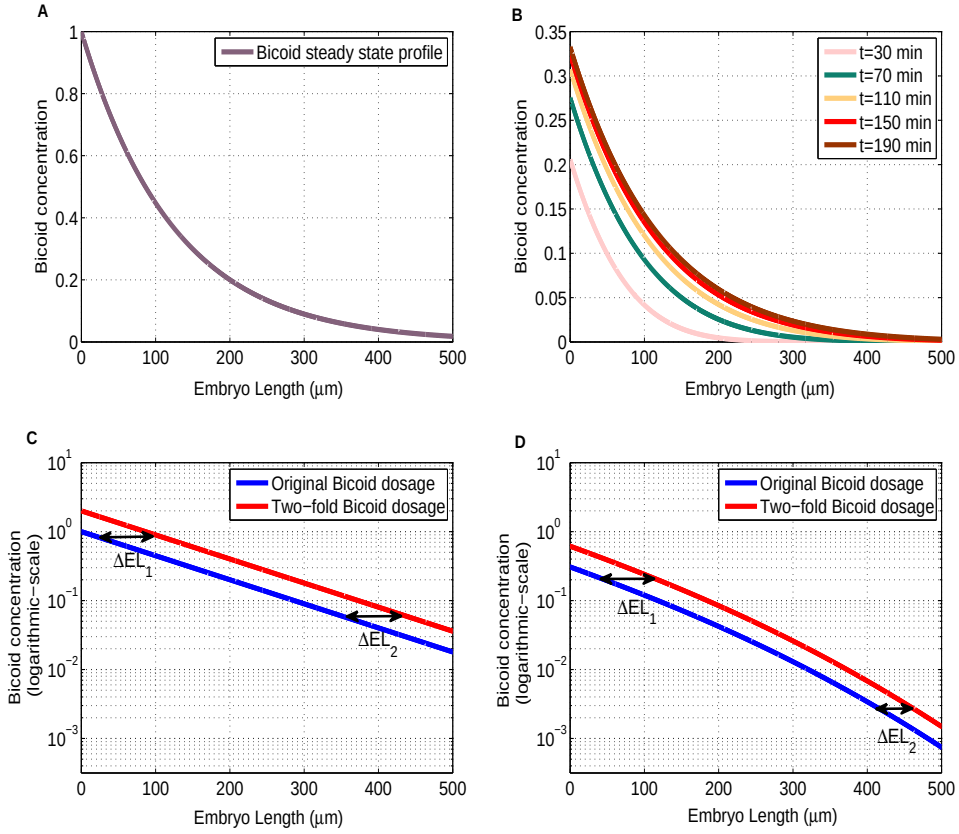


Figure 2.3: Bicoid steady state and pre-steady state profiles with Bicoid dosage changing. **A** shows traditional exponential steady state profile described in Equation 2.4. The pre-steady state model from Bergmann et al. (2007)'s work has been reproduced in **B** with developmental time increasing. The analytical derivation is shown in Appendix A.1. The Bicoid gradient shifts of these two models, caused by dosage changing, are shown in **C** (steady state) and **D** pre-steady state logarithmically.

alternative mechanism based on an additional cytoplasmic flow term proposed by (Hecht et al., 2009) showed how such a disagreement may be addressed when the flow velocity has been included. It was believed that the cytoplasmic flow is caused by the viscous cytoplasm dragged from nuclei motion due to axial expansion and cortical migration. This was later argued by Von Dassow and Schubiger (1994) who showed that the cytoskeletal forcing drives the flow. The fountain steaming pattern is resulted by deep cytoplasm flowing rapidly from centre of the embryo toward the poles and peripheral cytoplasm flowing toward the middle region of the embryo.

Alternatives to passive deterministic diffusion from a point source at the anterior end have been considered by some recent authors. Coppey et al. (2007, 2008) proposed such a mechanism for Bicoid gradient establishment based on the idea of nuclear trapping. Their model explicitly accommodates the growth in the number of nuclei in the embryo,

and the shuttling of Bicoid molecules in and out of nuclei at each cycle. This mechanism, in essence, serves as a substitute for degradation of the morphogen molecules assumed in other models. A more recent model was described by Kavousanakis et al. (2010) who considered an arrangement of periodic components representing nuclei, modelling very much the same nuclear trapping aspect studied by Coppey et al. (2007, 2008). Cheung et al. (2011) have considered Bicoid production rate as a variable, i.e. the quantity of maternally deposited mRNA being a function of the size of the embryo, as an explanation of morphogen gradient scaling.

2.2.2 Spatial distributed *bicoid* mRNA model

Although several models have been proposed, the exploration of Bicoid gradient formation has not been suspended. *bicoid* mRNA, maternally provided and normally assumed localised in the anterior pole of *Drosophila* embryo, encodes Bicoid proteins which diffuse along the A-P axis (St Johnston et al., 1989, 1991). The mRNA-based approach has attracted attention recently because mRNA provides a better way to understand morphogen gradient formation.

Contradicting with the traditional SDD model proposed in (Driever and Nüsslein-Volhard, 1988a), a totally different perspective for Bicoid profile establishment in (Spirov et al., 2009) showed that Bicoid profile is formed by *bicoid* mRNA gradient which is established by the transport of mRNA along the embryo cortex. By the FISH method and confocal microscopy, the authors suggested that the inconsistency between the small diffusion coefficient and gradual protein gradient measured by Gregor et al. (2007b,a) can be addressed by this active RNA transport and synthesis (ARTS) model. This experimental observation is also numerically simulated by Dilão and Muraro (2010) who proposed a computational mRNA diffusion model, in which, mRNA has a spatial diffusion along the A-P axis and the protein gradient forms without any protein degradation.

Is the spatial distributed *bicoid* mRNA sufficient to provide protein concentration gradient during embryo development? Capturing the dynamics of *bicoid* mRNA particles by the FISH method, Little et al. (2011) found most mRNA is localised within the anterior 20% of the embryo and protein diffusion towards the posterior is still required to establish the gradient. The mRNA spatial distribution can not account for the whole protein gradient formation. Meanwhile, by evaluating the least square error between the numerical model and experimental measurements, the authors showed that the error can be reduced when the non-localised source cooperated with the SDD model. The various sources such as point and non-localised mRNA have been discussed in (Dalessi et al., 2011) and the effects on Bicoid morphogen formation have also been analysed.

2.2.3 Stochastic reaction-diffusion model

The computational chemical reaction-diffusion system, described by the deterministic differential equation, was introduced above. However, such a continuous system has limitations when the uncertainties analysis is required. Therefore, a discrete chemical reaction system, based on stochastic simulation, has been proposed and analysed by a number of studies (Gillespie, 1977; Gibson and Bruck, 2000; Andrews and Bray, 2004; Hattne et al., 2005; Erban et al., 2007; Erban and Chapman, 2009). This stochastic system provides a more detailed understanding of the chemical reactions in biological systems.

The stochastic simulation algorithm proposed by Gillespie (1977) represents exactly the distribution of the master equation and correctly accounts for the inherent fluctuations and correlations that are ignored in the deterministic system. The basic idea of stochastic simulation can be summarised as two questions: which reaction occurs next and when does it occur? Sometimes, however, this simulation is computationally expensive when the system contains many reactions. The modified method proposed by Gibson and Bruck (2000) renders the Gillespie algorithm more efficiently by only updating the changed chemical reactions at each step. A practical guide to the stochastic simulations based on the Gillespie algorithm are provided by Erban et al. (2007); Erban and Chapman (2009).

Focusing on the discrete nature of the molecular system, Wu et al. (2007) presented a probabilistic formulation, treating the embryo as a finite number of compartments, and formulating the chemical master equation for molecules undergoing transitions between them. Because it is hard to obtain the analytical solutions of the reaction diffusion master equation, numerical simulations by a software MesoRD (Hattne et al., 2005) have been used in their work. An elegant approximate inference method for such stochastic models is presented in Dewar et al. (2010), drawing on statistical physics literature. They proposed a Bayesian approach, based on formulating a Markov jump process, for estimating parameters from observational data, along with uncertainties in these estimates. While inference from such a system is usually achieved via stochastic simulations, the authors use approximate inference to circumvent the associated computational complexities. Another stochastic approach can be found in (Fomekong-Nanfack et al., 2009), in which the gap gene model was simulated stochastically and robustness of this noisy model was analysed.

2.2.4 mRNA stability regulation model

To the best of our knowledge, the computational models reviewed above assume that the *bicoid* mRNA, as a source to translate morphogen proteins, is only a constant supply. While the control of stability and translation during development has been discussed by

other authors (e.g. see review by Cooperstock and Lipshitz (1997)), these have not been included in computational models. The constant assumption is not realistic because it is common that in animal development, maternally encoded mRNAs degrade during transition from maternal to zygotic expression (Yasuda and Schubiger, 1992; Schultz, 1993).

A particularly novel insight into the process of Bicoid translation comes from the experimental work of Surdej and Jacobs-Lorena (1998). By analysing *bicoid* mRNA abundance during the first three hours of embryogenesis, the authors suggested that the stability of the maternal mRNA may be systematically regulated; i.e. kept stable for a period of time (two hours) during which mRNA is translated and morphogen synthesised, and subsequently rapidly killed off within 30 min by *bicoid* instability element (BIE). BIE is contained within the first 43 nucleotides of the $3'$ untranslated region ($3'$ UTR) on *bicoid* mRNA. This observation matches our natural expectation as there is no need for the organism to continue to produce Bicoid protein beyond the point in time when it is decoded. Another experimental finding from Spirov et al. (2009) also confirmed *bicoid* mRNA disappeared rapidly during early nuclear Cycle 14.

To our surprise, however, modelling literature over the 30 years since the discovery of Bicoid ignores this possibility and assumes a constant production rate at the anterior pole. In the recent Bicoid modelling reviews (Grimm et al., 2010; Dalessi et al., 2011), the degradation of *bicoid* mRNA still remains an open question in modelling morphogen gradient formation. We showed recently in (Liu and Niranjan, 2009, 2011) that it is possible to model Bicoid production in a manner similar to Surdej and Jacobs-Lorena (1998)'s experimental findings and computationally extract the time at which mRNA decay begins, and the rate at which it is killed off, to match data measured on real fly embryos and archived in `FlyEx` database (Pisarev et al., 2009). An explicit model of mRNA stability regulation and a least squares fitting procedure between model output and observed data are shown in (Liu and Niranjan, 2011).

2.3 `FlyEx` database

In order to understand gene segmentation expression in *Drosophila*, accurate quantitative datasets on gene expression are the key sources of information about the development of the organism. Meanwhile, the availability of these accurate data is of critical importance for developing new mathematical models and inferring regulatory interactions in the genetic network.

There are some databases that can provide details of gene functions and developmental processes, i.e. FlyBase (Tweedie et al., 2009; Gelbart et al., 2003) which provides *Drosophila* genetic and genomic information, such as molecular function, biological processes and subcellular location, etc. Another database, FlyMove (Weigmann et al.,

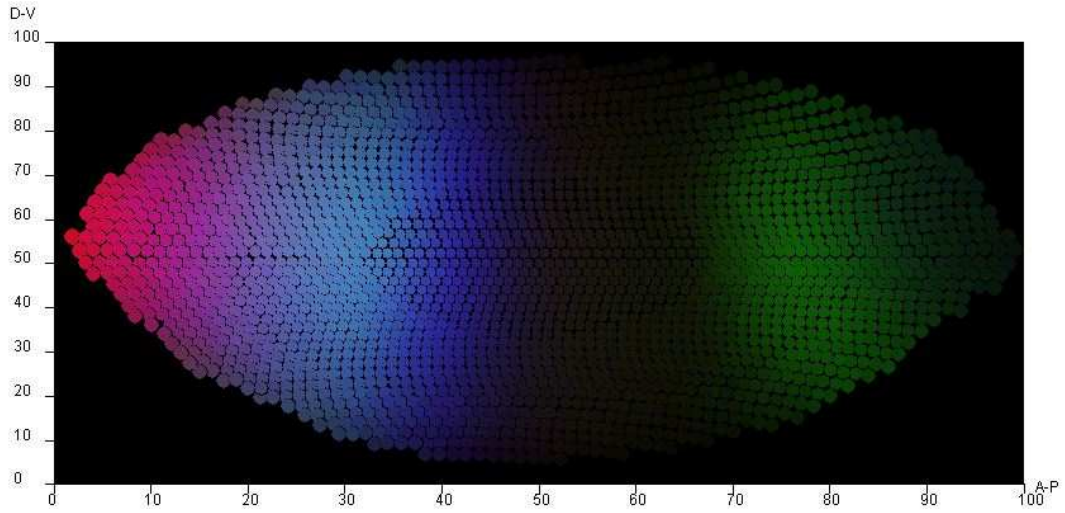


Figure 2.4: A two dimensional image of gene expression patterns: Bicoid (Pink pattern), Hb (Blue pattern) and Gt (Green pattern) from `FlyEx` (http://urchin .spbcas.ru/flyex).

2003), aims to use multimedia presentations like 3-D images to facilitate understanding of *Drosophila* development.

In this thesis, the experimental measurements we use are published in `FlyEx` (Poustelnikova et al., 2004; Pisarev et al., 2009), which provides high resolution quantitative gene expression data, extracted from confocal images on gene expression patterns of the *Drosophila* embryo. As noted above, during early embryonic development, there are 14 nuclear divisions. While there is some variability in how these developmental stages map onto real time, on average, Cycle 14A (temporal classes 1 – 8) lasts for around 50 min Foe et al. (1993). Cycle 11 starts around 100 min from fertilisation, and the three cycles 11, 12 and 13 last an average of 10 min each. In `FlyEx`, Bicoid data are available for the 11 temporal classes from cycles 11 to 14A.

The quantitative expression data of 14 genes have been published in `FlyEx`, i.e. *bicoid*, *hb* and *gt*, etc. These data are obtained from images of gene expression patterns in 1580 wild-type *Drosophila* embryos by confocal scanning microscopy of fixed embryos with immunostained proteins. All the images are obtained in 8-bit format with maximum intensity at 255. These images then need to be segmented by constructing a binary nuclear mask to determine the x and y coordinates and estimate the mean fluorescence of each nucleus. The x and y coordinates for each nucleus correspond to the A-P and D-V axes, respectively and the coordinates are expressed as a percentage of the maximum size of the embryo in x and y directions. This compensates for size differences among embryos (Poustelnikova et al., 2004). The background noise needs to be removed by data normalisation, whereby fluorescence is transformed to zero if at or below background level, and maximum possible fluorescence to itself.

Because segmentation gene expression is largely a function of position along the A-P axis in the embryo, this can therefore be represented well in one dimension. The one dimensional integrated data is extracted from the central 10% of y values on the mid-line of an embryo in the A-P direction (Myasnikova et al., 2001). The data here are grouped in 100 bins along A-P axis and averaged within each bin. The gene expression data are mapped to one row of nuclei and a single nucleus is close to 1% of the embryo length in diameter.

2.4 Gap gene regulatory network

The major developmental activity, that takes place following the establishment and interpretation of maternal morphogen Bicoid, is the formation of the gap gene expression patterns, consisting of a closely coupled network of six transcription factors cross-regulating each other. These have been studied extensively by experimentally (e.g. ChIP-chip experiments by Perry et al. (2011)), and computationally in a series of papers spanning a period of two decades, starting from Reinitz and Sharp (1995) to Jaeger et al. (2004b,a). Bicoid is known to initiate the gap gene expression by binding to the corresponding enhancers in the regulatory regions of target genes and it is considered as an input to all gap genes in the computational models.

The establishment of the gap gene formation happens immediately, about two hours after fertilisation, precisely the time duration in which translation of the maternal mRNA is switched off. Hence, we are interested in asking what effect rapidly decaying Bicoid concentration might have on the gap gene expressions. In this section, we give a review of various aspects of the gap gene network, and later in Chapter 5 present and analyse our own work on optimising the network parameters with time-varying input data taken from `FlyEx` dataset.

The gap genes, as part of segmentation-determined system, are the first zygotic genes expressed along the A-P axis and all of them are transcription factors. They are normally recognised by their mutant phenotypes, in which the contiguous body segments are missing and gaps of the body pattern are formed. The first three gap genes were identified by Nüslein-Volhard and Wieschaus (1980) by genetic screen: *hb*, *knirps (kni)* and *Krüppel (Kr)*. Further gap genes, *giant (gt)*, *tailless (tll)* and *huckebein (hkb)*, were identified in later works (Petschek et al., 1987; Weigel et al., 1990).

Initially, the maternal provided mRNAs are placed in the oocyte, where the *bicoid* and *nanos (nos)* mRNAs are localised in the two opposing ends. The located information is determined by $3'$ UTR of each gene (Frigerio et al., 1986; Berleth et al., 1988; Ferrandon et al., 1997). The maternal *hb* and *caudal cad* mRNAs are uniformly distributed along the A-P axis (see Figure 2.5).

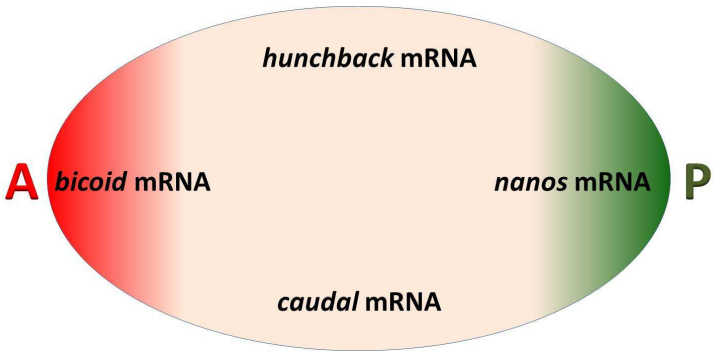
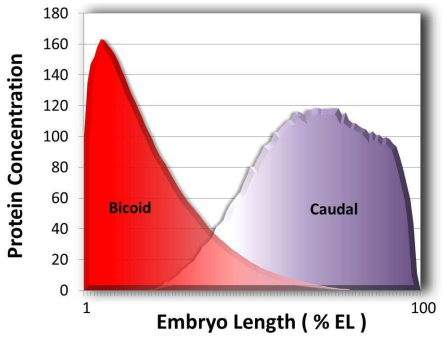
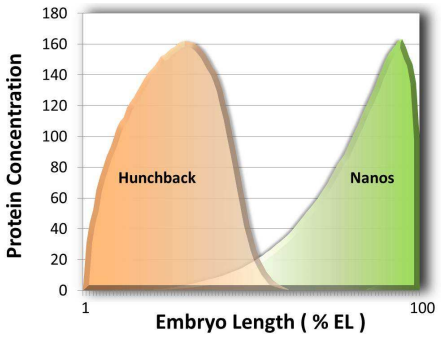


Figure 2.5: mRNA distribution during oogenesis.

After fertilisation, *bicoid* mRNA spreads to the posterior and forms a gradient within 20% of the embryo anterior part (Frigerio et al., 1986; Spirov et al., 2009). Bicoid protein, translated from *bicoid* mRNA as stated in previous sections, will bind to a specific region of 3′ UTR on *cad* mRNA and inhibits the translation of Cad protein from mRNA in the anterior regions (Chan and Struhl, 1997). This is important for embryo development since Cad will result in an improper anterior formation if it has not been repressed in this region. Similarly, Nos prevents *hb* translation in the posterior area (Tautz, 1988); this is the only contribution of Nos found during embryonic development so far (Jaeger, 2011). The connections between Bicoid ⊣[1] Cad and Nos ⊣ Hb are shown in Figure 2.6(a) and 2.6(b). All the protein concentrations are drawn at Cycle 14*A*.1 from `FlyEx` except Nos, which is not included in this database and we use mirrored Bicoid concentration to replace it.



(a) Bicoid and Cad protein distribution after fertilisation



(b) Hb and Nos protein distribution after fertilisation

Figure 2.6: Proteins distribution at Cycle 14*A*.1

Bicoid, Hb and Cad, as the earliest transcription factors during *Drosophila* embryonic development, activate or repress zygotic genes expression in the concentration-dependent manner when the embryo is a syncytium. Gap genes can diffuse away from where they

[1] ⊣ represents the repression

are synthesised in such a stage and their early localised regulation is dependent on those maternal genes.

The whole gap gene network is initialised by Bicoid. An embryo which lacks *hb* will have a severe phenotype such as deletion of gnathal and thoracic segments. Zygotic Hb, activated by Bicoid binding sites in the *hb* regulatory region (Driever and Nüsslein-Volhard, 1989), forms a gradient in the anterior regions of the embryo after Cycle 12. This protein, associated with maternal Hb, acts as a morphogen to control other gap gene expression with Bicoid. For example, *Kr* may be activated by Bicoid by molecular studies as in (Hoch et al., 1990) and it may also be activated by low level Hb (Struhl et al., 1992). *gt* is activated by Bicoid in the anterior domain and is repressed by Hb in the posterior domain (Tautz, 1988; Rivera-Pomar et al., 1995). The expression of *gt* in posterior regions is activated by another maternal gene, *Cad* (Rivera-Pomar et al., 1995). There are, however, still some problems in gap gene regulation that need to be explored. For example, it is unclear which gene controls the expression of Kr and how the Kr boundaries are defined. The evidence so far is not sufficient to characterise proper early gap gene regulations. In a recent gap gene review, Jaeger (2011) conjectured that other unknown maternal genes may still be missing to regulate the gap genes.

This later stage of development, considering the cross-regulation between gap genes, turns the framework into a much more complicated one. Some powerful computational tools need to be applied to analyse the connections between genes. Starting from the computational genetic model proposed by Reinitz and Sharp (1995), Jaeger et al. (2004b,a) built a mathematical gap gene network during cycles 13 and 14A based on quantified gene expression data from `FlyEx` to reveal the connections between such gap genes. A linear ODE system has been established within the main embryo body axis from 35% $\sim$ 92% of embryo length with six genes, *hb*, *cad*, *tll*, *kni*, *Kr* and *gt*. *bicoid* gene expression data, as an external input in this system, are averaged as constant. The key role of this model may explain the gap gene patterns anterior shift, caused by the asymmetrical repression. An improved gap gene model included terminal gap gene *hkb* was proposed in (Ashyraliyev et al., 2009). The main idea of this computational model remains the same in this work but with a smaller model size, in which there are four regulated genes (*hb*, *kni*, *Kr* and *gt*) and four external inputs (*bicoid*, *cad*, *tll* and *hkb*).

The gap genes connections in the mathematical models are described by the model parameters, of which the maximum number is 62 in (Jaeger et al., 2004b) and 44 in (Ashyraliyev et al., 2009). An global optimisation method parallel lam simulated annealing (PLSA) (Lam and Delosme, 1988a,b) has been used to fit the model to quantitative data in both of the models. Such a fitting process took between 8 and 160 CPU hours on 10 parallel 2.4-GHz Pentium P4 Xeon processors. The computational cost is dramatically decreased in (Fomekong-Nanfack et al., 2007) by applying another global approach – evolution strategy combined with a local search strategy, by which the computational time is around 8 to 11 CPU hours, much faster than PLSA.

2.5 Gaussian processes in systems biology

2.5.1 Bayes' theorem

The Bayes' rule, yielding the posterior density, is given by:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{2.5}$$

where $p(\theta)$ gives the prior distribution of θ being observed and likelihood is given by $p(y|\theta)$. The marginal likelihood $p(y)$, defined as $\sum_\theta p(y|\theta)p(\theta)$ for discrete θ and $\int p(y|\theta)p(\theta)d\theta$ for continuous θ, is independent of θ. Hence, $p(y)$ can be considered as a constant and Equation 2.5 becomes

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{2.6}$$

* *Prediction.* Following Bayes' rule above, we can infer the unknown observable y_* from the given data y with parameter θ. The posterior predictive distribution of y_* is given by:

$$p(y_*|y) = \int p(y_*, \theta|y)d\theta = \int p(y_*|\theta, y)p(\theta|y)d\theta \tag{2.7}$$

* *Likelihood.* As can be seen in Equation 2.6, the data y affect the posterior distribution only through $p(y|\theta)$, known as likelihood function. Generally, such a function is from the exponential family (Duda and Hart, 1973; Bernardo et al., 1994)

* *Prior.* The widely used Gaussian distribution prior, is a specific example of the conjugate prior, in which the posterior distribution has the same form as the prior. Moreover, the conjugate prior exists for any member of the exponential family.

2.5.2 Gaussian processes

In biological systems, given the observations, we can never know the underlying true values due to the noise. In machine learning, Bayesian approaches, as powerful tools, can deal with the uncertainties probabilistically. Bayesian inference has been shown to be useful in a range of applications including systems biology and bioinformatics. Successful examples range from the identification of genetic regulatory networks (Pe'er et al., 2001; Husmeier, 2003; Perrin et al., 2003; Beal et al., 2005) and causal protein-signaling networks (Sachs et al., 2005) by dynamic Bayesian networks, inference of transcription regulation using a state space model (Sanguinetti et al., 2006a,b), Bayes factors estimating by MCMC and thermodynamic integration methods in non-linear ODE models (Calderhead and Girolami, 2009), approximate inference using variational methods for

the spatio-temporal Bicoid system (Dewar et al., 2010), to parameter estimation using Monte Carlo simulations to understand stochastic dynamics of bacterial gene regulation (Wilkinson, 2011).

Gaussian processes (GPs) (Rasmussen and Williams, 2006) provide an effective framework for inferring the action of latent functions from partially observed data and learning the parameters from the model. In particular, such a framework has been successfully applied to infer unobserved chemical species of tumour repressor $p53$ in (Lawrence et al., 2007; Gao et al., 2008), which provides an efficient alternative to the Markov chain Monte Carlo (MCMC) approach advanced earlier by Barenco et al. (2006). GPs have been applied to infer the transcription factor given mRNA concentrations because it is very hard to measure the active concentration of transcription factor proteins which drive the process and the sensitivity of target genes to these concentrations. This non-parametric Bayesian probabilistic inference methodology has also been used in (Honkela et al., 2010) to identify potential targets of a transcription factor using time series expression data.

A GP is defined as

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x')), \tag{2.8}$$

where the mean and covariance functions are given by:

$$\mu(x) = E(f(x)), \quad \text{and} \tag{2.9}$$

$$k(x, x') = E((f(x) - \mu(x))(f(x') - \mu(x'))). \tag{2.10}$$

The latest software packages to implement GPs can be found on `The Gaussian Processes Web Site`[2]. Generally, the packages such as `NETLAB`[3] (Nabney, 2001) and `BUGS`[4] (Lunn et al., 2009) have made Bayesian inference easy to access for practitioners.

In line with previous works (Gao et al., 2008), in this thesis, we use the squared exponential covariance function (radial basis function (RBF)) to define the covariance between pairs of two input points:

$$k(x_i, x_j) = \text{cov}(f(x_i), f(x_j)) = \sigma_r^2 \exp\left(-\frac{(x_i - x_j)^2}{l^2}\right), \tag{2.11}$$

and to generate smooth function $f(x)$. This covariance function varies inversely with the distance between two inputs and becomes almost unity when two time points are very close. The σ_r^2 and length scale l are the parameters in the squared exponential covariance function, where l is constant for all inputs. In general, the RBF is a smooth kernel, which produces smooth functions.

[2] http://www.gaussianprocess.org/
[3] http://www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/
[4] http://www.mrc-bsu.cam.ac.uk/bugs/

Apart from Dewar et al. (2010)'s work, Bayesian methods mentioned above address purely temporal phenomena. In the non-biological spatio-temporal system, GPs have been applied substantially in spatial analysis as a prediction scheme, such as kriging (Cressie, 1993; Stein, 1999). As a powerful modelling example, kriged kalman filter proposed by Mardia et al. (1998) provides spatio-temporal prediction of environmental problems. Another successful application proposed by Alvarez et al. (2009) describes a attempt at combining a data driven model (GP) with a latent process explaining the known physics of a system. The authors consider the heat equation, which is a simplified form of a diffusion system, to model the spatio-temporal profile of pollution.

Modelling biological spatio-temporal system using GPs has not been addressed in literature so far. In (Liu and Niranjan, 2012), we focused on the inference of *bicoid* mRNA regulation, which is an unobserved time-varying function in Bicoid gradient establishment, and placed a Gaussian prior distribution over it. Moreover, we extended the previous GP works into the spatio-temporal field by implementing the inference for mRNA regulation and spatio-temporal Bicoid concentration given simulated model output and real datasets from `FlyEx` database (Pisarev et al., 2009). This work will introduced in Chapter 4.

Chapter 3

Establishing Bicoid gradient with regulated mRNA stability

In this chapter, we introduce the biological spatio-temporal reaction-diffusion system, starting from Fick's law in Section 3.1.1. As explained in Section 2.2, it is noted that the assumption of constant Bicoid proteins supply is unrealistic because the maternal mRNA is known to decay after a certain time following fertilisation (Surdej and Jacobs-Lorena, 1998). In Section 3.1.2, we incorporate a realistic model of the morphogen source, in which the stability of *bicoid* mRNA is regulated. We explicitly regulate *bicoid* mRNA stability as a constant supply followed by exponential decay and solve the reaction-diffusion equation numerically for morphogen propagation. In Section 3.1.3, this work is extended to further show such mRNA regulation framework combined with a recently published flow model (Hecht et al., 2009) that takes into account a cytoplasmic flow term. Moreover, a Bicoid stochastic model based on Gillespie algorithm (Gillespie, 1977; Erban et al., 2007) which includes regulated mRNA, is shown in Section 3.2. By minimising the squared error between model outputs and measurements published in `FlyEx` database, we will show in Section 3.3 how parameters of diffusion rate, protein and mRNA degradation times, and the onset of maternal mRNA decay can be assigned sensible values. In line with recent thinking on the subject (Spirov et al., 2009; Little et al., 2011), we will also analyse the Bicoid gradient establishment with a spatial gradient of maternal mRNA, rather than being fixed at only the anterior pole.

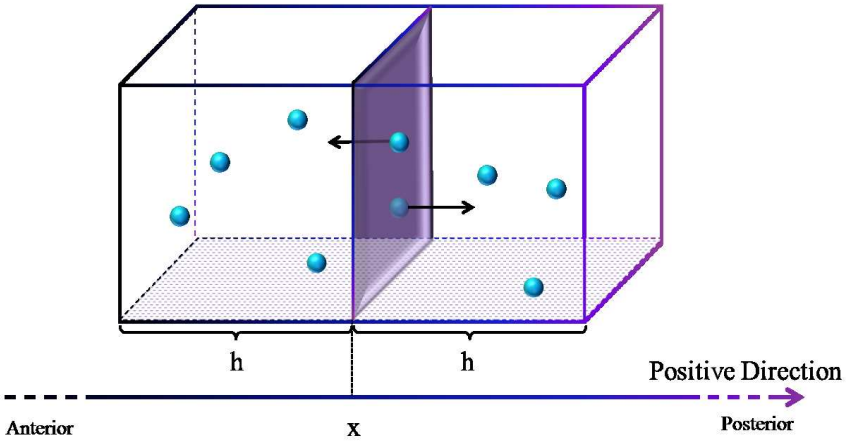


Figure 3.1: Diagram of Fick's first law. The Bicoid molecules, shown as blue spheres, diffuse between two cubes along the A-P axis in the *Drosophila* embryo.

3.1 Bicoid deterministic reaction-diffusion model

3.1.1 Fick's law

The diffusion of Bicoid molecules in the early development of the *Drosophila* embryo is introduced in this section. The establishment of the protein concentration gradient involves three processes: diffusion so that molecules can move around between nuclei, degradation of molecules themselves and production of molecules from maternal provided mRNA. Generally, the Bicoid reaction-diffusion computational model is represented by a second-order PDE, which is a reaction-diffusion equation to simulate Bicoid protein concentration gradient establishment. Let a spatio-temporal function $m(x,t)$ be defined as the Bicoid protein concentration at position x ($x \in [0, L]$) along embryo A-P axis and developmental time t ($t \in [0, T]$). We start from the basic concept of diffusion which is explained by Fick's law, derived by Fick (1855).

3.1.1.1 Fick's first law

Figure 3.1 shows a representation of two adjacent regions of an embryo across which Bicoid diffusion is illustrated. For the scenario we consider, i.e. morphogen propagation in the embryo, diffusion is considered an acceptable model since cell walls around each nuclei have not formed yet, as introduced in Chapter 2.2. The length of each cube side, h, is as small as the diameter of a nucleus and the concentrations at those two cubes can be represented as $m(x - \frac{h}{2})$ and $m(x + \frac{h}{2})$. The number of molecules in each cube at time t is given by $m(x \mp \frac{h}{2}, t)h^3$.

The transport of molecules, without any degradation and production considered firstly, is defined by a flux term J, which describes how many molecules move from regions

of high to low concentration per unit area and unit time. Axis x stands for a one-dimensional A-P axis with the positive direction from anterior to posterior. Therefore, the flux is positive if molecules are going right and is negative when they are going left.

During a small time interval $(t, t+\Delta t)$, the number of molecules at position x is defined by a simplest random walk model: a random molecule starts from a point (x,t) and takes $\frac{h}{2}$ as one step length towards left or right during Δt. Since the chances of the molecule going right or left are equal, the probability of going to each direction is 0.5. That means the molecule will either be at $x-\frac{h}{2}$ or $x+\frac{h}{2}$ at time $t+\Delta t$. Assuming that all of the molecules along the A-P axis are based on the random walk model, the number of molecules crossing the area at $(x, t+\Delta t)$ is given by:

$$\underbrace{\frac{1}{2}m\left(x-\frac{h}{2},t\right)h^3}_{A\rightarrow P}\ \underbrace{-\frac{1}{2}m\left(x+\frac{h}{2},t\right)h^3}_{A\leftarrow P}. \tag{3.1}$$

Arrows represent the directions of the molecules and the minus means the molecules are moving from the right cube to the left one.

The flux $J(x,t)$, defined as a number of molecules (Equation 3.1) cross the unit area at x over unit time, is given by:

$$\begin{aligned} J(x, t\rightarrow t+\Delta t) &= -\frac{1}{\Delta t}\frac{1}{h^2}\left[\frac{1}{2}m\left(x+\frac{h}{2},t\right)h^3-\frac{1}{2}m\left(x-\frac{h}{2},t\right)h^3\right] \\ &= -\frac{h^2}{2\Delta t}\frac{\left[m\left(x+\frac{h}{2},t\right)-m\left(x-\frac{h}{2},t\right)\right]}{h}. \end{aligned} \tag{3.2}$$

As Δt and h approach zero, the limit of Equation 3.2 is given by:

$$\begin{aligned} J(x,t) &= -\frac{h^2}{2\Delta t}\frac{\partial}{\partial x}m(x,t) \\ &= -D\frac{\partial}{\partial x}m(x,t), \end{aligned} \tag{3.3}$$

where D $(\mu m^2/s)$ is the diffusion coefficient and Equation 3.3 is Fick's first law (Fick, 1855).

3.1.1.2 Fick's second law

Fick's second law, shown in Figure 3.2, describes the variation of the number of molecules in one cube (h^3) during the time interval $(t, t+\Delta t)$. This change, counting the incoming and outgoing molecules at each end of the cube, is defined by the flux (Equation 3.3) at $x-\frac{h}{2}$ and $x+\frac{h}{2}$:

$$h^3[m(x,t+\Delta t)-m(x,t)] = h^2\Delta t\left[J\left(x-\frac{h}{2},t\right)-J\left(x+\frac{h}{2},t\right)\right]. \tag{3.4}$$

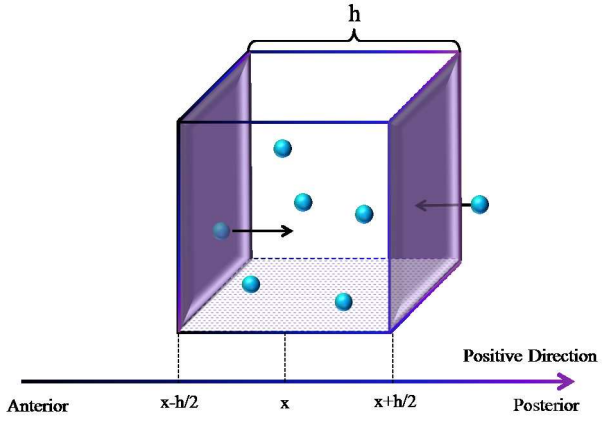


Figure 3.2: Diagram of Fick's second law, which describes the concentration changes in one cube.

Being divided by the volume of the cube (h^3) and duration Δt, Equation 3.4 becomes:

$$\frac{m(x,t+\Delta t)-m(x,t)}{\Delta t} = -\frac{J\left(x+\frac{h}{2},t\right)-J\left(x-\frac{h}{2},t\right)}{h}. \tag{3.5}$$

Let Δt and h approach zero, the limit of Equation 3.5 which is Fick' second law is given by:

$$\begin{aligned}\frac{\partial}{\partial t}m(x,t) &= -\frac{\partial}{\partial x}J(x,t)\\ &= D\frac{\partial^2}{\partial x^2}m(x,t).\end{aligned} \tag{3.6}$$

An alternative way of deriving the diffusion equation is to consider the Divergence Theorem, also known as Gauss's Theorem, which equates surface integrals and volume integrals (Morse and Feshbach, 1953). Let $\Omega = h^3$ be the volume of the cube and $\partial\Omega$ be the boundary of this region. Combining Divergence Theorem, the total out-flux in this region is:

$$\int_{\partial\Omega} \overrightarrow{J}(x,t)\cdot\overrightarrow{n}(x)dx = \int_{\Omega} \mathrm{div}(\overrightarrow{J}(x,t))dx \tag{3.7}$$

where $\overrightarrow{n}(x)$ is unit outward normal direction at x. The rate of the number of molecules changing is given by:

$$\begin{aligned}\frac{\partial}{\partial t}\int_{\Omega} m(x,t)dx &= -\int_{\partial\Omega} \overrightarrow{J}(x,t)\cdot\overrightarrow{n}(x)dx\\ &= \int_{\Omega} \mathrm{div}\left(D\frac{\partial}{\partial x}m(x,t)\right)dx.\end{aligned} \tag{3.8}$$

Rewriting Equation 3.8, we get the diffusion equation as:

$$\begin{aligned}\frac{\partial}{\partial t}m(x,t) &= \text{div}\left(D\frac{\partial}{\partial x}m(x,t)\right)\\ &= D\frac{\partial^2}{\partial x^2}m(x,t).\end{aligned} \tag{3.9}$$

3.1.2 *bicoid* mRNA stability regulation

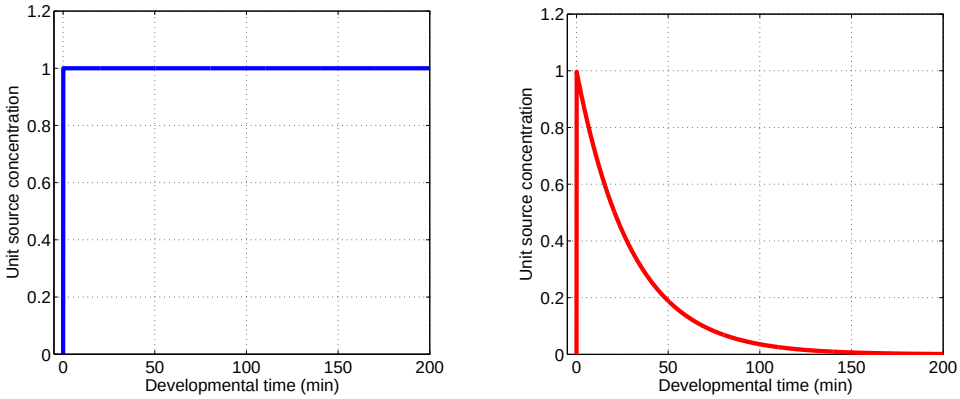
Section 3.1.1 shows how the concentration of molecules varies with diffusion. As introduced in Chapter 2.2, by combining protein reactions such as destruction and creation, the one dimensional reaction-diffusion equation used to model Bicoid concentration gradient establishing is given by:

$$\frac{\partial}{\partial t}m(x,t) = D\frac{\partial^2}{\partial x^2}m(x,t) - \tau_p^{-1}m(x,t) + S(x,t), \tag{3.10}$$

where $m(x,t)$ is the morphogen concentration as a spatio-temporal function, D, the diffusion constant and τ_p, the degradation time of the morphogen protein. $S(x,t)$, constant source at the anterior pole of embryo. Equation 3.10, which involves localised constant protein synthesis, diffusion and linear degradation, sets up an exponentially decaying Bicoid gradient along the embryo A-P axis.

Although the variations in gap gene expression may be reduced according to the stable morphogen profile, inconsistencies between observed and expected results appear and the traditional morphogen gradient model has been called into question. Gregor et al. (2007b) measured Bicoid diffusion coefficient D and found that with this small value, the observed morphogen profile can not be achieved, and faster protein diffusion is required. Several morphogen models have been proposed towards this contradiction, i.e. a pre-steady state decoding of gap gene patterning (Bergmann et al., 2007, 2008), a modified reaction-diffusion model with cytoplasmic streaming (Hecht et al., 2009), and an active RNA transport and synthesis (ARTS) model incorporating mRNA spatial distribution (Spirov et al., 2009).

To the best of our knowledge, however, all the computational models assume that the translation of maternal mRNA takes place at a constant rate in the anterior pole of the *Drosophila* embryo, resulting in a constant supply of morphogen to diffuse in the system. While mathematically convenient, in that it leads to easy closed-form solutions, this is an unrealistic assumption, for there is no need for the embryo to continue to maintain a constant supply of morphogen beyond that is needed for downstream decoding. In this chapter, we focus on how Bicoid propagates when the *bicoid* mRNA stability is regulated.



(a) Constant mRNA source model (b) Unregulated mRNA decay model

Figure 3.3: Two different approaches modelling the source concentrations which are provided by maternal deposited *bicoid* mRNA. Figure 3.3(a) shows constant source which supplies proteins over the whole time scale ($0 \sim 200\,\text{min}$). Another source model, in which mRNA is decaying from the very beginning without any regulation is shown in Figure 3.3(b)

3.1.2.1 *bicoid* mRNA without regulation

The usual assumption in solving the Bicoid reaction-diffusion model is that the protein synthesis process is continuous, which is shown in Figure 3.3(a):

$$S_{\text{con}} = S_0 \delta(x) \Theta(t), \tag{3.11}$$

where S_0 is the source concentration, $\delta(x)$ is the Kronecker delta function and $\Theta(t)$ is Heaviside step function.

With the constant protein synthesis, the Bicoid morphogen reaction-diffusion system is given by:

$$\frac{\partial}{\partial t} m(x,t) = D \frac{\partial^2}{\partial x^2} m(x,t) - \tau_p^{-1} m(x,t) + S_{\text{con}}(x,t). \tag{3.12}$$

Bergmann et al. (2007) provided the analytical solution as

$$m(x,t) = \frac{S_0}{2\beta D} \left(\exp(-\beta x) - \frac{\exp(-\beta x)}{2} \text{erfc}\left(\frac{2\beta D t - x}{\sqrt{4Dt}} \right) - \frac{\exp(\beta x)}{2} \text{erfc}\left(\frac{2\beta D t + x}{\sqrt{4Dt}} \right) \right), \tag{3.13}$$

where β is:

$$\beta = 1/\sqrt{D\tau_p}. \tag{3.14}$$

More derivation details can be found in Appendix A.1.

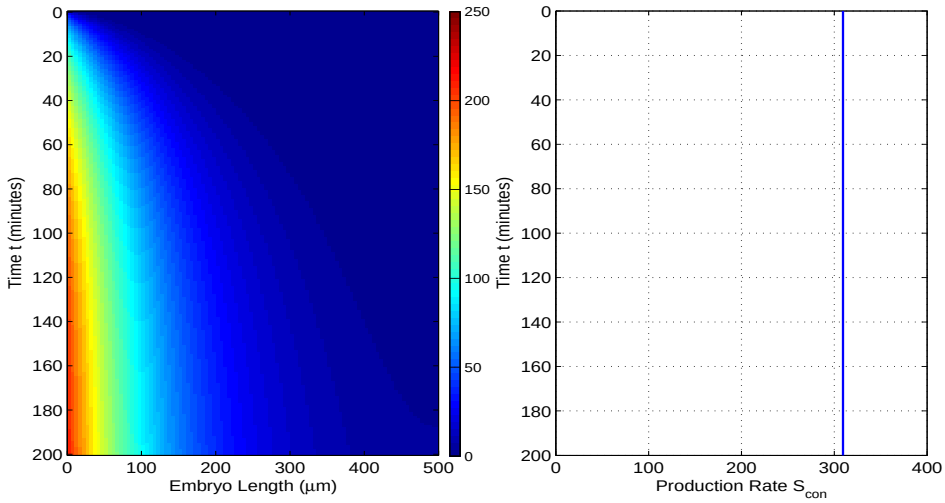


Figure 3.4: The spatio-temporal Bicoid morphogen profile with constant mRNA supply, time scale from $t = 0 : 200\,\text{min}$ and embryo length from: $L = 0 : 500\,\mu m$.

The left panel in Figure 3.4 shows the solution to the reaction diffusion model of morphogen propagation, where the production of proteins is a constant. We obtained this solution by numerically integrating the differential equation using the `pdepe` Toolbox in MATLAB with $D = 3\mu m^2/s$ and $\tau_p = 87\,\text{min}$. The source production rate is shown in the right panel.

3.1.2.2 Deterministic model with mRNA regulation

In this thesis, following Surdej and Jacobs-Lorena (1998)'s experimental findings that *bicoid* mRNA is kept constant in the first two hours then degrades quickly, the mRNA regulation of stability has been incorporated with morphogen propagation. As shown in Figure 3.5, such a regulation model is given by

$$S_{\text{con}-\text{dec}} = S_0\delta(x)\,[\Theta(t) - \Theta(t - t_0)] + S_0\delta(x)\,\Theta(t - t_0)\,\exp\left(-\frac{t - t_0}{\tau_m}\right), \qquad (3.15)$$

where τ_m is mRNA degradation time and t_0 is the decaying onset time. These two are key parameters for *bicoid* mRNA regulation. Unlike the analytic solution shown above, the solution of this regulated reaction-diffusion system has the imaginary part when the protein degradation rate is smaller than mRNA decaying rate due to the inverse Fourier transform (details are shown in Appendix A.1). We instead chose an alternative option of numerically integrating them with the MATLAB Toolbox `pdepe`, which is a PDE solver for both parabolic and elliptic PDEs. In `pdepe` toolbox, the approximate solutions of our parabolic PDEs system are integrated temporally by the spatial discretised ODEs.

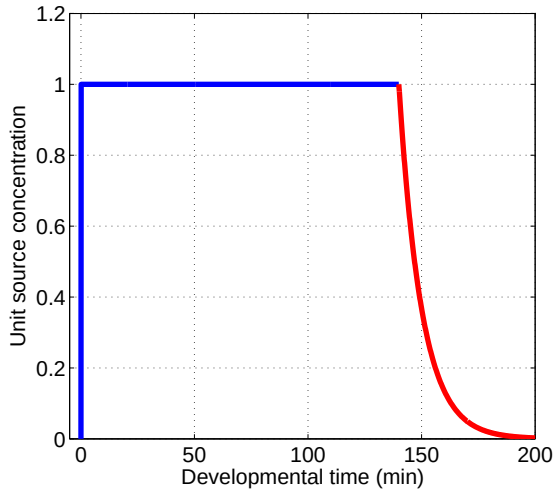


Figure 3.5: *bicoid* mRNA regulation model, which is kept stable for a period of time during mRNA being translated and subsequently rapidly killed off.

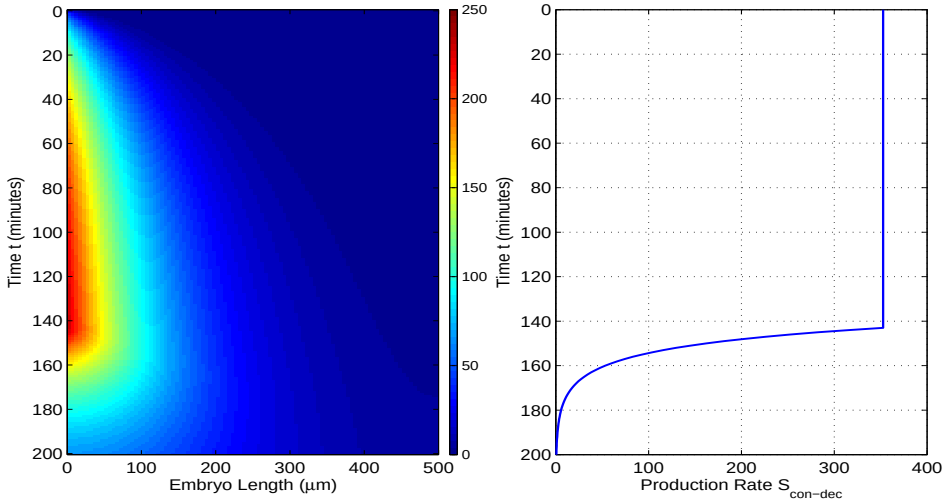


Figure 3.6: The intensity of the Bicoid spatio-temporal profile with mRNA stability regulation shown in the right panel.

Figure 3.6 shows the morphogen profile with mRNA regulation, in which the source is a combination of a constant supply followed by an exponential decay (shown on the right). As expected, the morphogen intensity in this system, evaluated numerically using the `pdepe` Toolbox, sets up a spatially decaying profile which subsequently decays to zero. For this simulation, D and τ_p are set to the same values as above while the decay rate of maternal *bicoid* mRNA was set as $\tau_m = 9\,\text{min}$ which is much smaller than τ_p. mRNA degradation was set to start at 143 min (More details of the model parameter estimation are discussed in Section 3.3).

3.1.3 Deterministic model with cytoplasmic flow

As reviewed in Chapter 2.2, the alternative morphogen propagation model with an additional cytoplasmic flow term is proposed by Hecht et al. (2009). It is motivated by the argument that, with passive diffusion, the quantitative properties of the morphogen profiles establishment require higher values of diffusion constant than have been experimentally measured Gregor et al. (2007b).

The one-dimensional flow model is given by:

$$\frac{\partial}{\partial t}m(x,t) = D\frac{\partial^2}{\partial x^2}m(x,t) - \tau_p^{-1}m(x,t) - V\frac{\partial}{\partial x}m(x,t) + S(x,t), \tag{3.16}$$

where V is fluid flow velocity. In the original formulation of this model, the flow term was permitted to be active only for a short duration in time, nuclear cleavage cycles 4 to 6, depending on the motion of the nuclei in the viscous cytoplasm. In our implementation, we allowed this term to be present throughout the developmental time period considered, to increase its difference from the standard diffusion model.

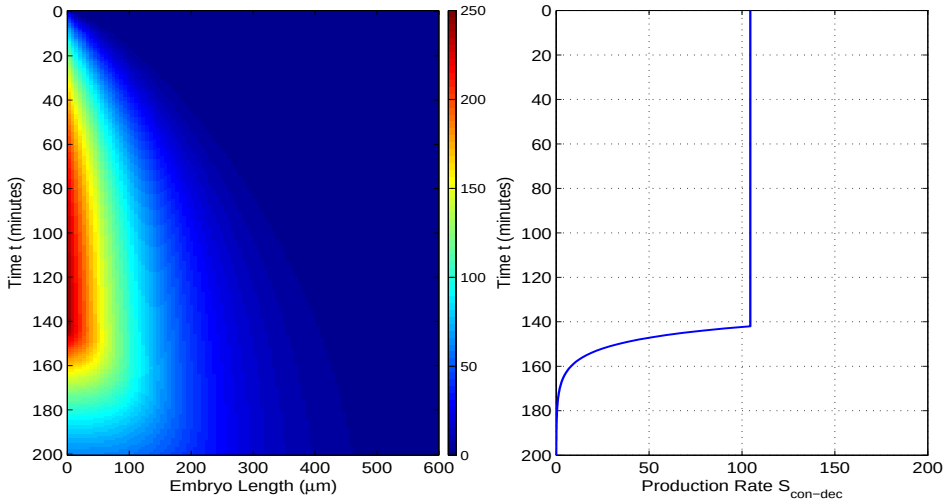


Figure 3.7: The spatio-temporal Bicoid reaction-diffusion model profile including cytoplasmic streaming, time scale from: $t = 0 : 200$ and embryo length from $L = 0 : 500\mu m$ with regulated source . The parameters here are set as: diffusion constant $D = 0.9\mu m^2/s$, Bicoid proteins decaying time $\tau_p = 42\,\text{min}$ and mRNA decaying rate $\tau_m = 7\,\text{min}$. Finally, flow velocity V is set to be $0.04m/s$.

3.2 Stochastic reaction-diffusion model

In the previous section, we have discussed the deterministic Bicoid reaction-diffusion model with mRNA stability regulation. In order to analyse the dynamical behaviour of

Bicoid molecules cooperated with mRNA regulation in this section, we model the Bicoid reaction-diffusion process based on the Gillespie stochastic algorithm, which provides a more detailed and precise picture of the molecule interactions. Due to the intractability of the analytical solution in the stochastic systems, some software packages have been used to implement stochastic reaction-diffusion simulation numerically (Andrews and Bray, 2004; Hattne et al., 2005; Erban and Chapman, 2009). Wu et al. (2007) used publicly available software MesoRD (Hattne et al., 2005) to simulate the stochastic Bicoid molecules reaction-diffusion processes to analyse intrinsic fluctuations of the Bicoid gradient. The realisation of the Bicoid reaction-diffusion system in our work is implemented in `MATLAB` based on the Gillespie algorithm (Gillespie, 1977) which is introduced in detail in Section 3.2.1.

3.2.1 Stochastic simulation

3.2.1.1 When does the next reaction occur?

Consider an example of a single chemical reaction:

$$m \xrightarrow{k} \emptyset, \tag{3.17}$$

where m is the chemical species we are interested in. k is degradation rate which describes the speed of the molecule decaying. During a small time interval $[t, t+dt)$, the probability of a random molecule of m degradation is defined by kdt. The propensity function a which gives the probability that the reaction Equation 3.17 occurs during $[t, t+dt)$ is given by:

$$adt = m(t)kdt, \tag{3.18}$$

where $m(t)$ is the number of the molecules at time t. Now we need to ask the first question: when does the next reaction occur? The easiest way to answer this is to generate some random numbers to decide when the next reaction happens (see Algorithm 1).

Figure 3.8 shows 10 iterations based on Algorithm 1 for degradation reaction, where the initial number of molecules $m(0)$ is 30, degradation rate k is $0.1\,\mathrm{min}^{-1}$ and time step dt is $0.005\,\mathrm{min}$. The accuracy of Algorithm 1 could be achieved if dt is decreased; however, the computational cost will be increasing along with such a smaller time step.

Sometimes, Algorithm 1 may be computationally expensive since we could generate a lot of random numbers and find that no reaction happens. In order to design a more efficient method, we can use only one random number to control when the reaction occurs. We already know that the probability of the next reaction taking place is defined as a

Algorithm 1 Stochastic simulation – Degradation

Initialisation: $m(0) = 0$.
while $m > 0$ **do**
  Generation:
  `Generate a uniformly distributed number` r, $r \in (0, 1)$.
  Reaction:
  if $r < m(t)kdt$ **then**
    $m(t + dt) = m(t) - 1$ (Degradation takes place).
  else
    $m(t + dt) = m(t)$.
  end if
end while

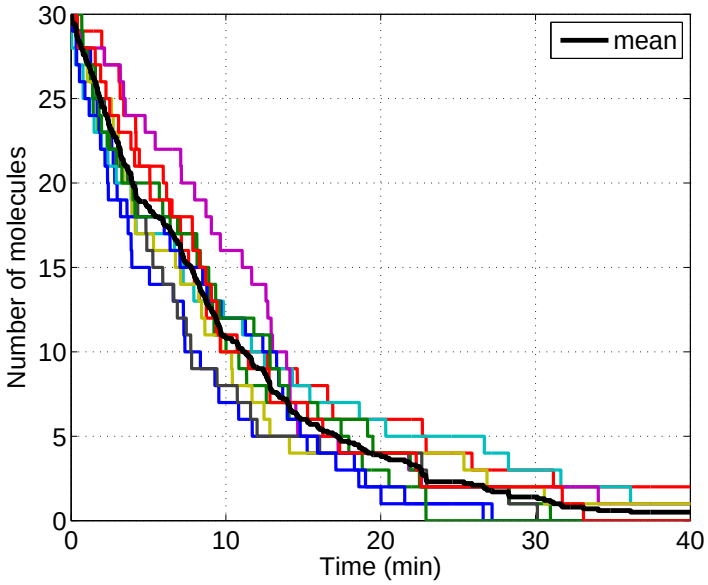


Figure 3.8: Ten realisations of degradation stochastic simulation (lines in colour) and the averaged profile over the realisations (black line).

(Equation 3.18). Gillespie (1977) showed that time step for the next reaction occurring is defined as:

$$\tau = \frac{1}{a} \ln(\frac{1}{r}), \tag{3.19}$$

where r is a uniformly distributed number. The derivation details are shown in Appendix A.3.

An example of production reaction is introduced here to describe how the algorithm is changed with the time step τ. This reaction is given by

$$\emptyset \xrightarrow{g} m, \tag{3.20}$$

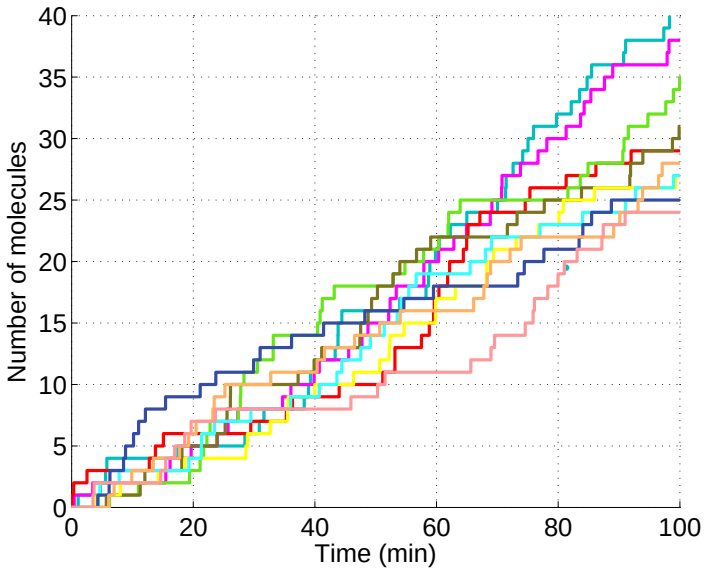


Figure 3.9: Ten realisations for production stochastic simulation (lines in colour) of Algorithm 2. The number of molecules in all the realisations starts from zero at initial time point and the simulation stops at 100 min. The production rate g is $0.3\,\mathrm{min}^{-1}$.

where g is the production rate to generate molecules of m. The probability that one molecule has been created during time interval $[t, t+dt]$ is gdt. The Algorithm 1 has been changed and shown in Algorithm 2.

Algorithm 2 Stochastic simulation – Production

Initialisation: $m(0) = 0$; $t = 0$.
while $t < T$ **do**
Generation:
Generate a uniformly distributed number r, $r \in (0, 1)$.
Calculation:
Calculate when the next reaction happens:
$t + \tau$, where $\tau = \frac{1}{a}\ln(\frac{1}{r})$, $a = g$.
Reaction:
$m(t+\tau) = m(t) + 1$
$t = t + \tau$
end while

Simulated results of Algorithm 2 are shown in Figure 3.9. The computational cost of Algorithm 1 has been reduced by τ because the reaction happens only at time point $t + \tau$. The comparison between these two algorithms for the same degradation reaction is shown in Figure 3.10, in which the time points in A (6974) are much larger than B (31).

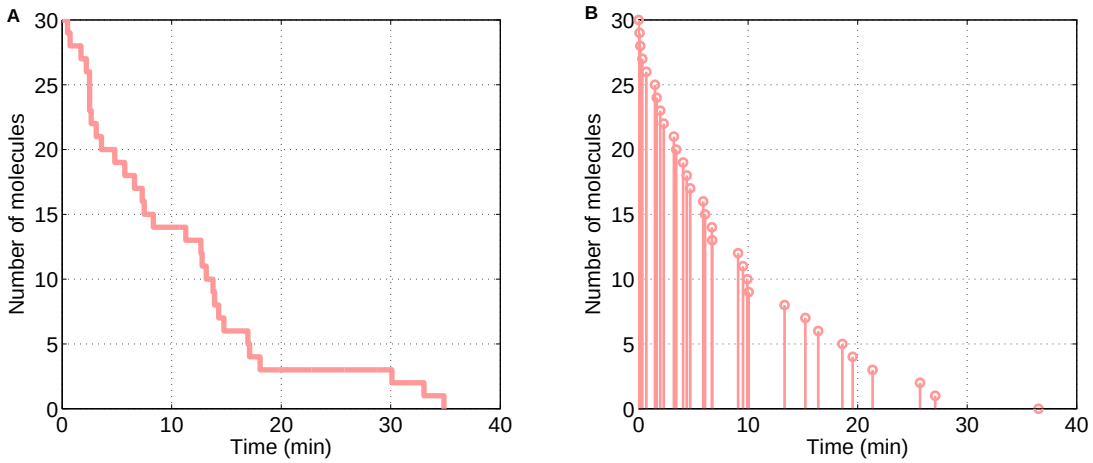


Figure 3.10: Simulation of degraded molecule concentration. Figure 3.10**A** is implemented by Algorithm 1, in which the computational cost is high. **B** shows the same reaction simulated by Algorithm 2, which is more effective.

The time step τ can also be expanded into multi-reactions system. Suppose there are Q reactions in a chemical system. The time step τ is similar to Equation 3.19:

$$\tau = \frac{1}{a}\ln(\frac{1}{r}) \tag{3.21}$$

where a is given by all the propensity functions in the system:

$$a = \sum_{j=1}^{Q} a_j \tag{3.22}$$

3.2.1.2 Which reaction occurs?

We have introduced how a random number controls when the next reaction happens in the last section. However, if there are multi-reactions occurring in one chemical system, how can we decide which reaction will take place? Consider two independent chemical reactions (degradation and production) appearing in one system:

$$m \xrightarrow{k} \emptyset, \quad \emptyset \xrightarrow{g} m. \tag{3.23}$$

The probabilities of the reactions taking place here, dependent on Equation 3.18, are defined as $m(t)kdt$ for degradation and gdt for production. The propensity function a_j is $m(t)k$ for degradation and g for production. The total propensity function of this system is given by:

$$a = m(t)k + g. \tag{3.24}$$

At time $t+\tau$, the probability of one molecule decaying is defined by $m(t)k/a$. Similarly, the probability of one molecule generated is given by g/a. The detailed pseudo-code is shown in Algorithm 3.

Algorithm 3 Degradation and Production stochastic simulation

Initialisation: $m(0) = 0$; $t = 0$
while $t < T$ **do**
  Generation:
  Generate two uniformly distributed numbers r_1 and r_2,
  r_1 and $r_2 \in (0, 1)$.
  Calculation:
  Calculate when the next reaction happens at $t + \tau$,
  where $\tau = \frac{1}{a}\ln(\frac{1}{r})$, $a = m(t)k + g$.
  Reaction:
  Decide which reaction takes place at $t + \tau$,
  if $r_2 < m(t)k/a$ **then**
    $m(t + \tau) = m(t) - 1$
  else
    $m(t + \tau) = m(t) + 1$
  end if
  $t = t + \tau$
end while

Figure 3.11 shows the several realisations of the chemical reactions given by Equation 3.23 and these stochastic results fluctuate around the corresponding ODE which will be discussed in the next section:

$$\frac{dm(t)}{dt} = -km + g. \tag{3.25}$$

3.2.1.3 Master equation and stochastic mean

Chemical master equations are used to describe the time evolution of the chemical system, in which the species movement is treated probabilistically. Assume that there is only one degradation reaction (Equation 3.17) in a chemical system. At time $t + dt$, the probability that there are n molecules of m in the system is defined as $P_n(t + dt)$, which is decided by the previous number of molecules at t. In order to achieve n molecules at $t + dt$, the degradation reaction happens if there are $n + 1$ molecules at time t and no reactions if there are n molecules at t:

$$P_n(t + dt) = P_{n+1}(t)k(n + 1)dt + P_n(t)(1 - kndt). \tag{3.26}$$

Let $x \to 0$, so the master equation is given by:

$$\frac{dP_n(t)}{dt} = k(n + 1)P_{n+1}(t) - knP_n(t). \tag{3.27}$$

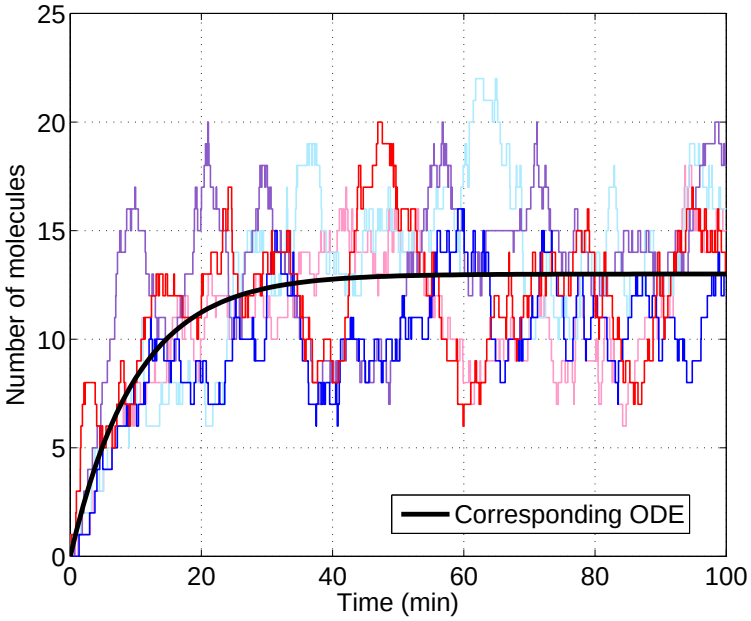


Figure 3.11: Five realisations for degradation and production stochastic simulation (lines in colour) of Algorithm 3. The number of molecules in all the realisations starts from zero at initial time point and the simulation stops at 100 min. The production rate g is 1.3 min^{-1} and degradation rate k is 0.1 min^{-1}.

In much the same way, by assuming that there is only one reaction occurring during a very small time interval dt, the master equation for the production and degradation reactions becomes:

$$\frac{dP_n(t)}{dt} = k(n+1)P_{n+1}(t) - knP_n(t) + gP_{n-1}(t) - gP_n(t). \tag{3.28}$$

Because the analytical solutions of the master equation are intractable in some complicated systems, such as Bicoid reaction-diffusion processes, the stochastic mean, which is the mean value of $m(t)$ over many realisations, can be used to describe the average performance of the system, i.e. average number of molecules. The derivation of the stochastic mean for the master equation in Equation 3.28 is shown below. The general stochastic mean is defined by:

$$\overline{m}(t) = \sum_{n=0}^{\infty} nP_n(t). \tag{3.29}$$

Moreover, since the $P_n(t)$ represents the probability that there are n molecules at time t, we have:

$$\sum_{n=0}^{\infty} P_n(t) = 1. \tag{3.30}$$

Instead of solving $P_n(t)$, we multiply n at both sides of Equation 3.28 and sum over n:

$$\sum_{n=0}^{\infty} \frac{dnP_n(t)}{dt} = \underbrace{\sum_{n=0}^{\infty} kn(n+1)P_{n+1}(t)}_{n+1 \to n} - \sum_{n=0}^{\infty} kn^2 P_n(t) + \underbrace{\sum_{n=0}^{\infty} gnP_{n-1}(t)}_{n-1 \to n} - \sum_{n=0}^{\infty} gnP_n(t). \tag{3.31}$$

In order to keep the consistence between left and right hand sides, the indices of the first term and third term should be changed as $\sum_{n=0}^{\infty} k(n+1)nP_n(t)$ and $\sum_{n=0}^{\infty} g(n+1)P_n(t)$, respectively, so we get:

$$\sum_{n=0}^{\infty} n\frac{dP_n(t)}{dt} = \sum_{n=0}^{\infty} k(n+1)nP_n(t) - \sum_{n=0}^{\infty} kn^2 P_n(t) + \sum_{n=0}^{\infty} g(n+1)P_n(t) - \sum_{n=0}^{\infty} gnP_n(t). \tag{3.32}$$

Combining Equation 3.29 and 3.30, the equation above becomes the ODE shown in Figure 3.11:

$$\frac{d\overline{m}(t)}{dt} = -k\overline{m}(t) + g. \tag{3.33}$$

This is the temporal production and degradation ODE without spatial diffusion. The Bicoid dynamic system will be introduced in the following section.

3.2.2 Bicoid stochastic reaction-diffusion model

Closely following Erban et al. (2007) and Wu et al. (2007), the stochastic Bicoid protein reaction diffusion system we implemented simulates 100 compartments along the A-P axis, each with length $h = 5\,\mu m$, which is approximately the average size of one nucleus.

The three chemical reactions involved in this description are:

$$\text{Bicoid}_1 \underset{d}{\overset{d}{\rightleftharpoons}} \ldots \underset{d}{\overset{d}{\rightleftharpoons}} \text{Bicoid}_i \underset{d}{\overset{d}{\rightleftharpoons}} \ldots \underset{d}{\overset{d}{\rightleftharpoons}} \text{Bicoid}_N, \quad \texttt{for} \quad i = 1, 2, \ldots, N \tag{3.34}$$

$$\text{Bicoid}_i \xrightarrow{\tau_p^{-1}} \emptyset, \quad \texttt{for} \quad i = 1, 2, \ldots, N \tag{3.35}$$

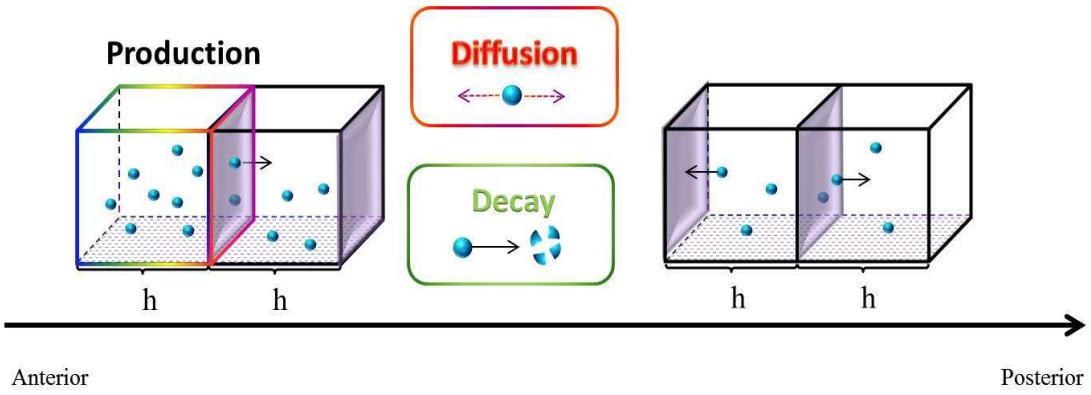


Figure 3.12: The diagram of Bicoid proteins chemical reaction-diffusion process along the A-P axis.

$$\emptyset \xrightarrow{S(t)} \text{Bicoid}_i, \quad \texttt{for} \quad i = 1 \tag{3.36}$$

The first of these, Equation 3.34, describes diffusion between neighbouring sub-volumes, allowed to take place in both directions, at a rate d, related to the diffusion constant of a deterministic model by $d = D/h^2$. The second, Equation 3.35, describes protein degradation, and the final, Equation 3.36, the source. Translation only takes place in the first bin, for $i = 1$.

Bicoid reaction-diffusion master equation is given by:

$$\begin{aligned}\frac{\partial}{\partial t}P(\mathbf{n},t) =& \underbrace{d\sum_{i=1}^{N-1}[(n_i+1)P(R_i^{\pm 1}\mathbf{n}) - n_iP(\mathbf{n})]}_{\texttt{Diffusion:}\,A\rightarrow P} + \underbrace{d\sum_{i=2}^{N}[(n_i+1)P(L_i^{\pm 1}\mathbf{n}) - n_iP(\mathbf{n})]}_{\texttt{Diffusion:}\,A\leftarrow P} \\ &+ \tau_p^{-1}\sum_{i=1}^{N}[(n_i+1)P(K_i^{+1}\mathbf{n}) - n_iP(\mathbf{n})] \\ &+ S(t)P[(K_1^{-1}\mathbf{n}) - P(\mathbf{n})],\end{aligned} \tag{3.37}$$

where $P(\mathbf{n},t)$ is joint probability of state vector $\mathbf{n} = [n_1, n_2, \ldots, n_i, \ldots, n_N]$ and $N = 100$. $R_i^{\pm 1}$, $L_i^{\pm 1}$, K_i^{+1} and K_i^{-1} are state operators, which are defined by:

$$R_i^{\pm 1}\mathbf{n} = [n_1, n_2, \ldots, n_i + 1, n_{i+1} - 1 \ldots, n_N], \quad i = 1, 2, \ldots, N-1 \tag{3.38}$$

$$L_i^{\pm 1}\mathbf{n} = [n_1, n_2, \ldots, n_{i-1} - 1, n_i + 1 \ldots, n_N], \quad i = 2, 3, \ldots, N \tag{3.39}$$

$$K_i^{+1}\mathbf{n} = [n_1, n_2, \ldots, n_i + 1, \ldots, n_N], \quad i = 1, 2, \ldots, N \tag{3.40}$$

$$K_i^{-1}\mathbf{n} = [n_1, n_2, \ldots, n_i - 1, \ldots, n_N], \quad i = 1, 2, \ldots, N \tag{3.41}$$

The first line in the chemical master equation corresponds to the Bicoid proteins diffusion throughout the A-P axis of the *Drosophila* embryo. The second line describes proteins

degradation while the final part is protein synthesis from *bicoid* mRNA. $S(t)$ is mRNA regulation function given by Equation 3.15.

Our implementation of the Gillespie algorithm for stochastic simulation of the master equation closely follows that of Erban et al. (2007) and is given in pseudo-code format in Algorithm 4. Similar to Algorithm 3, this process consists of the generation of two random numbers to select the time at which a reaction occurs, and which one that is. The probability that j-th chemical reaction taking place is given by: a_j/a, where a is a total propensity function, computed in step 2 (Algorithm 4). The vector $\mathbf{m}$ contains the number of molecules along the $N = 100$ bins while Equations 3.34 – 3.36 define a total of $R = 3N - 1$ reactions. The propensity functions for the reactions are:

$$\text{Bicoid}_1 \xrightarrow{d} \ldots \xrightarrow{d} \text{Bicoid}_N : \quad a_1 = \sum_{i=1}^{N-1} dm_i \tag{3.42}$$

$$\text{Bicoid}_1 \xleftarrow{d} \ldots \xleftarrow{d} \text{Bicoid}_N : \quad a_2 = \sum_{i=2}^{N} dm_i \tag{3.43}$$

$$\text{Bicoid}_i \xrightarrow{\tau_p^{-1}} \emptyset : \quad a_3 = \sum_{i=1}^{N} \tau_p^{-1} m_i \tag{3.44}$$

The propensity function for the source part is defined by $a_4 = S(t)m_1$ because this reaction occurs in the first bin.

The results for Bicoid stochastic reaction-diffusion in one stochastic simulation realisation based on the Gillespie algorithm *Direct Method* (Algorithm.4) are shown in Figure 3.14. This algorithm might be computationally intensive when the numbers of reactions increasing because it is waste to recalculate all the propensity functions at each time step. Therefore, Gibson and Bruck (2000) proposed more efficient implementation of the Gillespie stochastic simulation. For example, we can only update those propensity functions changed by reactions.

Algorithm 4 Bicoid reaction-diffusion stochastic simulation

Output: Bicoid molecular numbers, **m**

Initialisation: $\mathbf{m} \leftarrow 0$, $t \leftarrow 0$

while time<final time **do**

1. Generate two random numbers which are uniformly distributed in $(0, 1)$: $r(1)$ and $r(2)$.
2. Calculate propensity functions of all the reactions: $a = a_1 + a_2 + a_3 + a_4$.
3. Calculate the time when next reaction occurs: $t + \tau$, where $\tau = 1/a \ln(1/r(1))$.
4. Decide which reaction occurs at $t + \tau$: find $j \in R$ such that: $\sum_{i=1}^{j-1} a_i/a \leq r(2) < \sum_{i=1}^{j} a_i/a$
5. Update numbers of reactants and products in j-th reaction and set $t \leftarrow t + \tau$

end while

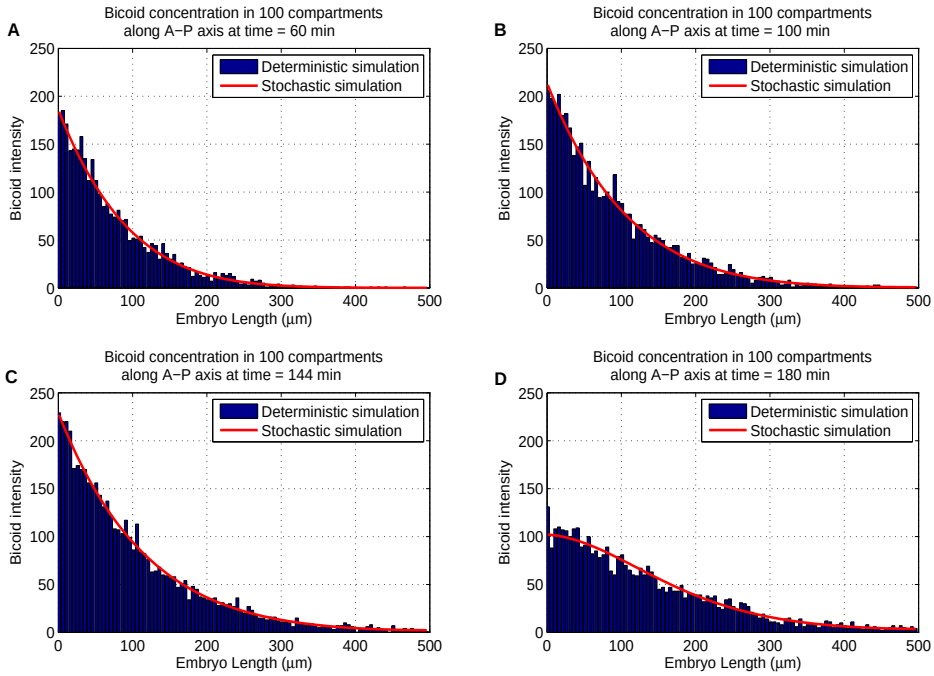


Figure 3.13: One realisation of the Bicoid stochastic simulation with mRNA regulation at 4 time points along the developmental period (blue bars), where mRNA decay onset time t_0 is 144 min. The deterministic simulation is implemented by ODEs given in Equations 3.47 – 3.49.

The results for Bicoid stochastic reaction-diffusion in one run of stochastic simulation based on the Gillespie algorithm, are shown in Figure 3.14. This model provides a more detailed understanding of the protein distribution, partitioned in compartments along

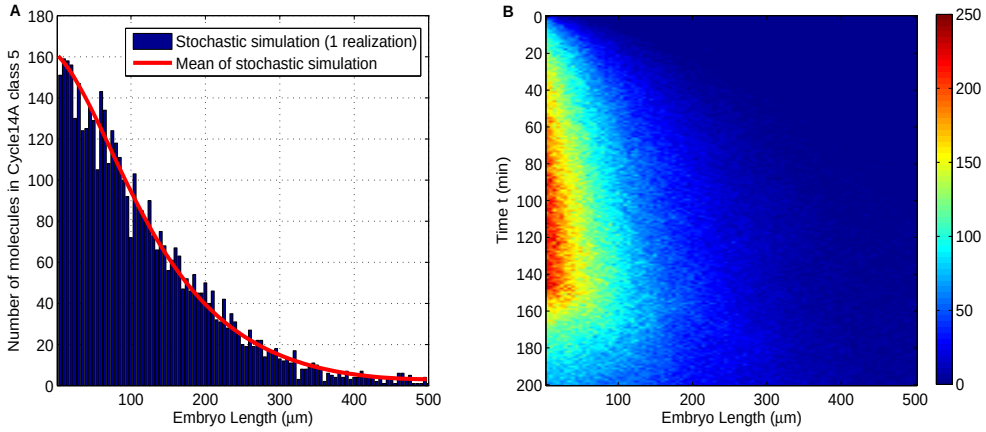


Figure 3.14: One realisation of stochastic simulation by Gillespie algorithm. Blue histogram, **A**, shows the numbers of Bicoid molecules along anterior and posterior axis in the embryo at a particular time point: Cycle 14*A* class 5. Average of several such simulations is used as model output to match against measurements. **B** shows the realisation jointly in space and time.

A-P axis. We note that such a stochastic model characterises a detailed view arising from molecular level variabilities.

In order to estimate parameters used in stochastic model, we have calculated vector $\overline{\mathbf{m}}$ by multiplying n_i and summing over the vector $\mathbf{n}$, $(i = 1, 2, \ldots, N)$ rather than averaging several Gillespie realisations of stochastic simulation.

$$\overline{\mathbf{m}} = \{\overline{m}_1, \overline{m}_2, \ldots, \overline{m}_i, \ldots, \overline{m}_N\}, \quad i = 1, 2, \ldots, N \tag{3.45}$$

$$\overline{m}_i(t) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \ldots \sum_{n_N=0}^{\infty} n_i P(\mathbf{n}, t) = \sum_{\mathbf{n}} n_i P(\mathbf{n}, t), \quad i = 1, 2, \ldots, N \tag{3.46}$$

where $\overline{m}_i(t)$ gives the mean number of Bicoid molecules at time t in i-th sub-volume. The details are introduced in Appendix A.4. Then we have a system of equations for $\overline{\mathbf{m}}$ followed as:

$$\frac{\partial}{\partial t}\overline{m}_1 = d(\overline{m}_2 - \overline{m}_1) - \tau_p^{-1}\overline{m}_1 + S(t), \qquad i = 1 \tag{3.47}$$

$$\frac{\partial}{\partial t}\overline{m}_i = d(\overline{m}_{i+1} + \overline{m}_{i-1} - 2\overline{m}_i) - \tau_p^{-1}\overline{m}_i, \quad i = 2, 3, \ldots, N-1 \tag{3.48}$$

$$\frac{\partial}{\partial t}\overline{m}_N = d(\overline{m}_{N-1} - \overline{m}_N) - \tau_p^{-1}\overline{m}_N, \qquad i = N \tag{3.49}$$

Equations 3.47 – 3.49 are ODEs for the Bicoid reaction-diffusion system, where d is diffusion rate for each sub-volume ($d = D/h^2$). This is a discretised system of Equation 3.10.

As a averaged result of the stochastic simulation, this system is used to estimate parameters in the following section.

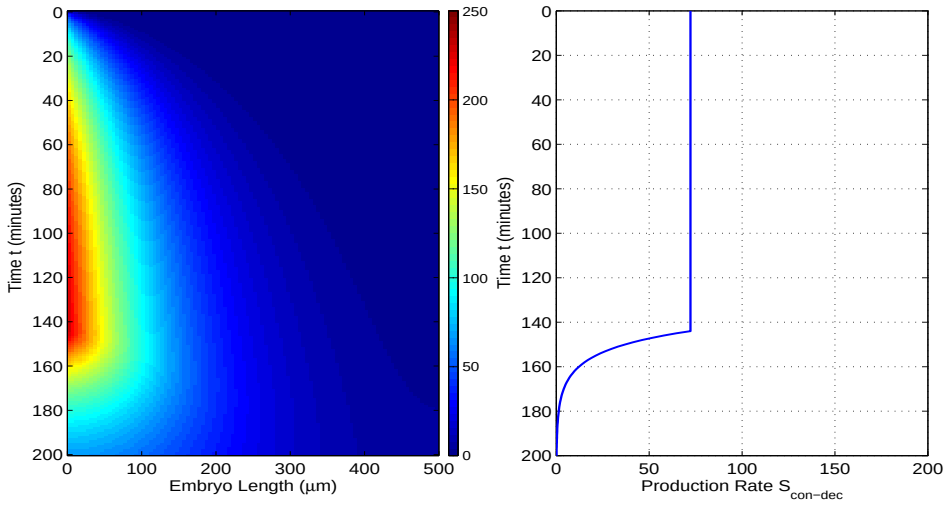


Figure 3.15: Spatio-temporal Bicoid morphogen profile of averaged stochastic model (ODEs) given by Equations 3.47 – 3.49. The mRNA regulated stability is shown in the right panel.

3.3 Parameter estimation

3.3.1 Matching models to quantitative measurements

In this section, we use the experimental measurements from `FlyEx` to estimate parameters in the three computational models we discussed above. It is notice that the measurements published in `FlyEx` are nuclear concentrations of Bicoid. The models we use, however, correspond to the total Bicoid. We make the assumption that the two concentrations are proportional across the developmental cycles. In recent work, Gregor et al. (2007b) published some measurements of nuclear and cytoplasmic Bicoid concentrations, showing the dynamical balance between the two during cycles of nuclear division. Their data are suggestive that the use of nuclear concentrations as proxy for total concentrations is reasonable. Once we assume the two are proportional, the parameters we infer by matching model outputs and data are unaffected, as any discrepancy will be absorbed by the source amplitude term S_0, computed by Equation 3.52.

The spatio-temporal data for Bicoid we use span 100 points uniformly spaced along the A-P axis, and cover 11 points in time. The temporal range of measurements starts from nuclear cleavage Cycle 11 to the end of Cycle 14A. Cycle 14A is of specific interest, because it is during this period that gap gene network is established and the established Bicoid profile begins to decay due to the decaying *bicoid* mRNA.

The squared error between model output and measured intensities is

$$Err = \sum_{t=T_1}^{T_2} \sum_{x=1}^{L} \{S_0 m\,(x,\,t)\;-\;m_d\,(x,\,t)\}^2\,, \tag{3.50}$$

$$\theta = \arg\min_{\theta \in \mathcal{R}}(Err(\theta)), \tag{3.51}$$

where $m(x,t)$ is the model output while $m_d(x,t)$ denotes the measured intensities from `FlyEx`. T_1 and T_2 are the boundaries of cleavage cycles $11-14A$. θ in Equation (3.51) represents a vector of all unknown model parameters and $\mathcal{R}$ is the space over which we search for optimum values.

Because the model output is linear in the source amplitude S_0 and is independent of the other parameters used in the three models, we calculate it in closed form rather than searching for an optimum in a grid. In order to minimise error Err in Equation (3.50), we differentiate it with respect to S_0 and equate it to zero. Then we have S_0 as following:

$$S_0 = \frac{\sum_{t=T_1}^{T_2} \sum_{x=1}^{L} m(x,t) m_d(x,t)}{\sum_{t=T_1}^{T_2} \sum_{x=1}^{L} m(x,t)^2}. \tag{3.52}$$

3.3.2 Results and discussion

For the deterministic diffusion and stochastic models, there are four parameters (diffusion constant D, protein degradation time τ_p, source mRNA degradation time τ_m and decay onset time t_0). For Hecht et al. (2009)'s cytoplasmic flow model, there is an additional parameter, the flow velocity V.

Table 3.1 shows the search spaces used in optimising the parameters of the three models considered. We used a coarse grid in the first round to get a rough estimate of the sensible range of parameters and followed it with a second round of search with a higher resolution and a reduced search range. Such a strategy is feasible, given that we have only five parameters to estimate. Further, given the noisy nature of the available data, searching over a finer grid to optimise parameters to a higher level of numerical precision does not make sense. If data of higher quality become available in the future, a scheme based on simulated annealing or population-based optimisation needs to be considered. With the grid sizes we chose, shown in Table 3.1, it was possible to do least squares fitting of all three models on a desktop PC, with at most three days of wall clock time.

Values of estimated parameters for the different models are shown in Tables 3.2 and 5.1, for the regulated stability model and a model in which source mRNA is permitted to decay from time zero (unregulated mRNA regulation, as shown in Figure 3.3(b)), respectively. We note that parameter values estimated by the fitting procedure are in sensible ranges used by previous authors.

Table 3.1: Parameter optimisation on a regular grid.

Space and resolution for the first round of search.			
Parameters	Diffusion model	Stochastic model	Flow model
$D(\mu m^2/s)$	0.1:0.2:5	0.1:0.2:5	0.1:0.3:5
t_0 (min)	120:2:160	120:2:160	120:3:160
τ_p (min)	40:2:100	40:2:100	40:3:100
τ_m (min)	1:2:30	1:2:30	1:3:30
$V(\mu m/s)$	Null	Null	0.01:0.02:0.1
Space and resolution for the second round of search.			
Parameters	Diffusion model	Stochastic model	Flow model
$D(\mu m^2/s)$	2:0.1:4	2:0.1:4	0.1:0.1:1.5
t_0 (min)	135:1:150	135:1:150	135:1:150
τ_p (min)	70:1:95	70:1:95	35:1:50
τ_m (min)	1:1:15	1:1:15	1:1:12
$V(\mu m/s)$	Null	Null	0.02:0.01:0.06

Table 3.2: Parameter estimation – Regulated stability.

Estimated Parameters	Diffusion	Stochastic	Flow
Diffusion constant D $(\mu m^2/s)$	3	3	0.9
mRNA decaying onset time t_0 (min)	143	144	142
Bicoid proteins degradation time τ_p (min)	87	86	42
bicoid mRNA degradation time τ_m (min)	9	9	7
Flow velocity V $(\mu m/s)$	N/A	N/A	0.04
Source intensity S_0	352	72	104

Parameter values estimated by matching model outputs to observed data from `FlyEx`. Least squares fitting of model outputs to `FlyEx` with exhaustive search for the best combination of parameters on a regular grid suggests sensible values for the mRNA decay onset time, t_0, in all three models. Regulated stability corresponds to an optimised period in time during which the mRNA is held stable and translated at a constant rate, followed by rapid decay.

Figure 3.16 shows how the models achieve a reduction of a quarter, in the root mean square error between model outputs and `FlyEx` measurements in the post-peak region of nuclear cleavage cycles 11 – 14A. This comparison between modelling errors, with and without our regulated source, confirms the merits of explicitly modelling the destruction of maternally deposited mRNA.

As seen in Figure 3.17, for the mRNA regulated stability estimates, there is strong agreement across the three different models with respect to the onset of source decay (t_0), and the speed at which it is decayed (τ_m^{-1}), the main focus of our investigation.

Table 3.3: Parameter estimation – Unregulated stability.

Estimated Parameters	Diffusion	Stochastic	Flow
Diffusion constant D $(\mu m^2/s)$	1.1	1.1	0.4
mRNA decaying onset time t_0 (min)	0.01	0.01	0.01
Bicoid proteins degradation time τ_p (min)	250	250	156
bicoid mRNA degradation time τ_m (min)	38	37	13
Flow velocity V $(\mu m/s)$	N/A	N/A	0.01
Source intensity S_0	901	188	980

Parameters values estimated cooperating with unregulated stability, where the mRNA is allowed to decay from the very beginning; these parameters were estimated by forcing $t_0 = 0.01$ min in the optimisation loop.

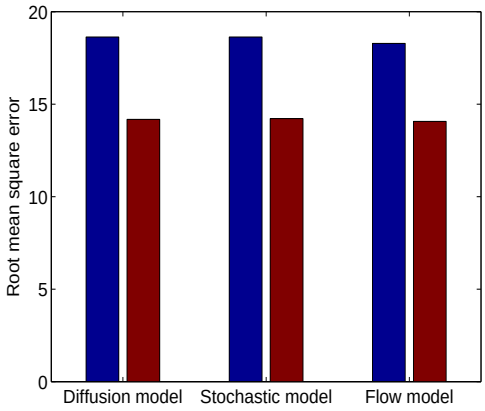


Figure 3.16: Reduction in squared error between model outputs and `FlyEx` measurements. In all three models, nearly a quarter reduction is achieved by the improved source whose parameters are optimised. Blue bars represent modelling errors for a constant source model and the red bars correspond to the regulated mRNA model.

As noted in Section 2.2, these observations confirm the experimental findings in (Surdej and Jacobs-Lorena, 1998) and (Spirov et al., 2009). Surdej and Jacobs-Lorena (1998) argued that the mRNA is developmentally regulated, i.e. being held stable for up to the first two hours and then rapidly killed off in the next 30 min. Spirov et al. (2009) also suggested that the rapid degradation takes place over a 15 – 20 min interval. The rapid decay of mRNA suggested in both these papers is consistent with degradation times of 9, 9 and 7 minutes inferred from our models.

We note that the diffusion constant estimated for Hecht et al. (2009)'s cytoplasmic flow model is smaller than the other two. This is to be expected since the motivation of this model is to use cytoplasmic flow as an additional trafficking mechanism that offsets a low diffusion constant. The value we estimated for flow velocity (0.04 $\mu m/s$) is close to

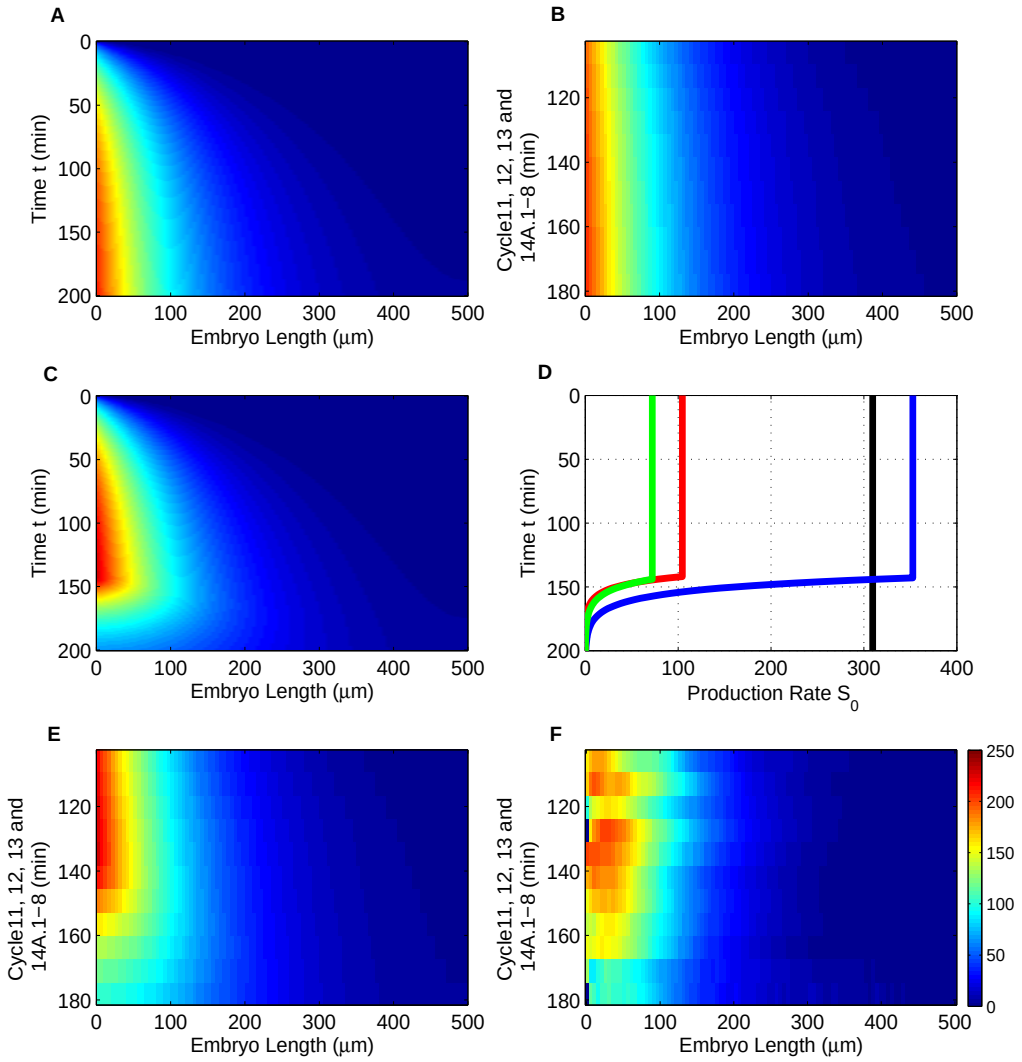


Figure 3.17: Spatio-temporal profiles of Bicoid and regulated anterior mRNA profiles inferred using three different computational models. **A** and **B** spatio-temporal profiles for a conventional model that assumes a constant source (drawn over two timescales). Inferred source profiles are in shown in **D**, for deterministic diffusion (blue), cytoplasmic flow (red) and the stochastic (green) models. They differ in the source amplitudes required to fit the data, but the estimated decay onset times are very close. The corresponding spatio-temporal profile is shown in **C** over the full time and space axes. **E** and **F**: model output and `FlyEx` data in the space-time range over which optimisation was carried out. Profile shown in **E** is only for the deterministic diffusion model for clarity.

what was used in (Hecht et al., 2009) (0.08 $\mu m/s$), who take this estimate from observed nuclear motions. They note a 20-fold large range of possible values for this parameter, and use an average value. It is encouraging that the parameter obtained by fitting to `FlyEx` happens to be quite close.

Table 5.1 shows the parameter estimation for an unregulated source which allows for mRNA decay from time zero (0.01 min in our simulation). This possibility is a natural expectation we need to explore, since mRNA molecules are inherently unstable. In order to match the measurements in the post-peak region, it turns out that this model not only has to amplify the source (S_0) to almost ten times that of the other models but also has to retain the protein in the medium for a much longer period ($\tau_p = 250\,\&156\,\text{min}$). These values of protein degradation time are significantly higher than what is thought to be of Bicoid proteins (Driever and Nüsslein-Volhard, 1988b). Further, the source amplitude being so high is inconsistent with the observation that Bicoid protein is often undetectable during the very early stages of development (Driever and Nüsslein-Volhard, 1988b; Grimm et al., 2010). Thus, it is reasonable to conclude that the source supply is regulated as suggested by Surdej *et al.*, rather than either being kept constant throughout or being subject to natural decay.

Figure 3.18 shows cross sections of the error function at the optimum point found by grid search. We have shown this with respect to all parameter combinations, taken pair-wise, setting the parameters not shown to their optimum values. The unimodal form of these error functions, shown here for the deterministic diffusion model, confirms that the optimisation strategy we chose was adequate for this purpose. Similar error surface plots for the other two models are given in Figures 3.19, 3.21 and 3.23. Over the parameter ranges considered for the search, the error surface turns out to be unimodal for all three models.

We note that previous authors working on Bicoid profiles have used a range of different values for diffusion and protein degradation parameters. For the diffusion constant, for example, values of, 0.3 (Gregor et al., 2007b), 7.0 (Little et al., 2011; Abu-Arish et al., 2010; Porcher et al., 2010) and $17\,\mu m^2/s$ (Gregor et al., 2005; Bergmann et al., 2007) have been used. With our models, we explored the effect of fixing one or more of the parameters at a value used by previous authors and optimising the remaining parameters. We found the dominant effect is one of the diffusion terms compensating for the protein degradation time, with the decay onset time and transcript degradation time we compute showing far less variation.

We have further quantified the uncertainties in our estimates of t_0 and τ_m by fitting the models to individual embryo measurements in `FlyEx` rather than their average profiles. We achieved this by constructing 50 reference datasets by uniformly bootstrapping from each temporal class in `FlyEx`. Figure 3.25 shows these uncertainties as box plots and confirms the fact that the estimated onset and decay rates are consistent across all three models.

Our models permit the exploration of other published hypotheses about potential mRNA regulation. For example, Salles et al. (1994), treating the poly(A) tail length of *bicoid* mRNA as proxy for its translational competence, suggest that protein production may be

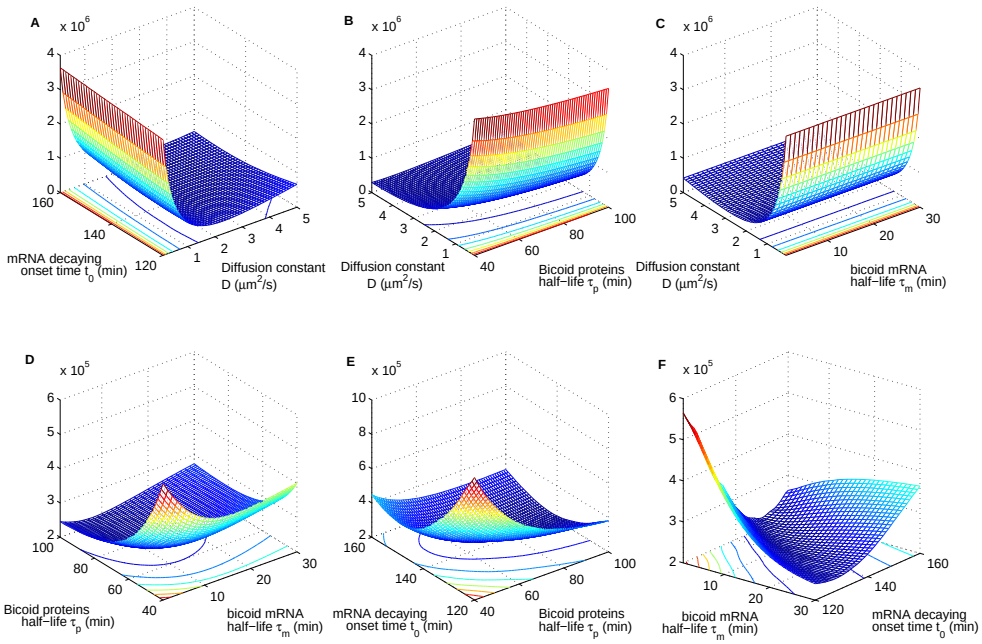


Figure 3.18: Cross sections through the error function between deterministic diffusion model output and measurements. Figures show the error function with respect to parameters taken pairwise, with those not shown held constant at their optimum values given in Table 3.2.

restricted in time, peaking between 1 to 1.5 hours in development. We have simulated this by implementing the source as a rectangular function between 60 min and Cycle 14*A*, and computing the resulting Bicoid profile, shown in Figure 3.26. We found the corresponding modelling error to be significantly higher, caused mainly by forcing the decay to be instantaneous.

Our implementation in deriving the main results for the stochastic model in 3.2, following the technique of Erban et al. (2007), via simultaneous ordinary differential equations corresponding to discrete bins along the spatial axis, captures average behaviour. Asymptotically (i.e. with increasing number of bins), this is the equivalent of averaging a large number of Gillespie simulations, and should also give the same solution as the deterministic differential equation. To estimate the effect of molecular level variation, we matched profiles generated by individual Gillespie simulations to bootstrap samples of Bicoid profiles from `FlyEx` (the same data used to derive uncertainties in Figure 3.25). As this process is computationally demanding, we restricted ourselves to estimating the variability in mRNA decay onset time only, with the remaining parameters fixed to their optimal values given in Table 3.2. Matching such individual simulations to data resulted in an increase in the standard deviation of estimation from 3.9 min to 5.5 min. While this increase suggests that the variability at the molecular level may be captured

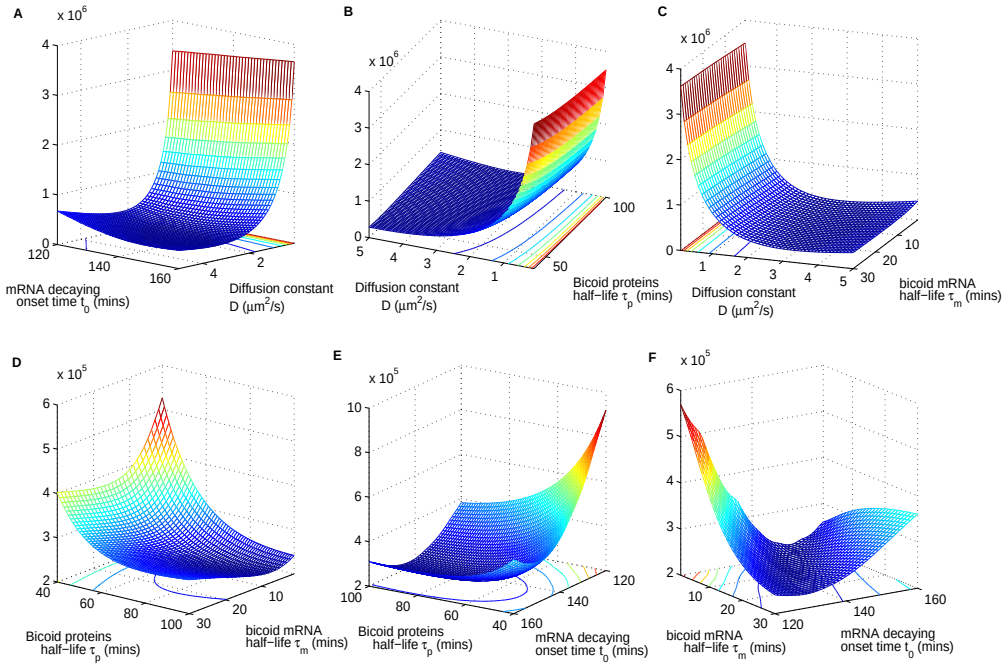


Figure 3.19: Modelling error displayed as functions of parameters taken pairwise. Stochastic simulation model.

by stochastic simulations, the resulting estimation uncertainties in both cases are still small for the mRNA decay onset.

3.3.3 *bicoid* mRNA with spatial distribution

While nearly all modelling works on Bicoid assume a spatial point source for *bicoid* mRNA, as noted in Chapter 2, Spirov et al. (2009) suggested that the *bicoid* mRNA may have spatial distribution which alone explains the morphogen gradient at the protein level. They argued for an active transport mechanism along a cortical microtubular network. This proposal is questioned by Little et al. (2011) who showed experimental evidence that a distributed spatial gradient of mRNA is not sufficient to achieve the required morphogen profile. Since computational modelling of active transport hypothesised by Spirov et al. (2009) is outside the scope of this study, we instead follow Dilão and Muraro (2010) who have postulated an mRNA diffusion model to achieve an effect similar to that of (Spirov et al., 2009). Instead of a reaction-diffusion equation (incorporating a term for natural mRNA decay (Dilão and Muraro, 2010)), we restrict ourselves to the heat equation given by:

$$\frac{\partial R}{\partial t} = D_r \frac{\partial^2 R}{\partial x^2}, \tag{3.53}$$

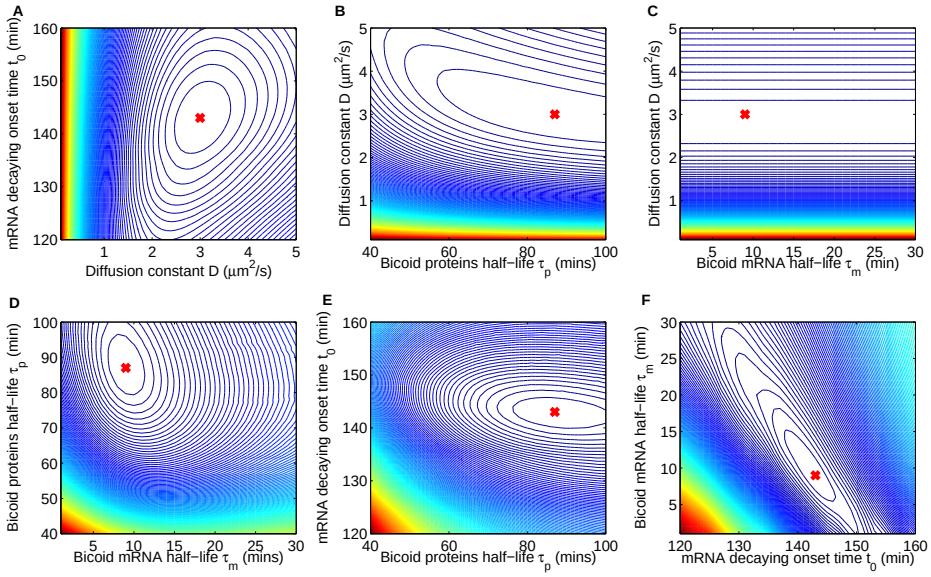


Figure 3.20: Contour plot for parameter estimation in diffusion and stochastic model according to 3D plot in Figure 3.18 and 3.19.

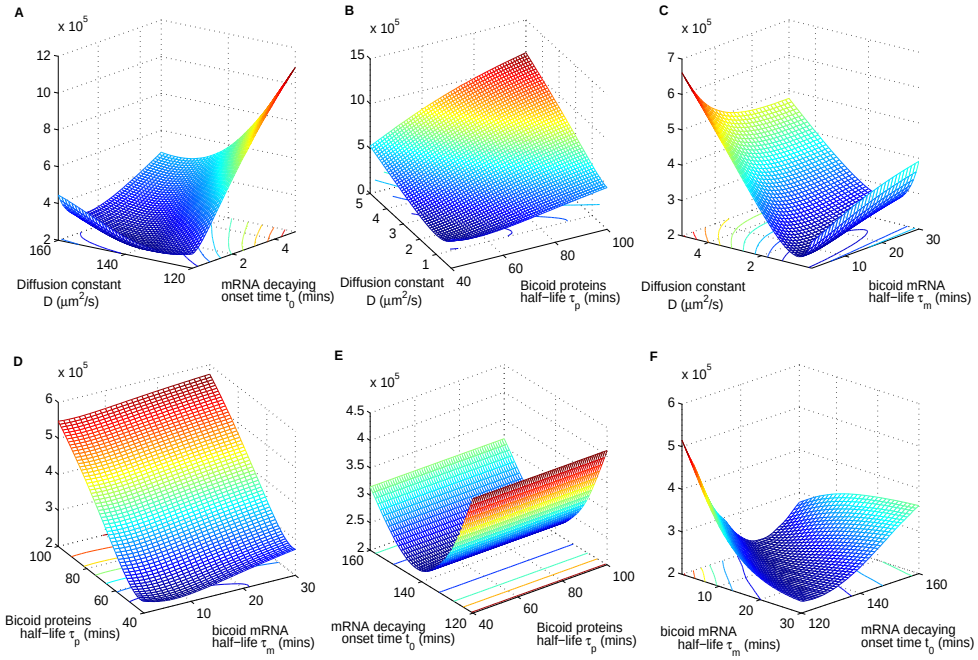


Figure 3.21: Modelling error displayed as functions of parameters taken pairwise: Cytoplasmic flow model.

where D_r is mRNA diffusion constant. This is justified because our model for the temporal regulation of *bicoid* mRNA is one which holds it stable up to t_0 followed by an active degradation.

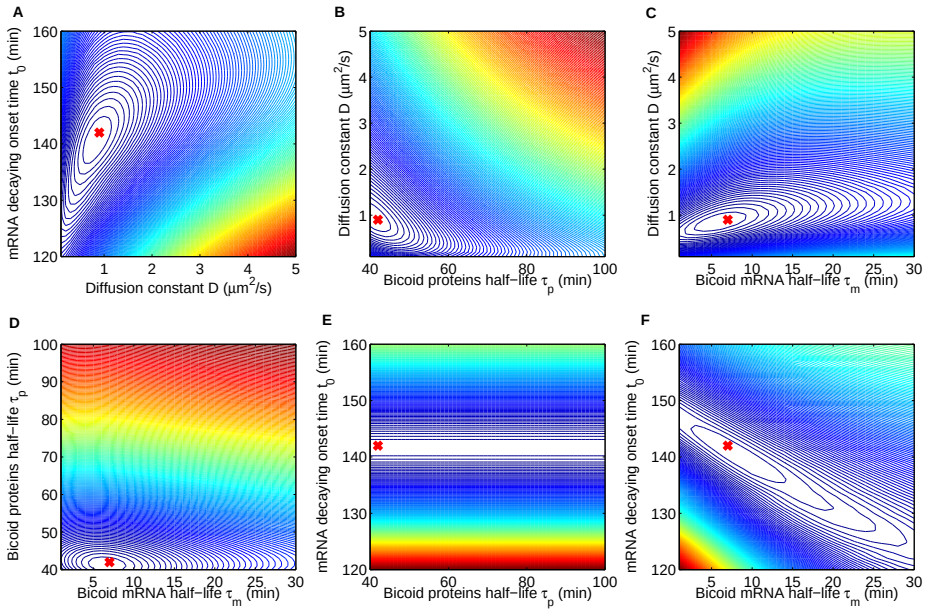


Figure 3.22: Contour plot for parameter estimation in flow model according to 3D plot in Figure 3.21.

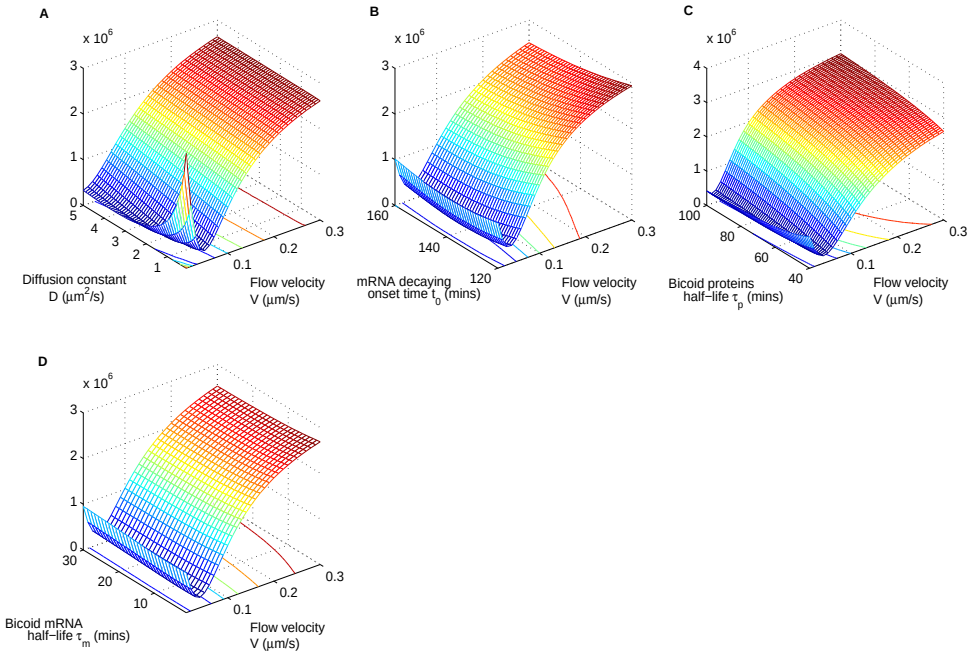


Figure 3.23: Modelling error surface for the cytoplasmic flow model as functions of flow velocity parameter and each of the other parameters.

Figure 3.27 shows protein intensities with a spatial distribution for *bicoid* mRNA. Figure 3.27A is profile obtained with only spatially distributed mRNA, while B is the result

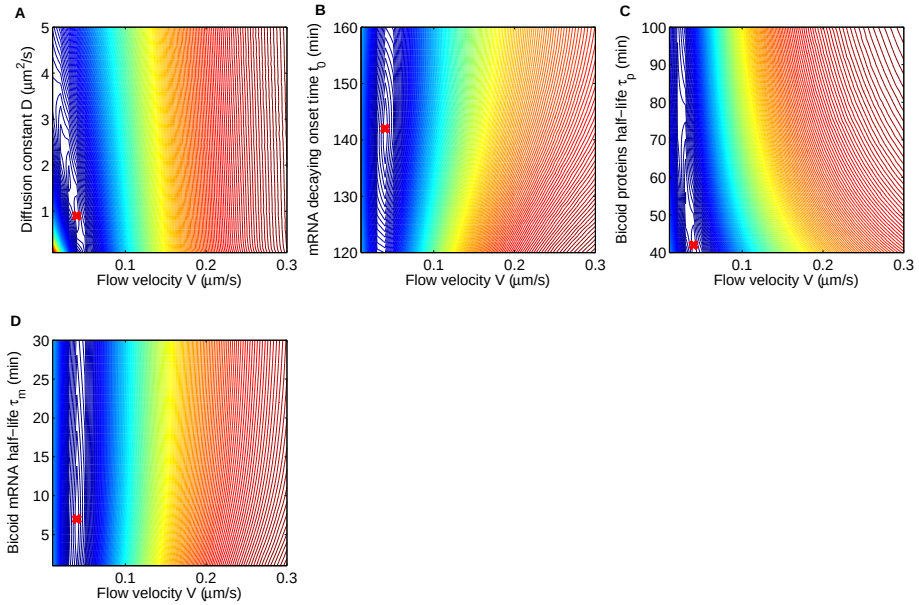


Figure 3.24: Contour plot for parameter estimation in flow model according to 3D plot in Figure 3.23.

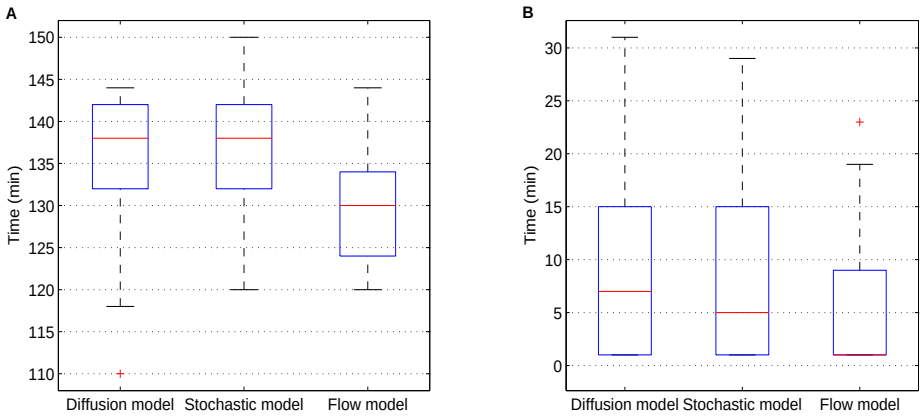


Figure 3.25: Uncertainty estimates of mRNA decay onset time t_0 in **A** and degradation time τ_m in **B** by fitting the models to 50 bootstrap samples of individual embryo measurements from `FlyEx`.

obtained with spatial distribution and temporal regulation and the post peak decay is clearly observed. Thus, even with simulated spatial distribution of maternal mRNA, our model finds a set of feasible parameter values that account for observed profiles in `FlyEx`. The corresponding parameter estimates are shown in Table 3.4. We find that the differences are in directions we would naturally expect, i.e. a spatially distributed maternal mRNA is compensated primarily by protein degradation. But it is encouraging to see that the onset of decay (t_0) changes only slightly.

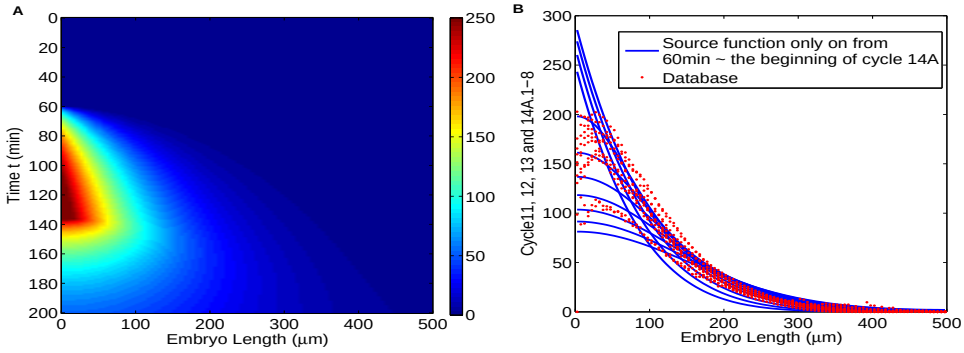


Figure 3.26: Spatio-temporal Bicoid profiles with source regulation as a step function, with constant rate of translation between 60 min and onset of Cycle 14*A*

Table 3.4: Parameter optimisation for the Bicoid stochastic mean model with mRNA regulation and spatial distribution.

Parameter:	$D(\mu m^2/s)$	t_0 min	τ_p min	τ_m min	$D_r(\mu m^2/s)$	S_0
	3.7	148	30	19	0.4	170

Parameter values estimated by matching model outputs to observed data from `FlyEx`. Least squares fitting of model outputs to `FlyEx` with exhaustive search for the best combination of parameters on a regular grid suggests sensible values for the mRNA decay onset time, t_0, in all three models. Regulated stability corresponds to an optimised period in time during which the mRNA is held stable and translated at a constant rate, followed by rapid decay.

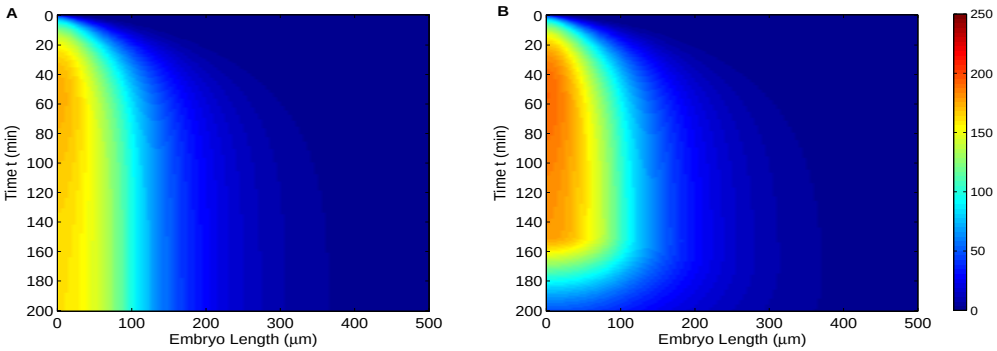


Figure 3.27: The effect of *bicoid* mRNA spatial gradient. **A** protein intensity without mRNA temporal regulation; **B** Bicoid profile with mRNA temporal regulation.

3.4 Summary

In this chapter, the *bicoid* mRNA stability regulation is proposed, which has been ignored for decades in literature since Bicoid was discovered. We have implemented three computational protein reaction-diffusion models cooperating such a regulated source function. The grid search has been used to estimate parameters for each model since the number of unknown parameters is small. According to the model parameters, our hypothesis of mRNA regulation has been confirmed that the source should be kept constant around two hours and followed by quickly decay. Comparing with the traditional models with constant source, the root mean square error between model outputs and database have been reduced with the regulated mRNA.

Chapter 4

Gaussian process modelling for *bicoid* mRNA regulation

In the previous chapter, we introduced the *bicoid* mRNA stability regulation mechanism, and how the protein profile changes with this source model. To match the measurement on real fly embryos from the database `FlyEx`, we have estimated the regulation parameters t_0, at which mRNA starts decaying and τ_m, the decaying rate, by exhaustive searching.

By looking at the noisy observations of the Bicoid spatio-temporal profile, this may generate the question as to how we can infer the mRNA regulation function, without making any strict assumptions as in Chapter 3. In this chapter, we bring a powerful Bayesian non-parametric machine learning methodology – Gaussian process (GP) regression – to study the establishment of the Bicoid morphogen profile by considering maternal mRNA as a latent function. By modelling the spatio-temporal morphogen profile as the output of a linear dynamical system, driven by the source function, we are able to exploit the published measurements in a novel way.

Our work builds on similar approaches by Gao et al. (2008) for temporal models, but takes a significant leap into a spatio-temporal problem. With synthetic data from a one dimensional diffusion model with a source simulated to model mRNA stability regulation, our results establish that the GP method can accurately infer the driving function and capture the spatio-temporal dynamics of embryonic Bicoid propagation. On real data from `FlyEx` database, too, the reconstructed source function, is indicative of stability regulation, but is temporally smoother than what we expected, partly due to the fact that the dataset is only partially observed. As similar in Chapter 3, we also analyse this model with a spatial gradient of maternal mRNA, rather than the mRNA being fixed at only the anterior pole.

We start this chapter with the basic concept of GPs, which are introduced in the beginning of Section 4.1. An toy problem of applying GPs on a simple time-varying biological model is shown in the end of this section. From Section 4.2, we show the details of *bicoid* mRNA GP inference in the spatio-temporal Bicoid profile. The corresponding results are discussed in Section 4.3.

4.1 Gaussian processes

4.1.1 Multivariate Gaussian

An univariate Gaussian distribution is defined as $x \sim \mathcal{N}(\mu, \sigma^2)$, in which the probability density function of x is given by:

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \left(-\frac{(x-\mu)^2}{2\sigma^2} \right), \tag{4.1}$$

where μ and σ are mean and standard deviation. Given a n-dimensional random vector $\mathbf{x}$ ($\mathbf{x} \in \Re^n$), of which each variate has an univariate Gaussian distribution, the probability density function of such a vector, which has a multivariate Gaussian distribution, is given by:

$$p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \tag{4.2}$$

where $\boldsymbol{\mu}$ ($\boldsymbol{\mu} \in \Re^n$) is the mean vector. $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma} \in \Re^{n \times n}$), a covariance matrix, must be symmetric positive semidefinite in order to be valid.

To make the concept of multivariate Gaussian more clear, we show a bivariate example here by generating a two-dimensional vector $\mathbf{a}$ ($\mathbf{a} \in \Re^2$, $\mathbf{a} = [a_1, a_2]^T$) where

$$\mathbf{a} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{4.3}$$

The mean and covariance matrix are given by:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}. \tag{4.4}$$

$\rho\sigma_1\sigma_2$ is the cross-covariance of the two variates a_1 and a_2. The correlation between a_1 and a_2 becomes high with a large correlation parameter ρ, and is zero when the two components are independent.

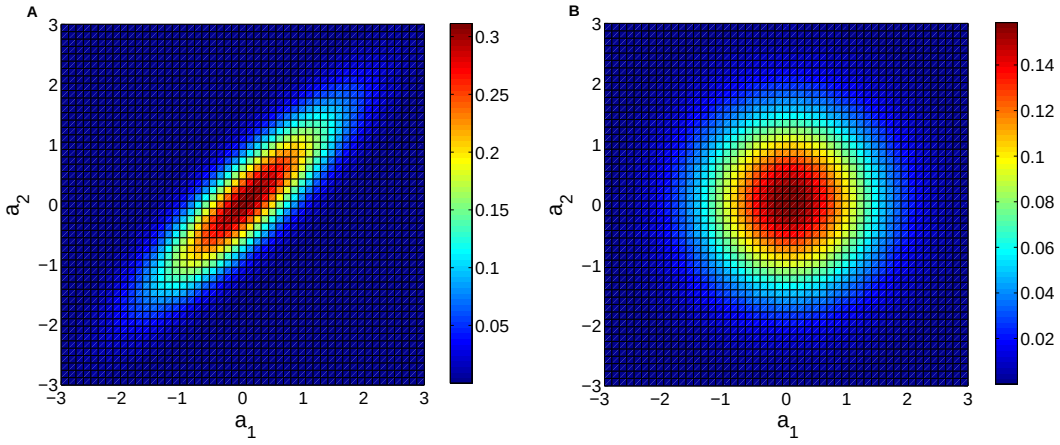


Figure 4.1: Contours of the multivariate Gaussian joint distribution over two variates a_1 and a_2. The covariance matrix is $[1\,0.86; 0.86\,1]$ in Figure 4.1**A** and $[1\,0; 0\,1]$ in Figure 4.1**B**. There is no correlation between the two components in Figure 4.1**B**.

We now consider two vectors $\mathbf{a}$ ($\mathbf{a} \in \Re^n$) and $\mathbf{b}$ ($\mathbf{b} \in \Re^n$), which have joint Gaussian distribution:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{B} \end{bmatrix} \right). \tag{4.5}$$

The covariance matrix $\boldsymbol{\Sigma}$ ($[\mathbf{A}\,\mathbf{C}^T; \mathbf{C}\,\mathbf{B}]$) is separated into the block matrices above, in which, $\mathbf{A}$ and $\mathbf{B}$ are corresponding to the covariance matrices for $\mathbf{a}$ and $\mathbf{b}$, respectively. $\mathbf{C}$ is the cross-covariance matrix between $\mathbf{a}$ and $\mathbf{b}$ and becomes zero when $\mathbf{a}$ and $\mathbf{b}$ are uncorrelated. We simplify the representation by zero-mean assumption ($\boldsymbol{\mu}_a = \boldsymbol{\mu}_b = \mathbf{0}$) and the joint distribution can be rewritten as:

$$p(\mathbf{a}, \mathbf{b}) \propto \exp\left(-\frac{1}{2} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}^T \begin{bmatrix} \mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right). \tag{4.6}$$

By the Schur complement, the block matrix is given by (derivation details are shown in Appendix A.5):

$$\begin{bmatrix} \mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{B^{-1}C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A} - \mathbf{C}^T\mathbf{B}^{-1}\mathbf{C})^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{C}^T\mathbf{B}^{-1} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \tag{4.7}$$

The joint distribution in Equation 4.6 becomes:

$$p(\mathbf{a},\mathbf{b}) \propto \exp\left(-\frac{1}{2}\begin{bmatrix}\mathbf{a}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{b}\\ \mathbf{b}\end{bmatrix}^T\begin{bmatrix}(\mathbf{A}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{C})^{-1} & \mathbf{O}\\ \mathbf{O} & \mathbf{B}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{a}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{b}\\ \mathbf{b}\end{bmatrix}\right)$$
$$\propto \exp\left(-\frac{1}{2}(\mathbf{a}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{b})^T(\mathbf{A}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{C})^{-1}(\mathbf{a}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{b})\right)\exp\left(-\frac{1}{2}\mathbf{b}^T\mathbf{B}^{-\mathbf{1}}\mathbf{b}\right). \tag{4.8}$$

The conditional distribution of $\mathbf{a}$ given $\mathbf{b}$ can be calculated as:

$$p(\mathbf{a}|\mathbf{b}) = \frac{p(\mathbf{a},\mathbf{b})}{p(\mathbf{b})} = \mathcal{N}(\mathbf{C}^T\mathbf{B}^{-1}\mathbf{b}, \mathbf{A}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{C}) \tag{4.9}$$

With the non-zero means, equation above becomes:

$$p(\mathbf{a}|\mathbf{b}) = \mathcal{N}(\boldsymbol{\mu}_a + \mathbf{C}^T\mathbf{B}^{-1}(\mathbf{b}-\boldsymbol{\mu}_b), \mathbf{A}-\mathbf{C}^T\mathbf{B}^{-1}\mathbf{C}) \tag{4.10}$$

4.1.2 Gaussian processes

Nowadays, GPs are becoming increasingly popular in machine learning for both regression and classification problems (Rasmussen and Williams, 2006). Differing from Gaussian distribution, GPs specify distributions over the functions. With the input time points $\mathbf{x}_*$, known as test inputs, we can generate a sample function from GP prior with the covariance matrix $\mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*}$ generated from RBF.

$$\mathbf{f}_* \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}_*}, \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*}) \tag{4.11}$$

This random sampled function, which provides a functional view in Gaussian distribution, and the covariance matrix are both shown in Figure 4.2.

Instead of drawing random functions from the Gaussian prior above, we combine the information from the observation function $\mathbf{f}$, with the inputs given by $\mathbf{x}$, known as training inputs, to make predictions for the function $\mathbf{f}_*$ at the locations given by $\mathbf{x}_*$. Therefore, the joint distribution of training outputs $\mathbf{f}$ and test outputs $\mathbf{f}_*$ is given by:

$$\begin{bmatrix}\mathbf{f}\\ \mathbf{f}_*\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_{\mathbf{f}}\\ \boldsymbol{\mu}_{\mathbf{f}_*}\end{bmatrix}\begin{bmatrix}\mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{\mathbf{f},\mathbf{f}_*}\\ \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} & \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*}\end{bmatrix}\right) \tag{4.12}$$

If there are n training points in $\mathbf{x}$ and n_* test points in $\mathbf{x}_*$, we can generate covariance matrices $\mathbf{K}_{\mathbf{f},\mathbf{f}}$ $(n \times n)$ and $\mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*}$ $(n_* \times n_*)$ associated with training data and test data, respectively. $\mathbf{K}_{\mathbf{f},\mathbf{f}_*}$ $(n \times n_*)$ and $\mathbf{K}_{\mathbf{f}_*,\mathbf{f}}$ $(n_* \times n)$ are the corresponding cross-covariance matrices.

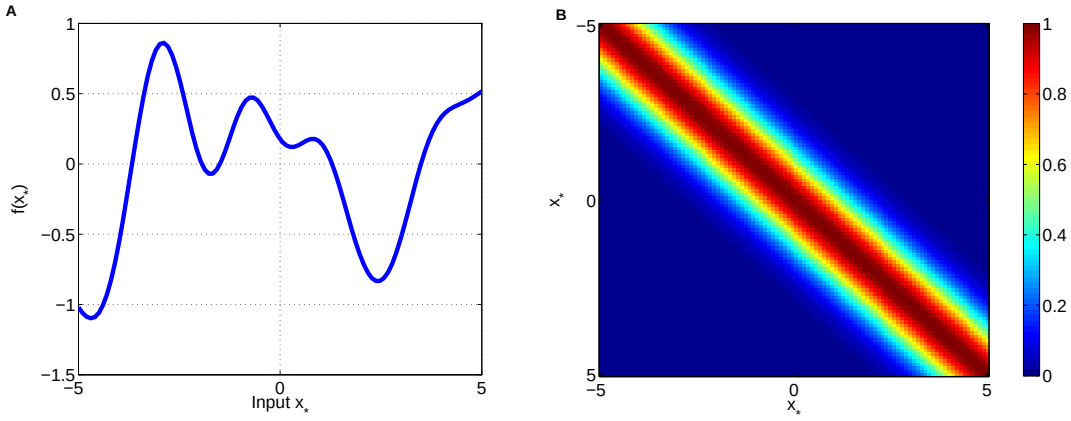


Figure 4.2: A sample function drawn from GP. **A** shows a function $f(x)$ drawn from Gaussian prior and the squared exponential covariance matrix between time points is shown in **B**, where $\sigma_r = 1$ and $l = 1$. The matrix is valued between 1 (red) and 0 (blue).

The conditional distribution is then given by

$$p(\mathbf{f}_*|\mathbf{f}) \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}), \tag{4.13}$$

with the mean and covariance function (derivations have been shown in Section 4.1.1 and Appendix A.5):

$$\mathbf{m} = \boldsymbol{\mu}_{\mathbf{f}_*} + \mathbf{K}_{\mathbf{f}_*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}(\mathbf{f} - \boldsymbol{\mu}_{\mathbf{f}}), \tag{4.14}$$

$$\mathbf{\Sigma} = \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*} - \mathbf{K}_{\mathbf{f}_*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},\mathbf{f}_*}. \tag{4.15}$$

Now the function $\mathbf{f}_*$ with test points $\mathbf{x}_*$ can be sampled by mean and covariance functions from the posterior distribution (shown above) by given the observations.

4.1.3 Noisy observations

In most cases, the realistic observations $\mathbf{y}$ are corrupted by noise. An easy way to deal with the noisy data in the GP framework is to add the independent identically distributed Gaussian noise ε, in which the covariance is defined by σ_n^2. We have

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon}, \quad \text{and} \tag{4.16}$$

$$\text{cov}(\mathbf{y}) = \mathbf{K}_{\mathbf{f},\mathbf{f}} + \sigma_n^2\mathbf{I}. \tag{4.17}$$

The joint probability of noisy observations and test function $\mathbf{f}$ is given by:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{f}} \\ \boldsymbol{\mu}_{\mathbf{f}_*} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{f},\mathbf{f}} + \sigma_n^2\mathbf{I} & \mathbf{K}_{\mathbf{f},\mathbf{f}_*} \\ \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} & \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*} \end{bmatrix} \right). \tag{4.18}$$

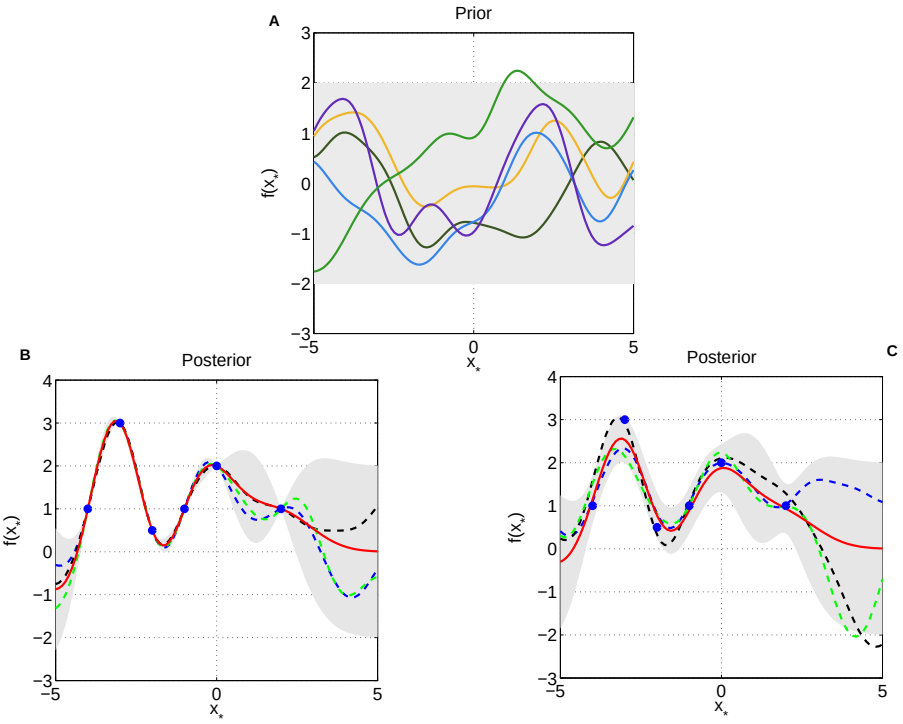


Figure 4.3: **A** shows five functions drawn from GP prior in different colours with input vector from $[-5, 5]$. **B** shows three time-varying functions (dashed lines) randomly drawn from the posterior which is conditioned on six noise free observations (blue points). The mean function is shown with red solid line. length scale l and signal standard deviation σ_r are both equal to 1. **C** shows three randomly drawn posterior functions (dashed lines) conditioned on the same six observation points with Gaussian noise ε. The mean function is shown with a red line. The hyperparameters l and σ_r remain in 1. In these figures, the grey area, known as the 95% confidence area, is defined by the mean function with two times the standard deviation.

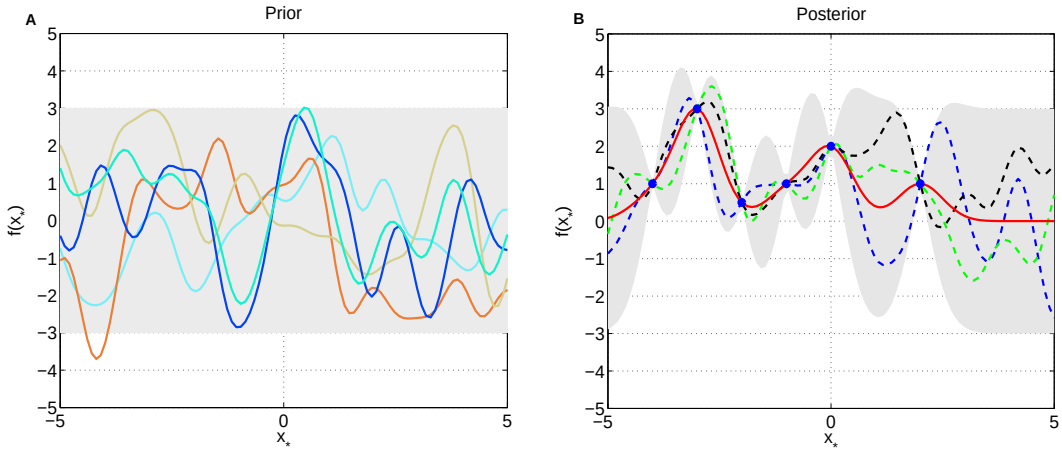


Figure 4.4: Random functions drawn from Gaussian prior (shown in **A**) and posterior (shown in **B**) with different hyperparameters: $l = 0.5$ and $\sigma_r = 1.5$.

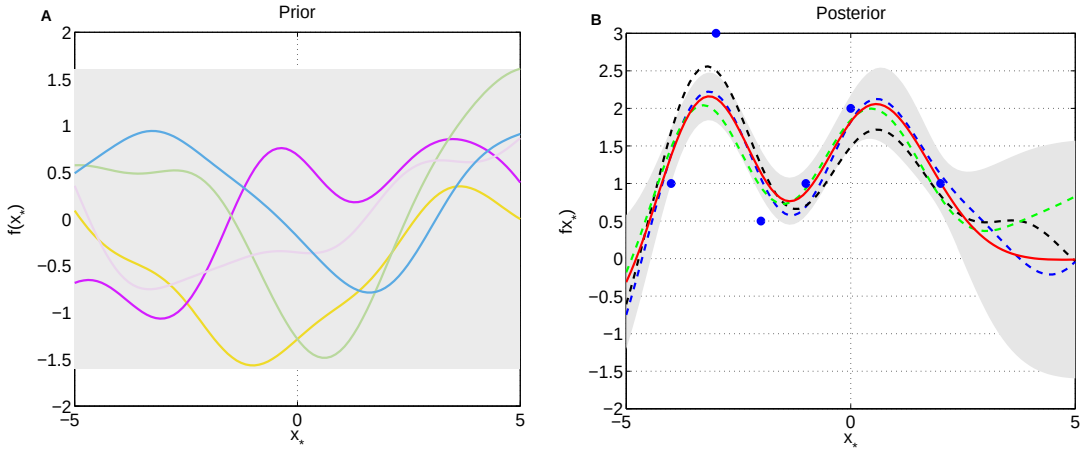


Figure 4.5: Random functions drawn from Gaussian prior (shown in **A**) and posterior (shown in **B**) with changed hyperparamters: $l = 0.5$ and $\sigma_r = 1.5$ and the noise standard deviation $\sigma_n = 0.2$.

The posterior of test function $\mathbf{f}_*$ given observations becomes:

$$p(\mathbf{f}_*|\mathbf{y}) \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}), \tag{4.19}$$

where

$$\mathbf{m} = \boldsymbol{\mu}_{\mathbf{f}_*} + \mathbf{K}_{\mathbf{f}_*,\mathbf{f}}(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \sigma_n^2\mathbf{I})^{-1}(\mathbf{f} - \boldsymbol{\mu}_{\mathbf{f}}), \quad \text{and} \tag{4.20}$$

$$\mathbf{\Sigma} = \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*} - \mathbf{K}_{\mathbf{f}_*,\mathbf{f}}(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \sigma_n^2\mathbf{I})^{-1}\mathbf{K}_{\mathbf{f},\mathbf{f}_*}. \tag{4.21}$$

4.1.4 Maximum likelihood estimation

Let θ be the parameters of $\mathbf{f}$. The marginal likelihood $p(\mathbf{y}|\theta)$ is given by

$$p(\mathbf{y}|\theta) = \int p(\mathbf{y}|\mathbf{f}, \theta)p(\mathbf{f}|\theta)d\mathbf{f}. \tag{4.22}$$

We already have the knowledge that prior $p(\mathbf{f}|\theta)$ is Gaussian, in which

$$p(\mathbf{f}|\theta) = (2\pi)^{-1/2}|\mathbf{K}_{\mathbf{f},\mathbf{f}}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{f} - \boldsymbol{\mu}_{\mathbf{f}})^T\mathbf{K}_{\mathbf{f},\mathbf{f}}{}^{-1}(\mathbf{f} - \boldsymbol{\mu}_{\mathbf{f}})\right). \tag{4.23}$$

Normally, the prior mean $\boldsymbol{\mu}_{\mathbf{f}} = 0$. The logarithm of the prior becomes

$$\log p(\mathbf{f}|\theta) = -\frac{1}{2}\mathbf{f}^T\mathbf{K}_{\mathbf{f},\mathbf{f}}{}^{-1}\mathbf{f} - \frac{1}{2}\log|\mathbf{K}_{\mathbf{f},\mathbf{f}}| - \frac{1}{2}\log(2\pi). \tag{4.24}$$

From Equations 4.16 and 4.17, the likelihood and probability density function are given by

$$p(\mathbf{y}|\mathbf{f}) = (2\pi)^{-1/2}|\sigma_n^2\mathbf{I}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^T(\sigma_n^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{f})\right). \tag{4.25}$$

Similarly, the logarithm of the likelihood is shown below:

$$\log p(\mathbf{y}|\mathbf{f},\theta) = -\frac{1}{2}(\mathbf{y}-\mathbf{f})^T(\sigma_n^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{f}) - \frac{1}{2}\log|\sigma_n^2\mathbf{I}| - \frac{1}{2}\log(2\pi). \tag{4.26}$$

Rewriting marginal likelihood integration in Equation 4.22, we have

$$\begin{aligned}\log p(\mathbf{y}|\theta) =& \log\int p(\mathbf{y}|\mathbf{f},\theta)p(\mathbf{f}|\theta)d\mathbf{f} \\ =& -\frac{1}{2}\mathbf{y}^T(\mathbf{K}_{\mathbf{f},\mathbf{f}}+\sigma_n^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}_{\mathbf{f},\mathbf{f}}+\sigma_n^2\mathbf{I}| - \frac{1}{2}\log(2\pi).\end{aligned} \tag{4.27}$$

The gradient of parameter θ can be obtained by

$$\begin{aligned}\frac{\partial\log p(\mathbf{y}|\theta)}{\partial\theta} =& -\frac{1}{2}\mathbf{y}^T(\mathbf{K}_{\mathbf{f},\mathbf{f}}+\sigma_n^2\mathbf{I})^{-1}\frac{\partial(\mathbf{K}_{\mathbf{f},\mathbf{f}}+\sigma_n^2\mathbf{I})}{\partial\theta}(\mathbf{K}_{\mathbf{f},\mathbf{f}}+\sigma_n^2\mathbf{I})^{-1}\mathbf{y} \\ & -\frac{1}{2}\mathrm{tr}\left((\mathbf{K}_{\mathbf{f},\mathbf{f}}+\sigma_n^2\mathbf{I})^{-1}\frac{\partial(\mathbf{K}_{\mathbf{f},\mathbf{f}}+\sigma_n^2\mathbf{I})}{\partial\theta}\right).\end{aligned} \tag{4.28}$$

The MLE can be obtained by solving

$$\frac{\partial\log p(\mathbf{y}|\theta)}{\partial\theta} = 0. \tag{4.29}$$

Normally, the closed-form solutions of such equations are intractable. In this regard, some numerical methods need to be resorted to, such as conjugate gradient (Shewchuk, 1994). In our work, the conjugate gradient is implemented by Rasmussen's minimise[1].

4.1.5 Gaussian process inference for time-varying model

From this section, the biological applications of GPs are introduced. We start with a toy problem, a time-varying example based on Gao et al. (2008)'s work. This model-based approach is popular to undertake inference and learning work, such as inferring the action of unobserved chemical species and learning the parameters from the model. A linear dynamical ODE system, as an toy problem here, is given by:

$$\dot{\mathbf{x}}(t) = -\mathbf{B}\mathbf{x}(t) + \mathbf{s}f(t), \tag{4.30}$$

where $\mathbf{x}(t) = [x_1(t), ..., x_N(t)]^T$, the concentrations of N genes driven by a latent function $f(t)$. In this biological system, the driven function $f(t)$ is not observed and we treat it

[1] http://www.gaussianprocess.org/gpml/code/matlab/util/minimize.m

as a latent function with a GP prior:

$$f(t) \sim \mathcal{N}(\mu(t), k_{f,f}(t,t')), \tag{4.31}$$

where $\mu(t)$ is the mean of $f(t)$ and $k_{f,f}(t,t')$ is the covariance function:

$$\mu(t) = E[f(t)] = 0, \tag{4.32}$$

$$\begin{aligned} k_{f,f}(t,t') &= E[(f(t)-\mu(t))(f(t')-\mu(t'))] \\ &= \exp\left(-\frac{(t-t')^2}{l^2}\right), \end{aligned} \tag{4.33}$$

and l is the time scale of the covariance function.

The model parameters, $\mathbf{s} = [s_1, ..., s_N]^T$ and decaying rate $\mathbf{B}$, a diagonal matrix with diagonal vector $[B_1, ..., B_N]^T$, and hyperparameters are estimated by MLE firstly. The solution of the model in Equation 4.30 is achieved by

$$\mathbf{x}(t) = \int_0^t \underbrace{\exp(-\mathbf{B}(t-u))}_{\text{matrix expontial}} \mathbf{s} f(u) du, \tag{4.34}$$

with the initial condition given by:

$$\mathbf{x}(0) = \mathbf{0}. \tag{4.35}$$

Because $\mathbf{x}(t)$ is a linear function of $f(t)$, it turns out to be a GP:

$$\mathbf{x}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\mathbf{x},\mathbf{x}}(t,t')) \tag{4.36}$$

where $\mathbf{K}_{\mathbf{x},\mathbf{x}}(t,t')$ is given by:

$$\mathbf{K}_{\mathbf{x},\mathbf{x}}(t,t') = \int_0^t \int_0^{t'} \exp(-\mathbf{B}(t-u))\mathbf{s}[\exp(-\mathbf{B}(t'-u'))\mathbf{s}]^T k_{f,f}(u,u') du du' \tag{4.37}$$

The cross covariance between $\mathbf{x}(t)$ and $f(t)$ is defined as:

$$\mathbf{k}_{\mathbf{x},f}(t,t') = \int_0^t \exp(-\mathbf{B}(t-u))\mathbf{s} k_{f,f}(u,t') du \tag{4.38}$$

From Equations 4.14 and 4.15, the posterior $p(f|\mathbf{x})$ of the latent function $f(t)$ given concentration $\mathbf{x}(t)$ becomes:

$$p(f|\mathbf{x}) \sim \mathcal{N}(\overline{f}^{\text{post}}, \mathbf{\Sigma}_{f,f}^{\text{post}}) \tag{4.39}$$

where

$$\overline{f}^{\text{post}} = \mathbf{K}_{f,\mathbf{x}}\mathbf{K}_{\mathbf{x},\mathbf{x}}^{-1}\mathbf{x} \tag{4.40}$$

$$\mathbf{\Sigma}_{f,f}^{\text{post}} = \mathbf{K}_{f,f} - \mathbf{K}_{f,\mathbf{x}}\mathbf{K}_{\mathbf{x},\mathbf{x}}^{-1}\mathbf{K}_{\mathbf{x},f} \tag{4.41}$$

The observation can be modelled at time t as:

$$y_i(t) = x_i(t) + e_i(t), \tag{4.42}$$

$$\mathbf{\Sigma}_{\mathbf{yy}} = \Sigma + \mathbf{K}_{\mathbf{xx}}, \tag{4.43}$$

where $e_i(t) \sim \mathcal{N}(0, \sigma_i^2(t))$ and $\Sigma = \text{diag}(\sigma_{11}^2, ..., \sigma_{1T}^2, ..., \sigma_{N1}^2, ..., \sigma_{NT}^2)$.

To predict the gene expression profiles corresponding to the testing time points, we let $\mathbf{f}_*$ be the latent function and $\mathbf{x}_i^*$ be the concentration of i-th gene. The testing vector $\mathbf{h}_* = [\mathbf{f}_*, \mathbf{x}_1^*, ..., \mathbf{x}_N^*]^T$. Similarly, we have the training vector containing observations given by $\mathbf{h} = [f, \mathbf{x}_1, ..., \mathbf{x}_N]^T$ with a fixed value for $f(0)$. Rewriting Equations 4.40 and 4.41, the mean and covariance matrix become:

$$\overline{\mathbf{h}}_*^{\text{post}} = \mathbf{K}_{\mathbf{h}_*,\mathbf{h}}\mathbf{K}_{\mathbf{h},\mathbf{h}}^{-1}\mathbf{y} \tag{4.44}$$

$$\mathbf{\Sigma}_{\mathbf{h}_*,\mathbf{h}_*}^{\text{post}} = \mathbf{K}_{\mathbf{h}_*,\mathbf{h}_*} - \mathbf{K}_{\mathbf{h}_*,\mathbf{h}}\mathbf{K}_{\mathbf{h},\mathbf{h}}^{-1}\mathbf{K}_{\mathbf{h},\mathbf{h}_*} \tag{4.45}$$

In this example, we created a five-gene model driven by a latent function $f(t)$ by adding a homogeneous noise. We inferred the functions by 10 noisy observations of each gene over 100 test time points. The inferred latent function $f(t)$ and five gene profiles are shown in Figure 4.6. By comparing the inferred and true functions, the GP inference framework provides good predictions of unobserved latent function $f(t)$ and gene concentrations given a few observations.

Model parameters ($\mathbf{B}$ and $\mathbf{s}$) and hyperparameters (l and σ_n) are estimated by maximising likelihood in Section 4.1.4. Results of the parameter maximum likelihood are shown in Figure 4.7. The covariance matrices are shown in Figures 4.8 – 4.10.

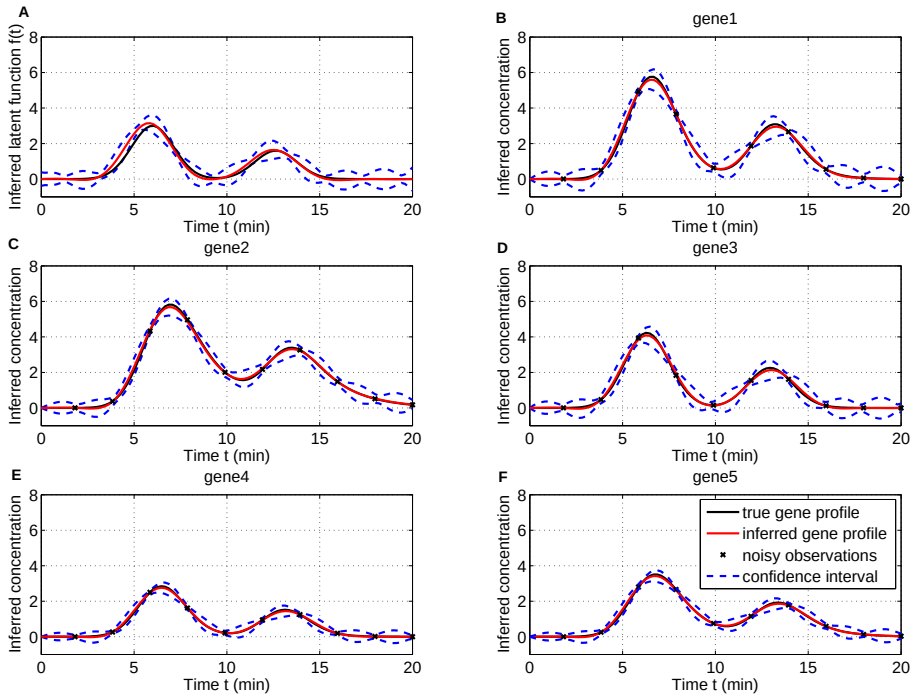


Figure 4.6: Inference for the toy model. Predictive mean functions are shown as red lines and the blue dashed lines are 95% confidence intervals. True values are shown as black lines in each sub-figure with noisy observations shown in black crosses. The latent function $f(t)$ is shown in Figure **A** while gene profiles are shown in Figures **B** – **F**.

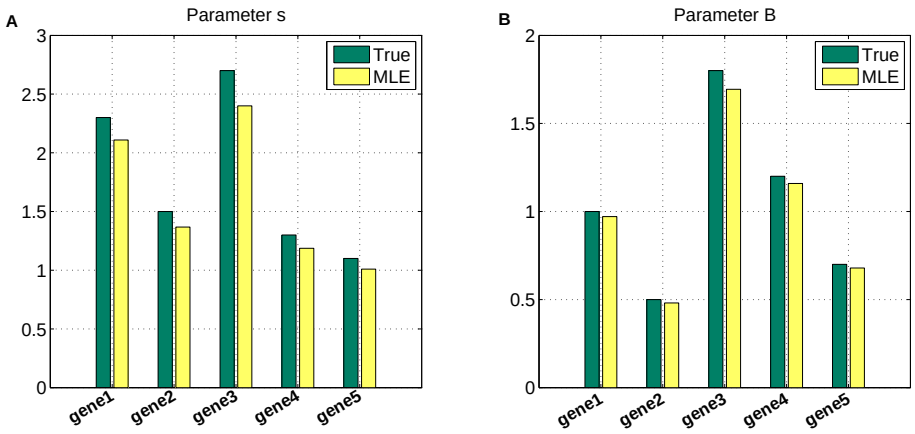


Figure 4.7: Model parameters estimated by maximum likelihood. The true values for **s** and **B** are shown in the green bars. Yellow bars show the corresponding estimated results for different genes ($\mathbf{x}(t)$).

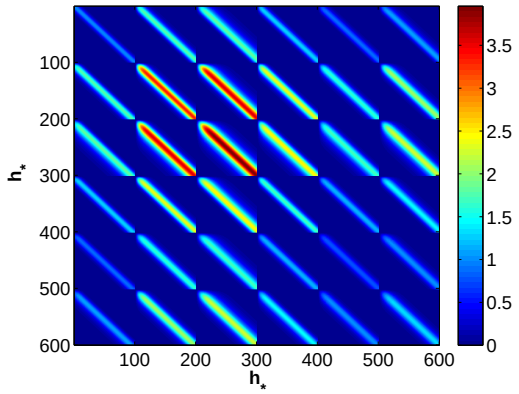


Figure 4.8: Covariance matrix $\mathbf{K}_{\mathbf{h}_*,\mathbf{h}_*}$ (600×600)

$$\mathbf{K}_{\mathbf{h}_*,\mathbf{h}_*} = \overbrace{\begin{bmatrix} \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*} & \mathbf{K}_{\mathbf{f}_*,\mathbf{x}_1^*} & \cdots & \mathbf{K}_{\mathbf{f}_*,\mathbf{x}_5^*} \\ \mathbf{K}_{\mathbf{x}_1^*,\mathbf{f}_*} & \mathbf{K}_{\mathbf{x}_1^*,\mathbf{x}_1^*} & \cdots & \mathbf{K}_{\mathbf{x}_1^*,\mathbf{x}_5^*} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{\mathbf{x}_5^*,\mathbf{f}_*} & \mathbf{K}_{\mathbf{x}_5^*,\mathbf{x}_1^*} & \cdots & \mathbf{K}_{\mathbf{x}_5^*,\mathbf{x}_5^*} \end{bmatrix}}^{100+5\times 100}$$

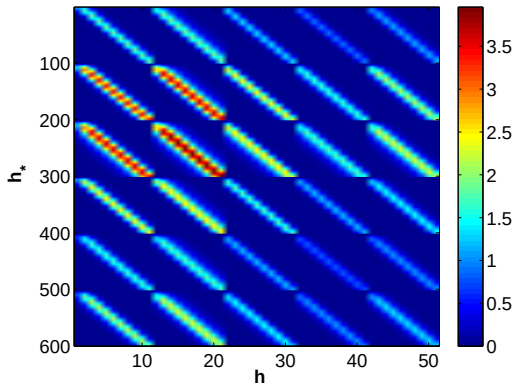


Figure 4.9: Covariance matrix $\mathbf{K}_{\mathbf{h}_*,\mathbf{h}}$ (600×51)

$$\mathbf{K}_{\mathbf{h}_*,\mathbf{h}} = \overbrace{\begin{bmatrix} \mathbf{K}_{\mathbf{f}_*,f} & \mathbf{K}_{\mathbf{f}_*,\mathbf{x}_1} & \cdots & \mathbf{K}_{\mathbf{f}_*,\mathbf{x}_5} \\ \mathbf{K}_{\mathbf{x}_1^*,f} & \mathbf{K}_{\mathbf{x}_1^*,\mathbf{x}_1} & \cdots & \mathbf{K}_{\mathbf{x}_1^*,\mathbf{x}_5} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{\mathbf{x}_5^*,f} & \mathbf{K}_{\mathbf{x}_5^*,\mathbf{x}_1} & \cdots & \mathbf{K}_{\mathbf{x}_5^*,\mathbf{x}_5} \end{bmatrix}}^{1+5\times 10}$$

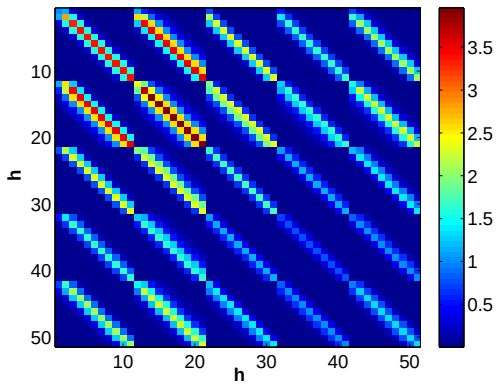


Figure 4.10: Covariance matrix $\mathbf{K}_{\mathbf{h},\mathbf{h}}$ (51×51)

$$\mathbf{K}_{\mathbf{h},\mathbf{h}} = \overbrace{\begin{bmatrix} \mathbf{K}_{\mathbf{f},f} & \mathbf{K}_{\mathbf{f},\mathbf{x}_1} & \cdots & \mathbf{K}_{\mathbf{f},\mathbf{x}_5} \\ \mathbf{K}_{\mathbf{x}_1,f} & \mathbf{K}_{\mathbf{x}_1,\mathbf{x}_1} & \cdots & \mathbf{K}_{\mathbf{x}_1,\mathbf{x}_5} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{\mathbf{x}_5,f} & \mathbf{K}_{\mathbf{x}_5,\mathbf{x}_1} & \cdots & \mathbf{K}_{\mathbf{x}_5,\mathbf{x}_5} \end{bmatrix}}^{1+5\times 10}$$

4.2 Inference of *bicoid* mRNA in spatio-temporal Bicoid profile

4.2.1 Bicoid linear dynamical system

we start from the Bicoid reaction-diffusion PDE which is the fundamental model in our work introduced in Chapter 3,

$$\frac{\partial}{\partial t}m(x,t) \;=\; D\,\frac{\partial^2}{\partial x^2}m(x,t) \;-\; \tau_p^{-1}\,m(x,t) \;+\; S_0 f(t), \tag{4.46}$$

where $f(t)$, the source, is the mRNA regulation function which we consider unknown in this chapter and place a prior distribution over. A linear gain term S_0 is included to allow scaling of the data to match observations.

As in the case of the stochastic model we discussed in Chapter 3, we consider the embryo as consisting of N cubes along the A-P axis in order to derive a linear dynamical system model from the continuous spatial diffusion equation. The basic idea for this stems from the work of Erban et al. (2007). With this discretisation into N cubes, the chemical reactions involved in the diffusion process are:

$$\text{Bicoid}_i \underset{d}{\overset{d}{\rightleftharpoons}} \text{Bicoid}_{i+1}, \quad \texttt{for} \quad i = 1, 2, \ldots, N-1 \tag{4.47}$$

$$\text{Bicoid}_i \xrightarrow{\tau_p^{-1}} \emptyset, \quad \texttt{for} \quad i = 1, 2, \ldots, N \tag{4.48}$$

$$\emptyset \xrightarrow{S_0 f(t)} \text{Bicoid}_1. \tag{4.49}$$

The first of these reactions, Equation 4.47, is Bicoid protein diffusion between neighbouring sub-volumes with rate constant d, where d is given by $d = D/h^2$ and h is length of each cube. The second process in Equation 4.48 describes Bicoid protein degradation. Finally, Equation 4.49 is the translation of Bicoid proteins from the maternal mRNA with $f(t)$ being the latent function that needs to be inferred.

We implemented a model in which source production occurs in a smaller first cube with length h_f ($5\mu m$ – the length of a nucleus) and the other $N-1$ cubes ($h = 10\mu m$) equally splitting the remaining A-P axis. The rate constants are different between the first cube (d_f), where mRNA is produced and its stability regulated, and the other cubes (d):

$$d_f \approx D/(h_f h), \tag{4.50}$$

$$d = D/h^2. \tag{4.51}$$

More details of the diffusion rate in the compartments with different sizes can be found in (Engblom et al., 2008).

In order to develop a linear dynamical system for Bicoid profile, following Erban et al. (2007), we rewrite the partial differential equation in Equation 4.46 as a system of ODEs for the morphogen concentration in each bin ($i = 1, ..., N$):

$$\frac{\partial}{\partial t}m_1(t) = d_f(m_2(t) - m_1(t)) - \tau_p^{-1}m_1(t) + S_0 f(t), \tag{4.52}$$

$$\frac{\partial}{\partial t}m_i(t) = d(m_{i+1}(t) + m_{i-1}(t) - 2m_i(t)) - \tau_p^{-1}m_i(t), \tag{4.53}$$

$$\frac{\partial}{\partial t}m_N(t) = d(m_{N-1}(t) - m_N(t)) - \tau_p^{-1}m_N(t). \tag{4.54}$$

The derivation details are shown in Appendix A.4.

Defining $\mathbf{m}(t) = [m_1(t), ..., m_N(t)]^T$, the linear dynamical system for the Bicoid reaction-diffusion system is then vectorised as:

$$\frac{\partial \mathbf{m}(t)}{\partial t} = \mathbf{A}\mathbf{m}(t) + \mathbf{s}f(t), \tag{4.55}$$

where the spatial transition matrix $\mathbf{A}$ $(N \times N)$ is defined by:

$$\begin{bmatrix} -(d_f + \tau_p^{-1}) & d_f & 0 & \cdots & 0 & 0 \\ d & -(2d + \tau_p^{-1}) & d & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & d & -(d + \tau_p^{-1}) \end{bmatrix}$$

and source production rate $\mathbf{s} = [S_0, 0, \ldots, 0]^T$.

The solution to Equation 4.55, $\mathbf{m}(t)$, giving the Bicoid spatio-temporal profile, is in terms of a matrix exponential and is given by:

$$\mathbf{m}(t) = \exp(t\mathbf{A})\mathbf{m}(0) + \int_0^t \exp((t-u)\mathbf{A})\mathbf{s}f(u)du, \tag{4.56}$$

where $\mathbf{m}(0)$ is zero at the beginning of the embryo's development.

4.2.2 Gaussian process modelling

We treat *bicoid* mRNA as a function drawn from a GP and extend it to the spatio-temporal application of the Bicoid dynamical system. The GP prior for the latent mRNA regulation is defined by mean and covariance functions:

$$f(t) \sim N(0, k_{f,f}(t, t')), \tag{4.57}$$

where $k_{f,f}(t,t')$ is given by squared exponential covariance function with the length scale l:

$$k_{f,f}(t,t') = \exp\left(-\frac{(t-t')^2}{l^2}\right). \tag{4.58}$$

Because the right-hand side of Equation 4.56 is linear, as noted in the previous section, $\mathbf{m}(t)$ turns out to be a multivariate function drawn from a GP:

$$\mathbf{m}(t) \sim N(\exp(t\mathbf{A})\mathbf{m}(0), \mathbf{K}_{\mathbf{m},\mathbf{m}}(t,t')). \tag{4.59}$$

The corresponding cross-covariance function $\mathbf{K}_{\mathbf{m},\mathbf{m}}(t,t')$ is given by:

$$\mathbf{K}_{\mathbf{m},\mathbf{m}}(t,t') = \tag{4.60}$$
$$\int_0^t \int_0^{t'} \exp((t-u)\mathbf{A})\mathbf{s}(\exp((t'-u')\mathbf{A})\mathbf{s})^T k_{f,f}(u,u')dudu'.$$

The cross covariance-function between $\mathbf{m}(t)$ and $f(t)$ becomes

$$\mathbf{k}_{\mathbf{m},f}(t,t') = \int_0^t \exp((t-u)\mathbf{A})\mathbf{s}k_{f,f}(u,t')du. \tag{4.61}$$

These expressions can be derived analytically and the derivation details are shown in Appendix A.6.

4.2.3 Predictive distribution

Let $\mathbf{f}_*$ be a vector of J_* values of the source function at equally spaced time points, and $\mathbf{m}_i^*$ be the corresponding protein profiles at these instances in time, within the i^{th} cube along the A-P axis. Concatenating these we define $\mathbf{h}_* = [\mathbf{f}_*, \mathbf{m}_1^*, ..., \mathbf{m}_N^*]^T$ corresponding to $J_* + NJ_*$ test points. The corresponding training data consists of the morphogen values in the N cubes, taken at J points in time, and a fixed source value f, contained in the vector $\mathbf{h} = [f, \mathbf{m}_1, ..., \mathbf{m}_N]^T$ of dimension $1 + NJ$.

With the above notation, the mean and covariance of the posterior distribution are given by:

$$\bar{\mathbf{h}}_*^{post} = \mathbf{K}_{\mathbf{h}_*,\mathbf{h}}\mathbf{K}_{\mathbf{h},\mathbf{h}}^{-1}\mathbf{y}, \tag{4.62}$$

$$\mathbf{\Sigma}_{\mathbf{h}_*,\mathbf{h}_*}^{post} = \mathbf{K}_{\mathbf{h}_*,\mathbf{h}_*} - \mathbf{K}_{\mathbf{h}_*,\mathbf{h}}\mathbf{K}_{\mathbf{h},\mathbf{h}}^{-1}\mathbf{K}_{\mathbf{h},\mathbf{h}_*}. \tag{4.63}$$

Each covariance matrix in Equations 4.62 and 4.63 is partitioned across the source and N Bicoid intensities in different cubes. Illustrating this for $\mathbf{K}_{\mathbf{h}_*,\mathbf{h}}$,

$$\mathbf{K}_{\mathbf{h}_*,\mathbf{h}} = \begin{bmatrix} \mathbf{K}_{\mathbf{f}_*,f} & \mathbf{K}_{\mathbf{f}_*,\mathbf{m}_1} & \mathbf{K}_{\mathbf{f}_*,\mathbf{m}_2} & \cdots & \mathbf{K}_{\mathbf{f}_*,\mathbf{m}_N} \\ \mathbf{K}_{\mathbf{m}_1^*,f} & \mathbf{K}_{\mathbf{m}_1^*,\mathbf{m}_1} & \mathbf{K}_{\mathbf{m}_1^*,\mathbf{m}_2} & \cdots & \mathbf{K}_{\mathbf{m}_1^*,\mathbf{m}_N} \\ \mathbf{K}_{\mathbf{m}_2^*,f} & \mathbf{K}_{\mathbf{m}_2^*,\mathbf{m}_1} & \mathbf{K}_{\mathbf{m}_2^*,\mathbf{m}_2} & \cdots & \mathbf{K}_{\mathbf{m}_2^*,\mathbf{m}_N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{\mathbf{m}_N^*,f} & \mathbf{K}_{\mathbf{m}_N^*,\mathbf{m}_1} & \cdots & \cdots & \mathbf{K}_{\mathbf{m}_N^*,\mathbf{m}_N} \end{bmatrix}$$

Therefore, the dimensions of the covariance matrices $\mathbf{K}_{\mathbf{h}_*,\mathbf{h}}$, $\mathbf{K}_{\mathbf{h},\mathbf{h}}$ and $\mathbf{K}_{\mathbf{h}_*,\mathbf{h}_*}$ are $(J_* + NJ_*) \times (1 + NJ)$, $(1 + NJ) \times (1 + NJ)$ and $(J_* + NJ_*) \times (J_* + NJ_*)$ respectively. The observations, collected in vector $\mathbf{y}$ of dimension $1 + NJ$, are assumed to be corrupted by additive noise as

$$y_i(t) = m_i(t) + e_i(t), \tag{4.64}$$

where $e_i(t)$ are drawn from $N(0, \sigma_i^2(t))$. Hence,

$$\mathbf{\Sigma}_{\mathbf{y},\mathbf{y}} = \mathbf{K}_{\mathbf{h},\mathbf{h}} + \mathbf{\Sigma}, \tag{4.65}$$

$$\mathbf{\Sigma} = \text{diag}(\sigma_f^2, \sigma_{11}^2, \ldots, \sigma_{1J}^2, \ldots, \sigma_{N1}^2, \ldots, \sigma_{NJ}^2). \tag{4.66}$$

With the hyperparameter vector $\boldsymbol{\theta}_h = [l, \sigma_f^2, \sigma_{11}^2, \ldots, \sigma_{1J}^2, \ldots, \sigma_{N1}^2, \ldots, \sigma_{NJ}^2]$, the likelihood is:

$$p(\mathbf{y}|\boldsymbol{\theta}_h) = \int p(\mathbf{y}|\boldsymbol{\theta}_h, \mathbf{f}) p(\mathbf{f}|\boldsymbol{\theta}_h) d\mathbf{f}. \tag{4.67}$$

The observations are taken from `FlyEx` database as the data used in Chapter 3. To get the estimate of the noise levels and the length scale l, we maximised the above likelihood following Section 4.1.4 and the results are given later in Section 4.3. When simulating synthetic data, we added zero mean Gaussian noise of standard deviation $\sigma = 0.1$, and assumed it to be known.

In this chapter, we set the model parameters by least squares fitting following Chapter 3, Equations 3.50 – 3.52. Ideally, one would like to exploit the elegance of the GP framework and estimate all of these by maximum likelihood. To achieve this, we need analytical expressions for the covariance matrices and their derivatives. While the covariance matrices of interest can be derived analytically in Appendix A.6, we encountered serious numerical issues in evaluating these expressions. This arises from the finite precision representation within `MATLAB` because the error function `erf(.)` is not evaluated to sufficient precision in the range of input values of its arguments that we needed[2]. The alternate approach we tried was to work with MPFR library written in `C` which has a

[2] e.g. `erf(9.5) = erf(10) = 1`

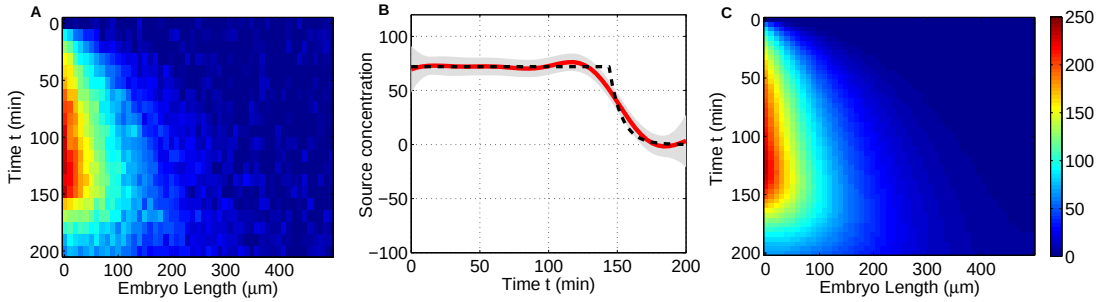


Figure 4.11: Inferred *bicoid* mRNA regulation and spatio-temporal protein concentration from a synthetic dataset; time scale from $0 \sim 200$ min in 51 cubes along the A-P axis. **A** training datasets from Bicoid reaction-diffusion ODE simulation with additive noise. **B** inferred mRNA regulation function (red line) with 95% confidence interval. Source function used in the simulation is shown as a black dashed line. **C** posterior mean GP approximation of the spatio-temporal profile.

MATLAB interface[3][4]. The results of this, in evaluating the covariance matrices, were consistent with the numerical evaluations reported in the main text. However, computing time for this toolbox was excessive, and we abandoned this approach.

The above are attempts at using the analytic expressions for evaluation of the covariance matrices. However, what we really need, to estimate the model parameters by maximum likelihood, are the derivatives of the expressions. These happen to include eigenvectors (Equations A.80 and A.89), derivatives of which we considered a significant distraction from the main point of this study. Therefore, in our work, model parameters D and τ_p are taken from Chapter 3 while the estimation of l and σ are implemented by Rasmussens minimise.

4.3 Results and discussion

4.3.1 Inference of mRNA regulation function

We first assume that *bicoid* mRNA is localised and its regulation occurs only in the anterior pole, the first cube in our model. Therefore, the source production amplitude vector is given by

$$\mathbf{s} = [S_0, 0, \ldots, 0]^T. \tag{4.68}$$

We examine the performance of our GP approach on two synthetic datasets (Figures 4.11 and 4.12) and an experimental dataset (Figure 4.13).

[3] http://www.mathworks.com/matlabcentral/fileexchange/6446
[4] http://www.mpfr.org/; http://gmplib.org/

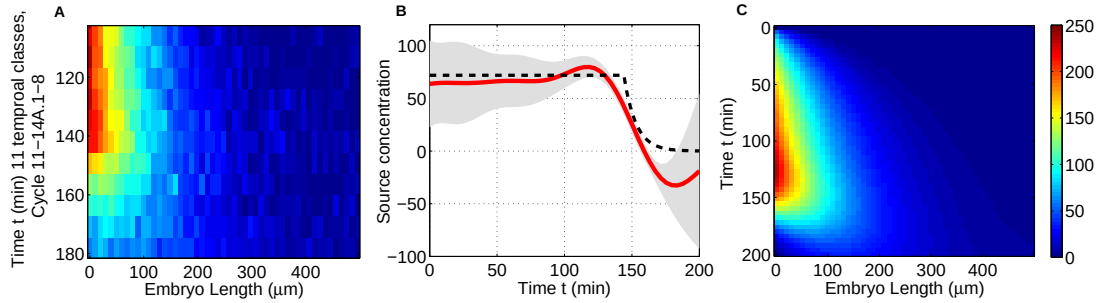


Figure 4.12: Predicted results for *bicoid* mRNA and Bicoid spatio-temporal profile using only partial data (106 – 178 min): cycles 11 – 13 and Cycle 14*A* class 1 – 8. **A** partial data used in training. **B** and **C** inferred source and spatio-temporal profiles respectively.

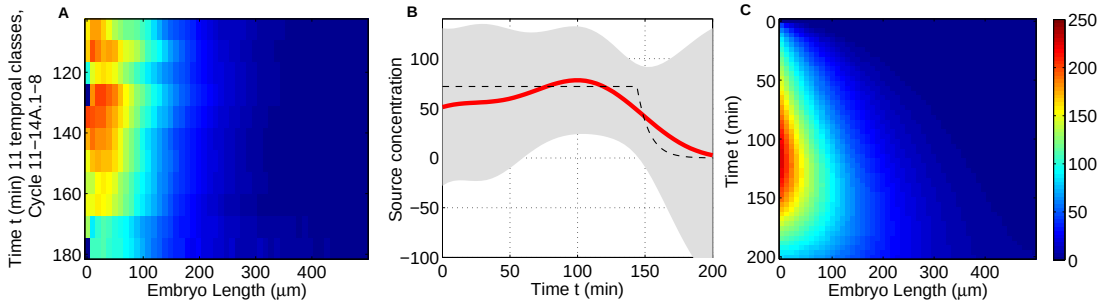


Figure 4.13: Source function and Bicoid profile inferred from `FlyEx` database. **A** average profile of the real data with time scale: cycles 11 – 13 and Cycle 14*A* class 1 – 8. **B** inferred source and **C** Bicoid concentration over the whole time scale (0 –200 min). The dashed line in **B** is the assumed source function.

As was the case in Chapter 3, to generate synthetic data, we implemented mRNA stability regulation by the function

$$f(t) = \delta(x)\ (\Theta(t) - \Theta(t - t_0)) + \delta(x)\,\Theta(t - t_0)\ \exp\left(-\frac{t - t_0}{\tau_m}\right), \tag{4.69}$$

which characterises mRNA, and hence the production of Bicoid protein, to be stable and constant to time t_0, followed by an exponential decay of time constant τ_m. The parameter values taken from Table 3.2, estimated by a least squares fit between model output and `FlyEx` measurements, are $t_0 = 144$ min and $\tau_m = 9$ min. The black dashed lines in Figures 4.11B, 4.12B and 4.13B show the true source function according to our hypothesis on regulation.

Figure 4.11A shows the synthetic training data using the ODE system of equations 4.52 – 4.54 with additive noise over the entire developmental period of 0 – 200 min, with 20 equally spaced time points and 51 cubes along the A-P axis. Figure 4.11B shows the estimated mRNA regulation function, $S_0 f(t)$, from the GP approach. We see that the GP is able to recover the regulated source function quite well, though the resulting

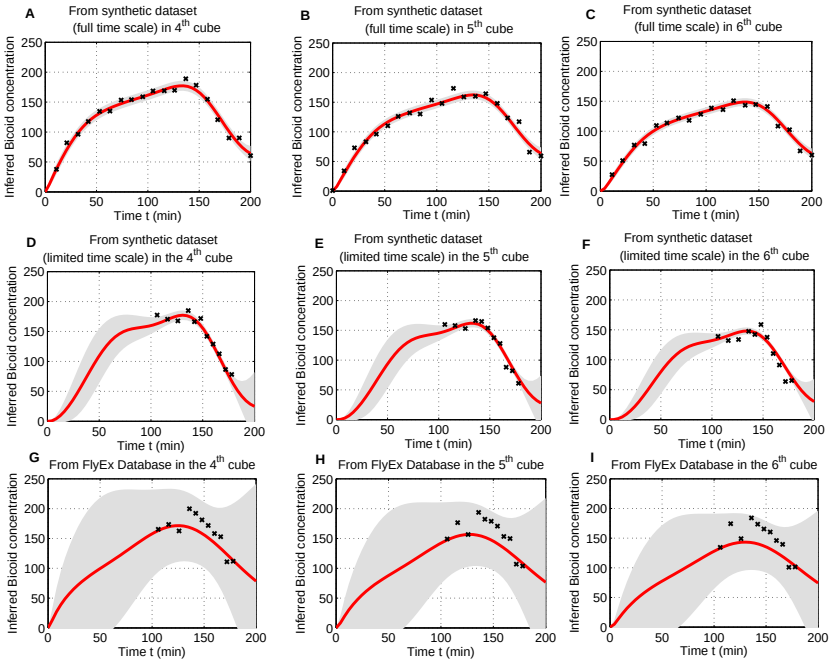


Figure 4.14: Predicted temporal posterior distribution of Bicoid protein concentrations in individual cubes on different training datasets. The mean inference and 95% confidence intervals are shown with red solid lines and grey area. The black crosses represent three different training datasets shown in Figures 4.11, 4.12 and 4.13. Figure 4.14**A**-**C** are inferred on the synthetic dataset with full time scale while the partial synthetic dataset is used in Figure s4.14**D**-**F**. Figures 4.14**G**-**I** are inferred on real dataset from `FlyEx`.

estimate is smoother. This is to be expected from a GP model, for which a function with a sharp discontinuity will have very low likelihood in the prior. Still, the decay beyond 2 hours is very rapid. The posterior mean of the inferred Bicoid concentration profile from the model is shown in Figure 4.11C. The temporal dynamics of a morphogen gradient being established and then killed off is clearly present in the model output.

In the above, shown in Figure 4.11, we have used the synthetic data over the full developmental time scale of interest. However, in `FlyEx` dataset, we do not have measurements available over the whole time scale and the source has to be inferred from partial data, starting from 100 min. In order to simulate this situation with synthetic data, we ran our GP models with only the partial data, shown in Figure 4.12A as input. As expected, in Figure 4.12B, the credible interval is wider at the early stages where no data are present and narrow during 106 – 178 min. Still, the GP posterior of the morphogen profile captures the spatio-temporal dynamics well and contains the sharp post-peak decay.

Figure 4.13 shows the behaviour of the GP model on real data from `FlyEx`, the three subfigures A, B and C showing the data, inferred source and model-based spatio-temporal

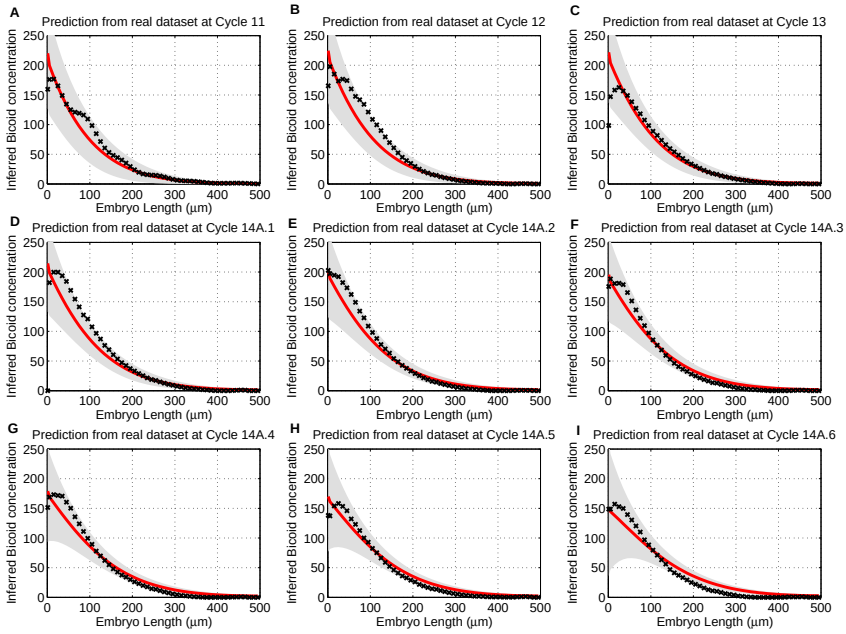


Figure 4.15: Inferred posterior distribution of spatial Bicoid protein concentrations in the fixed time points from Cycle 11 (Figure 4.15**A**) to Cycle 14*A*.6 (Figure 4.15**I**) on the real dataset. The black crosses show the observed protein spatial data in different developmental time points.

morphogen profile respectively. We see that the reconstructed profile accurately reproduces the establishment of Bicoid gradient and subsequent decay. The inferred source function is smoother than our hypothesised model, but contains the basic elements of an approximately constant part and subsequent decay. The noisy nature of the data causes the resulting uncertainties to be very high. Noise levels inferred from a maximum likelihood setting (Equation 4.67) is shown in Figure 4.16. These are much higher than the variance of the additive noise we used to construct the synthetic data of Figures 4.11 and 4.12. Further, we note that the source of uncertainty in the data is not purely additive instrument noise. `FlyEx` measurements do not come from observations on a single embryo. They are taken from populations of embryos, harvested at various stages of development. The effect of this is not modelled anywhere in our approach.

As noted, the GP-inferred source functions are smoother than hypothesised by our model. A consequence of smooth functions fitting the data is also that the temporal point at which mRNA begins to decay starts earlier. The rapid change between mRNA being translated and killed off is not explicitly modelled in the GP approach. Such rapid changes may well be better modelled in a probabilistic framework that explicitly incorporates switching behaviour, such as the two-state Markov Jump process (also known as a telegraph process) considered in (Sanguinetti et al., 2009).

Figure 4.14 shows the predicted temporal profiles of the Bicoid at three adjacent spatial points along the embryo[5]. The training data are also shown. The three rows in the figure correspond to cases illustrated in Figures 4.11, 4.12 and 4.13. We see that at the level of the GP model generating the data, reasonably good fits are obtained. With real data, we see high uncertainties in regions where data are not present.

4.3.2 Hyperparameter estimation

Since the hyperparameters, length scale l and noise standard deviation σ, are not the elements of matrix $\mathbf{A}$ and the corresponding derivatives are easily obtained, we estimate these parameters by maximum likelihood similar as the case in Section 4.1.5.

The estimated length scales were similar on the different training datasets used: 51.0 for synthetic data with full time scale (Figure 4.11), 50.3 for partial synthetic data (Figure 4.12) and 51.7 for real dataset (Figure 4.13). Estimated noise standard deviations $\{\sigma_{i,j}\}_{i=1,j=1}^{N\ J}$ (Equation 4.66) for the real dataset are shown in Figure 4.16. We note that the Bicoid expression data show spatio-temporal dependence and the noise levels we estimated, shown in Figure 4.16, are different at each point. The noise levels in the anterior part are much higher than in the posterior part of the embryo because the protein intensities are exponential decaying. In addition, such a noise level becomes higher in later stage, around $180\,min$, due to the measurements are not accurate in Cycle $14A.7$ and 8 in `FlyEx`.

To study the effect of non-homogeneous noise on synthetic data, we re-synthesised data with a noise profile similar to what was inferred from the real data (Figure 4.16), tapering down linearly from 0.4 to 0.1 in standard deviation along the A-P axis and repeated the estimation procedure. Figures 4.17 and 4.18 show the results of these.

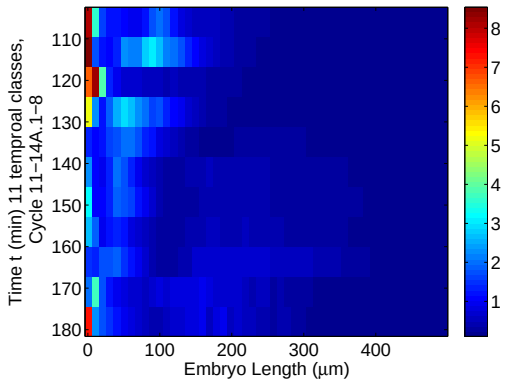


Figure 4.16: Estimated noise level at each spatio-temporal point. The intensities are shown to a logarithm scale.

[5]These cubes are chosen for illustration because these are locations where much of the variation is happening

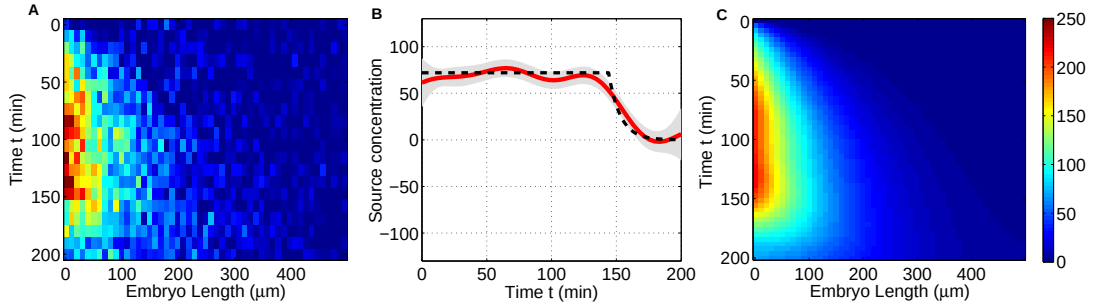


Figure 4.17: Inferred *bicoid* mRNA regulation and spatio-temporal protein concentration from a synthetic dataset with non-homogeneous noise; time scale from 0 – 200 *min* in 51 cubes along the A-P axis. **A** training datasets from Bicoid reaction-diffusion ODE simulation with additive noise. **B** inferred mRNA regulation function (red line) with 95% confidence interval. Source function used in the simulation is shown with black dashed line. **C** posterior mean GP approximation of the spatio-temporal profile.

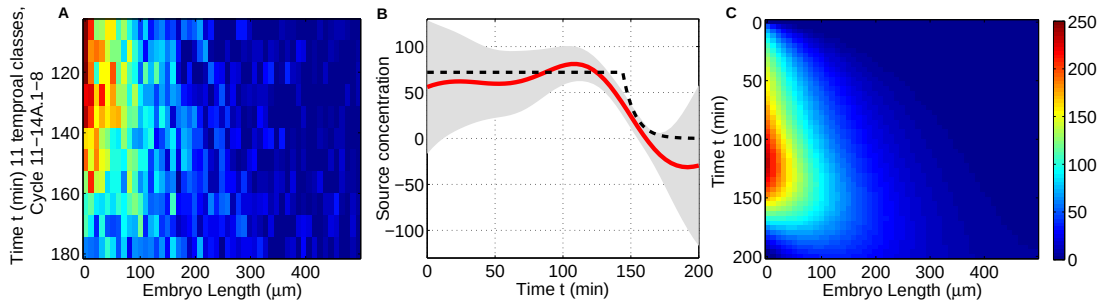


Figure 4.18: Predicted results for *bicoid* mRNA and Bicoid spatio-temporal profile using only partial data (106 – 178 *min*): cycles 11 – 13 and Cycle 14*A* class 1 – 8. **A** partial data used in training with non-homogeneous noise. **B** and **C** inferred source and spatio-temporal profiles respectively.

4.3.3 Inference of non-localised maternal mRNA

Additionally, in Figure 4.15 we show cross sections of the spatio-temporal profiles, taken along the A-P axis at different developmental cycles. Here, we see that the exponential spatial decays of `FlyEx` measurements are faithfully captured by the GP model. We also observe that most of the mismatch between model output and measured data is towards the anterior part of the A-P axis. This mismatch motivates one to question the use of a highly localised point source as the input to the diffusion system.

As reviewed in Chapter 2, Spirov et al. (2009) and Little et al. (2011) have discussed the scenario in which maternal *bicoid* mRNA itself has a spatial gradient. We also simulated this possibility in our GP models, with maternal mRNA being spatially distributed in the first 10 of the 50 cubes ($h = 10\mu m$) with an initial exponentially decaying spatial profile. The corresponding results of the inferred mean source (now a spatio-temporal

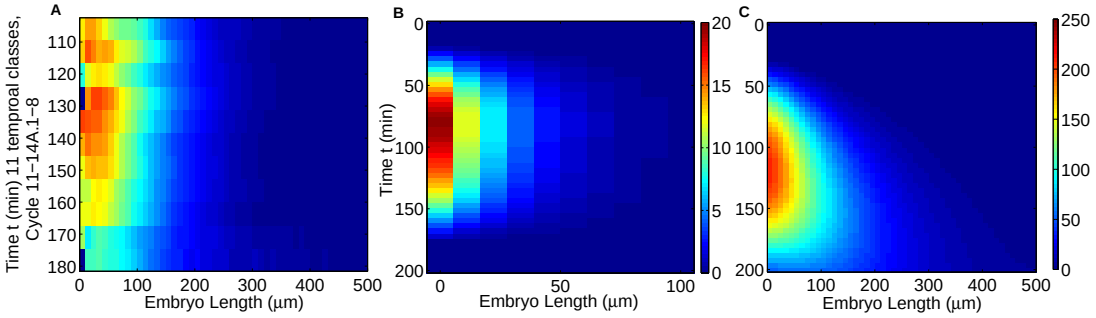


Figure 4.19: Inferred mRNA regulation and protein profile on real dataset from `FlyEx`. mRNA spatially distributed over 20% EL (10 cubes) rather than being localised. **A** training data from `FlyEx` database. **B** inferred disperse source in the first 10 cubes along developmental time. **C** inferred Bicoid profile.

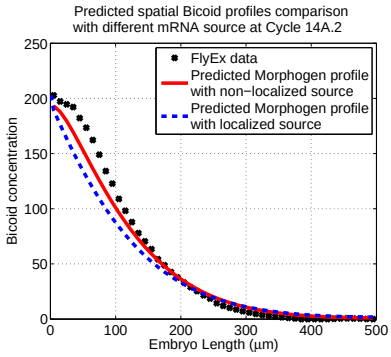


Figure 4.20: Inferred spatial protein profile with spatially distributed *bicoid* mRNA (red line) on real dataset (crosses) from `FlyEx` at the fixed developmental stage of Cycle 14*A*.2. For comparison, the predicted protein profile with localised mRNA is shown as a dashed line.

profile) and the GP mean morphogen profile are shown in Figure 4.19B and C. We note that in this case, the onset of mRNA decay begins slightly earlier than for the point source in the first bin, which is to be expected since the mRNA spatial distribution contributes to the generation of morphogen upto 20% EL, and the destruction has to start earlier to compensate. However, since the training dataset is partially observed and additional degree of freedom is introduced in this non-localised source model, the over-fitting problem occurs and the predicted results are negative during the beginning and end of development (0 – 20 min & 170 – 200 min in Figure 4.19B and 0 – 30 min in Figure 4.19C). Only the positive values are shown here.

As seen in Figure 4.20, the fit to the data does improve with spatially distributed mRNA. We include this for completeness, showing that a GP model can be applied in a flexible way in this manner, but do not think that the results can resolve the differences discussed by Spirov et al. (2009) and Little et al. (2011).

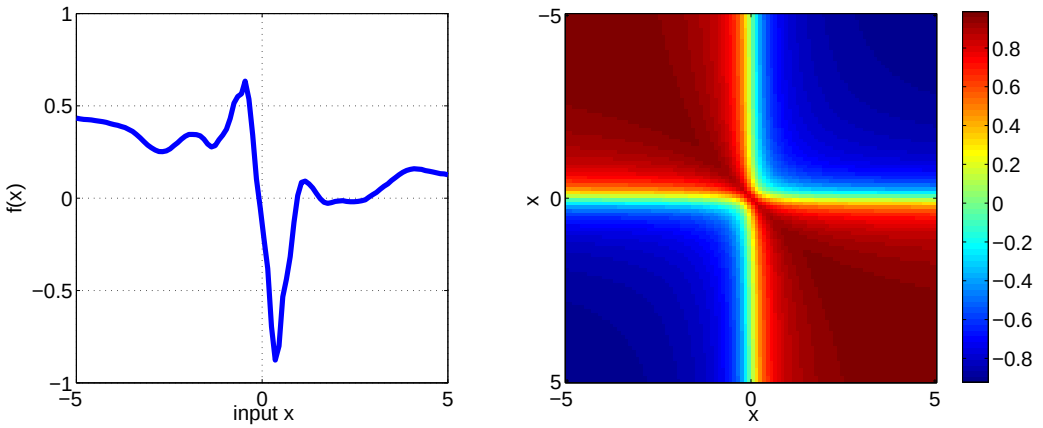


Figure 4.21: A sample drawn from the neural network covariance function with $\Sigma = \text{diag}([3^2, 10^2])$ is shown in the left panel and the corresponding covariance matrix is shown in the right panel.

4.3.4 Model and kernel selection

It is evident from our empirical results that GP with RBF kernel is not suitable for the inference of *bicoid* mRNA regulation. Functions based on RBF kernel are tend to be smooth, which are difficult to fit our hypothesis or the ground truth in the synthetic data. In turn, the prediction results are also not appealing. Within the GP framework, it is clear that one should explore non-stationary covariance functions such as neural network kernel (Rasmussen and Williams, 2006). This covariance function is given by

$$k_{NN}(\mathbf{x}, \mathbf{x}') = \frac{2}{\pi} \sin^{-1}\left(\frac{2\tilde{\mathbf{x}}^\top \Sigma \tilde{\mathbf{x}}'}{\sqrt{(1 + 2\tilde{\mathbf{x}}^\top \Sigma \tilde{\mathbf{x}})(1 + 2\tilde{\mathbf{x}}'^\top \Sigma \tilde{\mathbf{x}}')}}\right), \tag{4.70}$$

where $\tilde{\mathbf{x}} = (1, x_1, x_2, ..., x_d)^\top$ is an augmented input vector and $\Sigma = \text{diag}([\sigma_0^2, \sigma^2])$. It is notable that this kernel can produce functions with sharp transition, which could potentially match our hypothesis. An example is shown in Figure 4.21. However, it is not straightforward to integrate this kernel within our current GP framework.

When dealing with prediction, it is very likely that results from a single model choice is not good enough. It might be worthwhile to combine results from multiple models by Bayesian model averaging. On the other hand, it is also possible that our biological hypothesis is not true *in vivo*. The source function might be in fact behaves smoothly. Thus, assumptions on both modelling and biology shall be refined. An iterative process of Bayesian modelling and wet-lab experiment, might be preferred (Xu et al., 2010).

4.4 Summary

In this chapter, we have shown that the non-parametric GP regression model can be applied to the problem of inferring biologically useful information from the spatio-temporal distribution of the Bicoid morphogen in early *Drosophila* embryogenesis. Discretisation of the spatial domain, transforms the spatio-temporal problem into a dynamical system for which, with a GP prior imposed on the source, the solution can be obtained as a matrix exponential. With synthetic data obtained from a linear spatio-temporal dynamical system, our results show that the GP approach is able to recover the driving input and model the Bicoid distribution. On real world data, our results estimate a rough version of the driving input due to the data being available only during part of the developmental process, yet, the part of the source decay is estimated. In addition, such a smoothed source function is predicted in our work because the covariance function we used is RBF, which provides smooth functions and the sharp change of our regulation hypothesis is hard to capture. As result, some shortcomings have been shown when the data are partially available, such as the negative values of the predicted source function occurring in Figure 4.12, which is unrealistic for the protein concentration and the confidence interval is under estimated in the sharp decaying region. Moreover, the over-fitting problem occurs in Figure 4.19, when the source spatial diffusion has been considered.

These results imply that even with the sophisticated modelling and inference strategy, such as GP in our work, it is still essential to collect data as much as possible. Secondly, more kernel functions, alternative to RBF, could be considered in different applications in order to capture various driving functions. For prediction purposes, it is advisable to average out parameters and hyper-parameters with Bayesian approaches, to avoid over-fitting problem.

Chapter 5

Gap gene regulatory network with realistic Bicoid input

In this chapter we discuss an analysis of the gap gene model (reviewed in Section 2.4) with time-varying Bicoid input. First, we implement and exactly reproduce the results claimed in (Reinitz and Sharp, 1995; Jaeger et al., 2004b,a) using their model and the parameters they published. We then replace the regulatory Bicoid input, which they set to be constant, by a time-varying input taken from `FlyEx`. This replacement necessitates a different set of values for the parameters of the model. We show that this new set of parameter values can be obtained by a local search using Jaeger's published values as initial guess.

5.1 Gap gene model

In *Drosophila* embryonic development, the expression of gap genes occurs in cleavage cycles 13 and 14A, during which there are three main stages of nuclei behaviour[1] as shown below:

* Interphase. The related gene transcription is activated and chromatin is replicated to prepare mitosis.
* Mitosis. The chromatin is in a condensed state, known as chromosome, and gene transcriptions are inactivated. Only the protein diffusion and degradation occur.
* Division. It is instantaneous in this model, where chromosomes are separated and the nucleus is replaced by two equal daughter nuclei after division. The distance between two nuclei is halved.

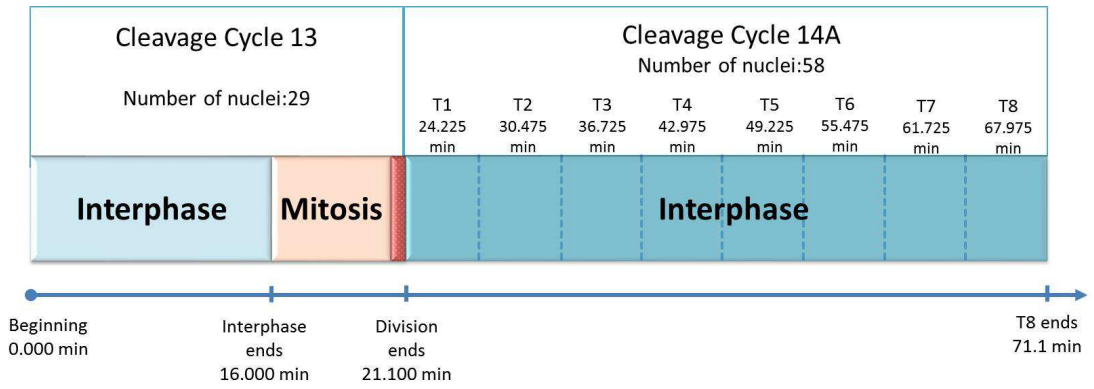


Figure 5.1: Temporal stages for gap gene model (modified from (Jaeger et al., 2004b))

The time schedule for the gap gene model is shown in Figure 5.1. This model covers the developmental period from the beginning of Cycle 13 (time = 0 min), when the gap genes are firstly detected clearly shown in `FlyEx`, to the end of Cycle 14*A* (time = 71.1 min), when the gastrulation starts. Mitosis takes 5.1 min at the end of Cycle 13 followed by transient division. The `FlyEx` quantitative gene expression data, shown in Figure 5.2, have one time class of Cycle 13 at 10.55 min and 8 classes during Cycle 14*A* as shown in the schedule. As noted in previous chapters, Cycle 13 in `FlyEx` is only staged around 10 min which is 10 min less than Jaeger et al. (2004b)'s schedule. Such a mapping variability between developmental stages and real time, found in literature, is expected to be unified in the future.

The mathematical simulated model, which is a widely used gap gene pattern formation model following Jaeger et al. (2004a)'s work, is defined by

$$\begin{aligned}\frac{dv_i^a}{dt} &= R_a\Phi\left(\sum_{b=1}^{N}\mathbf{W}^{ab}v_i^b + m^a v_i^{\text{Bicoid}} + h^a\right) && \texttt{Regulated synthesis}\\ &+ D^a(v_{i-1}^a - v_i^a + v_{i+1}^a - v_i^a) && \texttt{Diffusion}\\ &- \tau_a^{-1}v_i^a && \texttt{Degradation}\end{aligned} \tag{5.1}$$

Six ($N = 6$) transcription factors (*cad, hb, Kr, gt, kni,* and *tll*), indexed by a, are cross-regulated each other gene with the external maternal input – Bicoid. The maternal *hb* and *cad* expression data in Cycle 12 from `FlyEx` are included as the initial inputs. v_i is concentration of each nucleus of gene a and nuclei denoted by $i-1$ and $i+1$ are the neighbours of i. Because the D-V patterning system is mostly independent of the A-P axis in the trunk region of the blastoderm embryo (Jaeger et al., 2004b), all the

[1]More detailed stages such as prophase, prometaphase and metaphase in mitosis are not considered in this gap gene model.

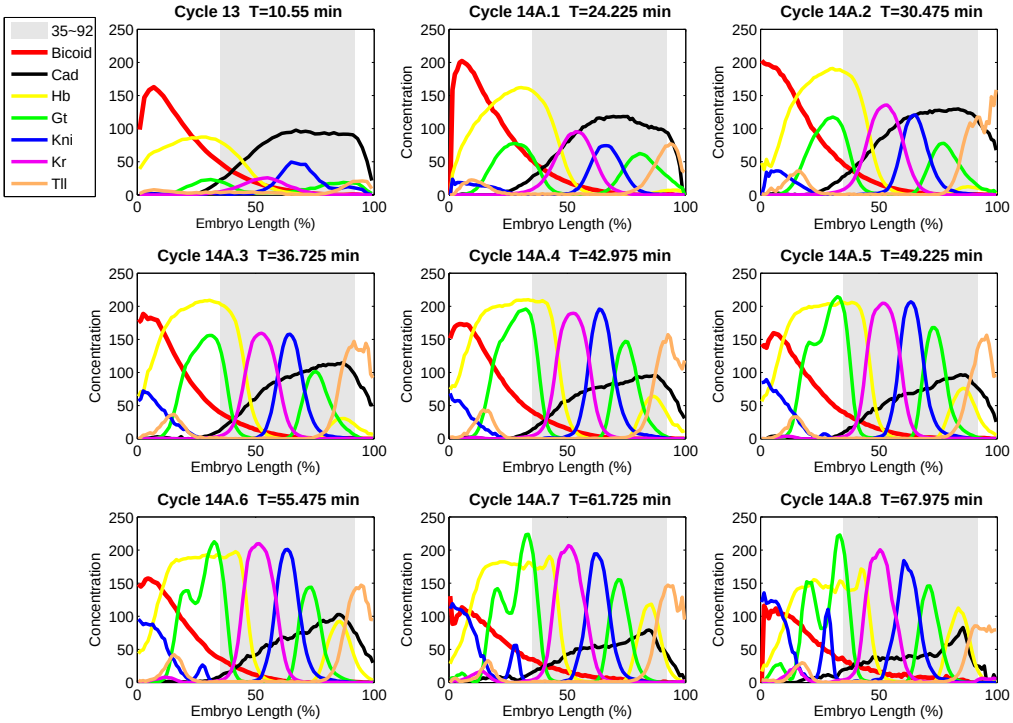


Figure 5.2: Observations from `FlyEx`, where the seven genes express along the A-P axis. The grey area represents partial axis within 35 – 92% EL. Bicoid gradient as an external input varies in the developmental stages.

nuclei in this model are distributed on the one-dimensional A-P axis only. Moreover, the regulatory domain is narrowed between 35% – 92% of embryo length since additional factors are required if the domain is extended to 0% – 100% (Jaeger et al., 2004a). For example, the regulation of head gap genes needs to be considered in the anterior of 35% A-P axis (Cohen and Jürgens, 1990; Grossniklaus et al., 1994) and *hkb*, a terminal gap gene, effects posterior region of 92% A-P axis (Weigel et al., 1990; Brönner and Jäckle, 1991).

The gap gene model in Equation 5.1 comprises three parts:

* Regulated synthesis, describing the interconnections between zygotic genes. $\mathbf{W}^{ab}$ $(N \times N)$ is a genetic regulation matrix, in which each element represents the regulatory effect of gene b on a. This matrix is shown below

$$\begin{bmatrix} w^{cad \leftarrow cad} & w^{cad \leftarrow hb} & w^{cad \leftarrow Kr} & w^{cad \leftarrow gt} & w^{cad \leftarrow kni} & w^{cad \leftarrow tll} \\ w^{hb \leftarrow cad} & w^{hb \leftarrow hb} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

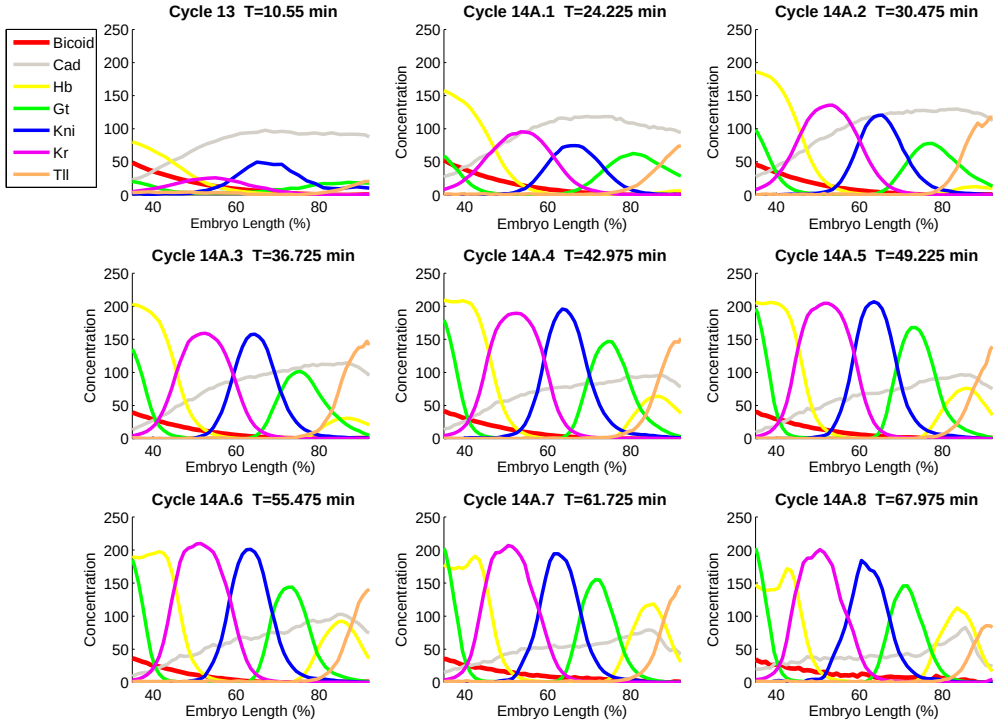


Figure 5.3: Observations from `FlyEx` and all the expression data are only shown within 35% – 92% of the A-P axis.

The regulatory effect for each gene a from input Bicoid concentration is represented by m^a. The total regulated synthesis, a sigmoid function, is then given by the genetic interactions, threshold parameter h^a and a maximum synthesis rate R^a:

$$R_a\Phi(x^a) = 0.5R_a\left(\frac{x^a}{\sqrt{(x^a)^2+1}}+1\right), \tag{5.2}$$

where $x^a = \sum_{b=1}^{N} \mathbf{W}^{ab} v_i^b + m^a v_i^{\text{Bicoid}} + h^a$.

* Diffusion, representing the diffusive exchanging of proteins between neighbouring nuclei. The diffusion coefficient D varies inversely with the squared distance between the nuclei after each division. In this model, the diffusion coefficient D in Cycle 14 is equal to four times the one in Cycle 13.

* The last term is protein degradation, where τ_a^{-1} is decaying rate.

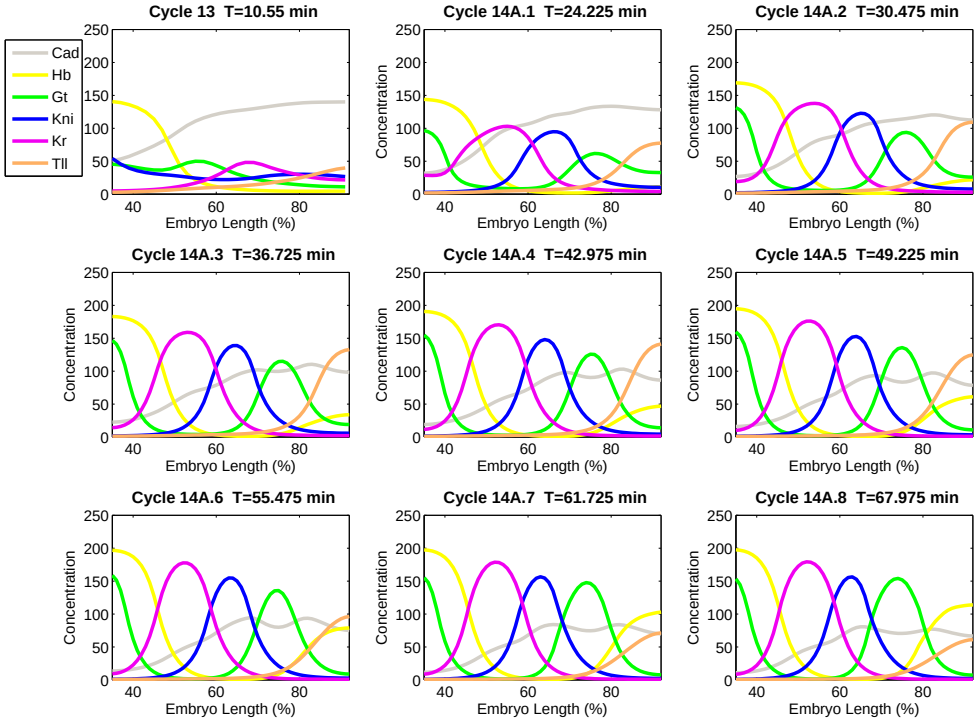


Figure 5.4: The simulated gap gene model with real Bicoid concentration during Cycle 13 and 8 temporal classes in 14*A*. Six regulated genes are shown with different colours.

5.2 Results and discussion

The Bicoid concentration has been considered as constant in previous works (Jaeger et al., 2004b,a; Fomekong-Nanfack et al., 2007). In this thesis, we implement the model with realistic Bicoid expression data from `FlyEx`.

Since the input has been changed, the 66 parameters of this gap gene model need to be re-estimated in order to find reasonable connections. A global optimisation algorithm, Parallel Lam Simulated Annealing (PLSA), has been used by Jaeger and colleagues in all of their works. This algorithm takes quite a long time, 8 – 160 h on 10 processors. Another method, evolution strategy, has been implemented on the same gap gene model (Fomekong-Nanfack et al., 2007, 2009) and the computational cost is much lower than PLSA.

In this thesis, we focus on the effect of the zygotic gene network from input Bicoid concentration. Due to the time limitations, instead of the global optimisation mentioned, we used a local search and started from the optimised parameter values from Jaeger et al. (2004a)'s work when Bicoid has been replaced by real data. The main contribution of Jaeger's well known gap gene model is that the gap genes' anterior shifts can be explained

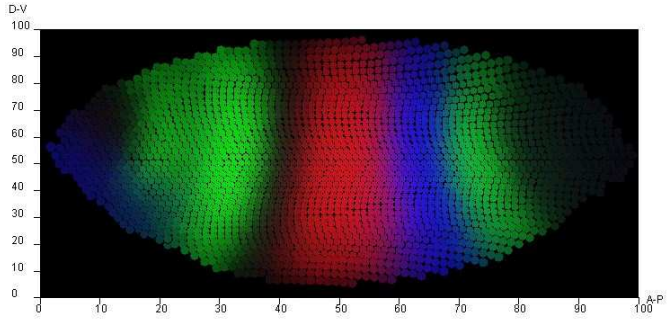


Figure 5.5: Pattern image of three gap genes: *gt* (green), *kni* (blue) and *Kr* (red) from `FlyEx`.

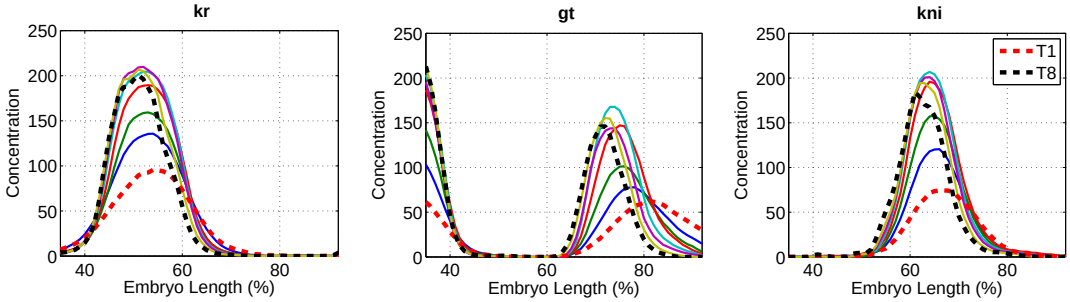


Figure 5.6: Three main gap genes *Kr*, *gt* and *kni* expression data drawn from `FlyEx` in Cycle 14*A* with 8 temporal classes. $T1$ and $T8$ are shown with red and black dashed lines, respectively.

by the asymmetric repression (see Figure 5.7). The same shift also occurs when the input Bicoid is changed to decay (see Figures 5.8 and 5.9).

The gap gene cross-connections are described by the genetic interconnectivity matrix W (see Section 5.1). It is worth asking how does such a regulatory mechanism change with the realistic Bicoid concentration? Differing from Jaeger et al. (2004b)'s network topology, the gene *kni* activates rather than represses *gt* according to our parameter estimation. It implies that the decaying Bicoid not only initialises the gene network, but also has a potential contribution to the cross-regulation in gap gene network. The model-based regulated Bicoid profile from Chapter 3 could also have the same contribution since this model fits the database well. However, in Chapter 3, the anterior fitting has a significant effect because the Bicoid concentration in the anterior domain is much larger than in the posterior part. In the gap gene model, only the main body domain, 35% – 92% EL, is considered. The synthetic Bicoid profile does not undergo distinguished degradation during this domain. In order to obtain the simulated decaying Bicoid in Cycle 14*A* at the middle and posterior parts, an improved weighted fitting procedure can be applied to match the database in future work, in which the higher weight factor is needed between 35% – 92% EL.

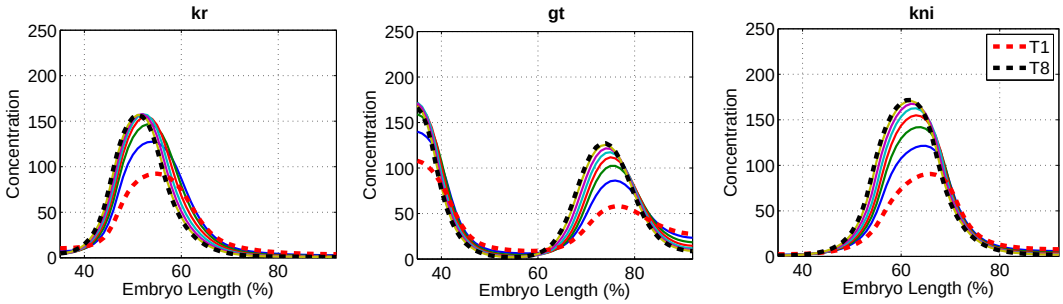


Figure 5.7: The reproduction of Jaeger et al. (2004a)'s work, in which Bicoid is an averaged constant vector. The parameters can be found in their work with numbered 28008.

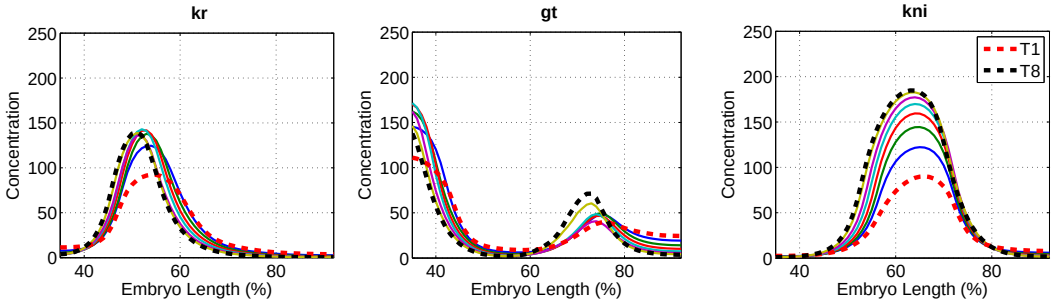


Figure 5.8: Simulated results for the gap gene model with the realistic Bicoid obtained from `FlyEx`, which shows a degradation during Cycle 14*A* (see Chapter 3 for more details) The parameters have remained the same as those set by Jaeger et al. (2004a).

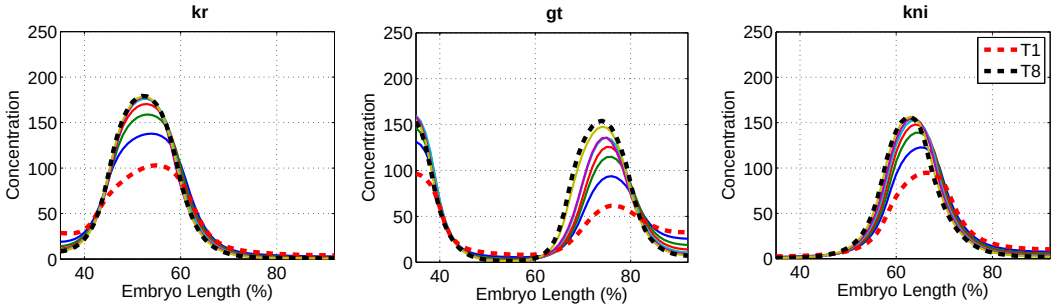


Figure 5.9: Simulated results for the gap gene model with real Bicoid concentration. The 66 parameters are estimated by local search starting from Jaeger et al. (2004a)'s estimation.

Table 5.1: Regulatory weight matrix **W**

	cad	*hb*	*Kr*	*gt*	*kni*	*tll*
cad	0/0/12	0/0/12	0/0/12	0/ 0/12	0/0/12	0/0/12
hb	12/ 0/0	12/0/0	**6/1/5** (8/2/2)	**12/0/0** (11/0/1)	0/0/12	**5/4/3** (8/2/2)
Kr	12/0/0	**0/1/11** (0/0/12)	12/0/0	0/0/12	0/0/12	0/0/12
gt	12/0/0	**2/0/10** (1/0/11)	0/0/12	**12/0/0** (10/1/1)	**10/0/2** (4/0/8)	0/0/12
kni	12/0/0	0/0/12	**2/2/8** (3/0/9)	0/0/12	**12/0/0** (11/1/0)	0/0/12
tll	6/0/6	0/0/12	0/0/12	1/0/11	1/0/11	**9/1/2** (7/1/4)

Parameter estimation by local search starting from Jaeger's 12 parameter sets. Each regulatory interaction is described by the number of interactions in triplet format: activation / no interaction / repression. The interactions in bold are the results from local search which are different from Jaeger's work.

5.3 Summary

The gap gene network establishes around two hours after fertilisation in embryo development. This is the same time duration that *bicoid* mRNA is decayed. Since the traditional gap gene model only considers constant Bicoid, it is interesting to ask what will happen if Bicoid is regulated. We use the Bicoid measurements from `FlyEx` as the external input to this network and did local search to estimate parameters by starting from the values in literature. The network topology has been changed in our result: *kni* activates *gt* rather than represses. More effects of regulated Bicoid need to be exploited in the future and some possible network parameter optimisations will be discussed in Chapter 6.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This dissertation has proposed computational analyses for modelling Bicoid gradient formation during *Drosophila* embryonic development. Its focus has been on the regulation of stability of the maternally provided *bicoid* mRNA. While there is some evidence in the biological literature that the stability may be regulated, computational models over the last three decades have not taken this into account. This work has presented two methods of modelling source regulation, and shows that model parameters can be inferred from a public domain dataset (`FlyEx`) of Bicoid measurements in space and time. The work further addresses a gap gene regulatory network which is activated precisely during the time when *bicoid* mRNA stability is degraded.

The first contribution in this thesis, described in Chapter 3, imposes a stylised regulation function on the production (source) of maternal Bicoid, based on the assumption that these are proportional. By matching the observations from a quantitative database `FlyEx`, our regulated Bicoid deterministic reaction-diffusion model results in better fit by searching the parameters exhaustively. A recent model, in which a cytoplasmic flow is included, has also been implemented with mRNA regulation. In addition to these deterministic models, we also implemented a stochastic simulation of Bicoid propagation and analysed it with the Gillespie algorithm. We made three observations from this study. Firstly the modelling error, i.e. the squared error between model outputs and `FlyEx` measurements turned out to be unimodal. Hence with the small number of unknown parameters in the models, it was possible to do parameter estimation by a grid search. Secondly, the parameters inferred by this process were largely consistent across the three models. In particular, the onset time of decay and its speed were very close. Finally, because the `FlyEx` measurements come from a population of embryos, we were

able to construct bootstrap samples of morphogen data and quantify the uncertainty in our inferred parameter values.

The alternative approach of modelling with non-parametric Bayesian methods, now popular in the machine learning literature, was described in Chapter 4. Here, we applied the GP method to infer source regulation in a spatio-temporal setting. While this approach of using GPs for gene regulation has been addressed by Gao et al. (2008) previously, our work is the first to study a spatio-temporal problem in developmental biology. For the synthetic data from the reaction-diffusion model of Chapter 3, our results showed that GP framework in our spatio-temporal system can accurately infer the source function and capture the protein propagation. On real data from `FlyEx`, the reconstructed source function was indicative of stability regulation, but was temporally smoother than what we expected, partly due to the fact that the dataset is only partially observed.

Chapters 3 and 4 focus on the regulated *bicoid* mRNA stability and the Bicoid concentration gradient formation. Bicoid, as a key maternal transcription factor, has contribution on regulating gap gene expression. The establishment of gap gene network occurs around two hours since fertilisation, which is the same time at which the translation of *bicoid* mRNA is switched off. In Chapter 5, we have analysed the dynamical topology of the gap gene network with the regulated Bicoid rather than the constant input from literature. In this contribution, we have estimated model parameters, which indicate the cross-regulatory network topology, by local search started from Jaeger et al. (2004b)'s global optimisation results. This early result from our work suggests potential change within the network structure.

6.2 Future Work

In this thesis, however, *bicoid* is the only morphogen has been studied systemically. Some other morphogens are still be attractive, such as *Decapentaplegic (Dpp)*, which is believed to make a contribution to the *Drosophila* D-V axis and the wing imaginal disc (Ferguson and Anderson, 1992), and *Sonic hedgehog homolog (Shh)*, which is crucial for the growth of digits on limbs and the brain (Roelink et al., 1995), etc. These morphogen gradients, which control the gene expression similar to the way that *bicoid* does, could be formatted in the future by the modelling tools introduced in our work.

Our ongoing work is to optimise the gap gene network parameters. Due to the large scale of the parameters in the gap gene model and the limitations of the implementation time, we have only finished the local search for the fitting parameters when the input has been changed. In order to investigate the effect of the maternal gene *bicoid* with mRNA regulated in gap gene model, a global search is required. Instead of PLSA, which has been successfully implemented by Jaeger and colleagues with the expensive

computational cost, we will follow Fomekong-Nanfack et al. (2007)'s work to estimate the 66 parameters by evolutionary strategy. The concept of such an efficient global algorithm includes crossover, selection and mutation.

There are many specialised evolution strategies (ESs) and the (μ, λ)-ES will be considered in our future work based on (Runarsson and Yao, 2005). The basic idea is that, during each iteration, a new population is generated only by the best μ of λ parents. We firstly assume $\boldsymbol{\theta}$ is a N-parameter vector ($N = 66$ in the gap gene network), $\boldsymbol{\sigma}$ is the step size and the total population is λ. Initially, there are λ combinations of $\boldsymbol{\theta}$ with the uniformly distributed matrix $\mathbf{U}$. The pseudo-code is shown in Algorithm 5

Algorithm 5 (μ, λ)-evolution strategy

Initialisation: $\boldsymbol{\sigma}'_j = (\boldsymbol{\theta}_{\text{up}} - \boldsymbol{\theta}_{\text{low}})/\sqrt{N}$; $\boldsymbol{\theta}'_j = \boldsymbol{\theta}_{\text{low}} + (\boldsymbol{\theta}_{\text{up}} - \boldsymbol{\theta}_{\text{low}})\mathbf{U}_j(0,1)$.
while termination criteria not satisfied **do**
1. Evaluate cost function $f(\boldsymbol{\theta}'_j)$ and the penalty function $g(\boldsymbol{\theta}'_j)$, $j \in (1, \lambda)$.
2. Rank the λ populations stochastically. The choice of the probability depends on the strength of penalisation.
3. Choose the best μ populations, $\boldsymbol{\sigma}'_k$ and $\boldsymbol{\theta}'_k$, as the new parents for the next generations, $k \in (1, \mu)$.
for $j = 1 : \lambda$ **do**
$k \leftarrow \text{mod}(j-1, \mu) + 1$
Apply Gaussian mutation on the step size:
$\sigma'_{j,i} \leftarrow \sigma'_{k,i} \exp(\tau' \mathcal{N}(0,1) + \tau \mathcal{N}_i(0,1))$, $i \in (1, N)$.
Generate new population:
$\boldsymbol{\theta}'_j \leftarrow \boldsymbol{\theta}_k + \boldsymbol{\sigma}'_j \boldsymbol{\mathcal{N}}(0,1)$, and
$\boldsymbol{\sigma}'_j \leftarrow \boldsymbol{\sigma}_k + \alpha(\boldsymbol{\theta}'_j - \boldsymbol{\theta}_k)$
end for
end while

Another interest in our future work is expression variation of gap gene network. Due to the limited measurements in `FlyEx`, the potential network connections (activation, no-interaction and repression) and gene expression patterns, are hard to be inferred. Furthermore, with the deterministic gap gene model, we can not guarantee that the inferred gap gene network topology is biologically realistic because the regulatory process is highly fluctuated. To address such problems, the variability and robustness of gap gene model should be analysed based on stochastic model rather than deterministic one. In particular, we can assume the real data are generated by the model output with additive Gaussian noise, which leads to a probabilistic gap gene network.

Appendix A

Appendix

A.1 Analytical solution for Bicoid reaction-diffusion model with constant source

Constant production rate

The usual assumption in solving the Bicoid reaction-diffusion model is that the protein synthesis process is continuous, which is shown in Figure 3.3(a):

$$S_{con} = S_0 \delta(x) \Theta(t), \tag{A.1}$$

where S_0 is the production rate, $\delta(x)$ is the Kronecker delta function and $\Theta(t)$ is Heaviside step function.

With the constant protein synthesis, the Bicoid morphogen reaction-diffusion system is given by:

$$\frac{\partial}{\partial t} m(x,t) = D \frac{\partial^2}{\partial x^2} m(x,t) - \tau_p^{-1} m(x,t) + S_{con}(x,t). \tag{A.2}$$

(Bergmann et al., 2007) provides the analytical solution by two-dimensional Fourier Transform:

$$\widetilde{m}(k,\omega) = \iint_{-\infty}^{\infty} m(x,t) e^{-i(kx+\omega t)} dx dt. \tag{A.3}$$

Therefore, Equation A.2 is transformed as follow:

$$i\omega \widetilde{m}(k,\omega) = -(k^2 D + \tau_p^{-1}) \widetilde{m}(k,\omega) + \widetilde{S}_{con}(k,\omega), \tag{A.4}$$

where $\widetilde{S}_{con}(k,\omega)$ is given by:

$$\widetilde{S}_{con}(k,\omega) = \iint_{-\infty}^{\infty} S(x,t)e^{-i(kx+\omega t)}dxdt = -\frac{iS_0}{\omega}. \tag{A.5}$$

The morphogen concentration in (k,ω) domain is shown as below:

$$\widetilde{m}(k,\omega) = \frac{S_0}{k^2D+\tau_p^{-1}}\left(\frac{1}{i\omega} - \frac{1}{i\omega+k^2D+\tau_p^{-1}}\right). \tag{A.6}$$

By one-dimensional inverse Fourier Transform (IFT), $\widetilde{m}(k,t)$ becomes as:

$$\widetilde{m}(k,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\widetilde{m}(k,\omega)e^{j\omega t}d\omega = \frac{S_0}{k^2D+\tau_p^{-1}}\left(1-e^{-(k^2D+\tau_p^{-1})}\right)\Theta(t). \tag{A.7}$$

Finally, the solution to Bicoid spatio-temporal model with constant source is given by:

$$m(x,t) = \frac{S_0}{2\beta D}\left(\exp(-\beta x) - \frac{\exp(-\beta x)}{2}\mathrm{erfc}\left(\frac{2\beta Dt - x}{\sqrt{4Dt}}\right) - \frac{\exp(\beta x)}{2}\mathrm{erfc}\left(\frac{2\beta Dt + x}{\sqrt{4Dt}}\right)\right), \tag{A.8}$$

where β is:

$$\beta = 1/\sqrt{D\tau_p^{-1}}. \tag{A.9}$$

Decaying production rate

Considering decaying source with exponential function:

$$S_{dec} = S_0\delta(x)\Theta(t)\exp^{-\frac{t}{\tau_m}} \tag{A.10}$$

where $1/\tau_m$ is decaying rate of *bicoid* mRNA.

$\widetilde{S}_{dec}(k,\omega)$ is transformed as:

$$\begin{aligned}\widetilde{S}_{dec}(k,\omega) &= \iint_{\infty}^{\infty} S(x,t)e^{-i(kx+\omega t)}dxdt \\ &= S_0\int_{\infty}^{\infty}\delta(x)e^{ikx}dx\int_{\infty}^{\infty}e^{-\frac{t}{\tau_m}}\Theta(t)e^{-i\omega t}dt \\ &= \frac{S_0}{i\omega+\tau_m^{-1}},\end{aligned} \tag{A.11}$$

and

$$\begin{aligned}\widetilde{m}(k,\omega) &= \frac{S_0}{\tau_m^{-1}+i\omega}\cdot\frac{1}{Dk^2+\tau_p^{-1}+i\omega} \\ &= \frac{S_0}{Dk^2+\tau_p^{-1}-\tau_m^{-1}}\cdot\left(\frac{1}{\tau_m^{-1}+i\omega} - \frac{1}{Dk^2+\tau_p^{-1}+i\omega}\right).\end{aligned} \tag{A.12}$$

With one-dimensional IFT, we have:

$$\begin{aligned}\widetilde{m}(k,t) &= \frac{S_0}{Dk^2+\tau_p^{-1}-\tau_m^{-1}} \cdot \left(e^{-\tau_m^{-1}t} - e^{-(Dk^2+\tau_p^{-1})t}\right)\Theta(t) \\ &= \frac{S_0 \cdot e^{-\tau_m^{-1}t}}{Dk^2+\tau_p^{-1}-\tau_m^{-1}} \cdot \left(1 - e^{-(Dk^2+\tau_p^{-1}-\tau_m^{-1})t}\right)\Theta(t) \end{aligned} \tag{A.13}$$

Protein decaying faster than source decaying ($\tau_p^{-1} - \tau_m^{-1} > 0$)

Equation A.13 is similar to the solution with constant source when $\tau_p^{-1} - \tau_m^{-1} > 0$, in which the protein degrades faster than source. The morphogen spatio-temporal model with decaying source is given by:

$$\begin{aligned} m(x,t) =& \frac{S_0 \cdot \exp\left(-\frac{t}{\tau_m}\right)}{2\beta_1 D} \\ & \left(\exp\left(-\beta_1 x\right) - \frac{\exp\left(-\beta_1 x\right)}{2}\mathrm{erfc}\left(\frac{2\beta_1 Dt - x}{\sqrt{4Dt}}\right) - \frac{\exp\left(\beta_1 x\right)}{2}\mathrm{erfc}\left(\frac{2\beta_1 Dt + x}{\sqrt{4Dt}}\right)\right) \end{aligned} \tag{A.14}$$

where β_1 is:

$$\beta_1 = \sqrt{\frac{\tau_p^{-1} - \tau_m^{-1}}{D}} \tag{A.15}$$

Source decaying faster than protein decaying $\tau_p^{-1} - \tau_m^{-1} < 0$

When $\tau_p^{-1} - \tau_m^{-1} < 0$, the Bicoid concentration.

$$\beta_2 = i\beta_1 \tag{A.16}$$

$$\begin{aligned} m(x,t) =& \frac{S_0 \cdot e^{-\frac{t}{\tau_m}}}{2i\beta_1 D} \\ & \left(\exp\left(-i\beta_1 x\right) - \frac{\exp\left(-i\beta_1 x\right)}{2}\mathrm{erfc}\left(\frac{2i\beta_1 Dt - x}{\sqrt{4Dt}}\right) - \frac{\exp\left(i\beta_1 x\right)}{2}\mathrm{erfc}\left(\frac{2i\beta_1 Dt + x}{\sqrt{4Dt}}\right)\right) \end{aligned} \tag{A.17}$$

A.2 Bicoid steady state

When the concentration remains unchanging in time, as an equilibrium maintained between the production and degradation of morphogen molecules, this reaction-diffusion system reaches steady state, in which $\partial m(x,t)/\partial t = 0$, and is defined as:

$$D\frac{\partial^2}{\partial x^2}m_s(x) = \tau_p^{-1}m_s(x). \tag{A.18}$$

As a second-order homogeneous linear equation, Equation A.18 has a general solution as follow:

$$m_s(x) = C_1 m_1(x) + C_2 m_2(x), \tag{A.19}$$

where C_1, C_2 are the constants to be determined and the two solutions $m_1(x)$ and $m_2(x)$ are linearly independent. The corresponding auxiliary equation is given by:

$$Dr^2 - \tau_p^{-1} = 0,$$

and the roots are:

$$r_{1,2} = \pm\sqrt{\frac{1}{D\tau_p}}. \tag{A.20}$$

Because r_1 and r_2 are real and distinct, the general solution is given by:

$$m_s(x) = C_1 \exp(r_1 x) + C_2 \exp(r_2 x). \tag{A.21}$$

Let λ ($\lambda = \sqrt{D\tau_p}$) be the decay length scale of the concentration gradient. We get:

$$m_s(x) = C_1 \exp\left(-\frac{x}{\lambda}\right) + C_2 \exp\left(\frac{x}{\lambda}\right). \tag{A.22}$$

The constants C_1 and C_2 are determined by two boundary conditions:

$$D\frac{\partial m}{\partial x}\bigg|_{x=0} = -J, \quad \text{and} \tag{A.23}$$

$$D\frac{\partial m}{\partial x}\bigg|_{x=L} = 0, \tag{A.24}$$

where J is the diffusive flux at anterior pole ($x = 0$) shown in Equation 3.3. The flux at the other end of embryo is defined as zero ($x = L$ by assuming that the length of embryo is L).

Substituting Equation A.22 into boundary conditions, we have:

$$C_1 = \frac{J\lambda}{D}\left(\frac{1}{1-\exp\left(\frac{-2L}{\lambda}\right)}\right), \tag{A.25}$$

and

$$C_2 = \frac{J\lambda}{D}\left(\frac{\exp\left(\frac{-2L}{\lambda}\right)}{1-\exp\left(\frac{-2L}{\lambda}\right)}\right). \tag{A.26}$$

If the embryo length L is much larger than length scale λ, C_2 becomes zero and C_1 is simplified as $\frac{J\lambda}{D}$. The finial solution for the steady state is defined by:

$$m_s(x) = C_1 \exp\left(-\frac{x}{\lambda}\right), \tag{A.27}$$

where $C_1 = \frac{J\lambda}{D}$, as a boundary condition, is morphogen concentration at source end ($x = 0$). λ, the decaying length, is given by $\sqrt{D\tau_p}$, which defines the distance between the source and the spatial point, at which the concentration is $\exp(-1)$ of C_1.

A.3 Gillespie stochastic simulation algorithm

This appendix shows the derivation details about how to obtain time step τ that the next reaction takes place in stochastic simulation (Gillespie, 1977; Erban et al., 2007). The basic idea of finding τ is dependent on the probability of the reaction occurring in the future. Suppose $P_{no}(A(t), \tau)$ is the probability that no reaction occurs during $[t, t+\tau]$, given $A(t)$ molecules at time t in this system. Let $P_{re}(A(t), \tau)d\tau$ be the probability that the next reaction takes place during $[t+\tau, t+\tau+d\tau)$, where $d\tau$ is a small time interval. Therefore, we can get:

$$P_{re}(A(t), \tau)d\tau = P_{no}(A(t), \tau)A(t+\tau)kd\tau. \tag{A.28}$$

Because there is no reaction during $[t, t+\tau)$, we have $A(t+\tau) = A(t)$ and Equation A.28 can be changed as follow:

$$P_{re}(A(t), \tau)d\tau = P_{no}(A(t), \tau)A(t)kd\tau. \tag{A.29}$$

Now we need to calculate $P_{no}(A(t), \tau)$. The probability that no reaction occurs in $[t, t+\tau+d\tau)$ is defined as no reaction happens in $[t, t+\tau)$ and $[t+\tau, t+\tau+d\tau)$:

$$\begin{aligned} P_{no}(A(t), \tau + d\tau) &= P_{no}(A(t), \tau)[1 - A(t+\tau)kd\tau] \\ &= P_{no}(A(t), \tau)[1 - A(t)kd\tau] \end{aligned} \tag{A.30}$$

After algebraic manipulation and let $d\tau \to 0$, we can obtain that:

$$\frac{dP_{no}(A(t), \tau)}{d\tau} = -A(t)kP_{no}(A(t), \tau) \tag{A.31}$$

Solving this ODE with initial condition $P_{no}(A(t), 0) = 1$, we get:

$$P_{no}(A(t), \tau) = \exp(-A(t)k\tau) \tag{A.32}$$

Combining Equation A.29 and A.32, we have:

$$P_{re}(A(t), \tau)d\tau = A(t)k\exp(-A(t)k\tau)d\tau. \tag{A.33}$$

Defining a function R of τ, in which

$$R(\tau) = \exp(-A(t)k\tau). \tag{A.34}$$

For an arbitrary $\tau \in (0, \infty)$, $R(\tau)$ a random number in $(0, 1)$. Given two random numbers $a \in (0, 1)$ and $b \in (0, 1)$, where $a < b$, the probability that $R(\tau) \in (a, b)$ is equal to $\tau \in (R^{-1}(b), R^{-1}(a))$ because R is a monotone decreasing function of τ. The

probability of $\tau \in (R^{-1}(b), R^{-1}(a))$ is given by:

$$\begin{aligned}\int_{R^{-1}(b)}^{R^{-1}(a)} P_{re}(A(t),\tau)d\tau &= \int_{R^{-1}(b)}^{R^{-1}(a)} A(t)k\exp(-A(t)k\tau)d\tau \\ &= -R(\tau)\big|_{R^{-1}(b)}^{R^{-1}(a)} \\ &= b-a. \end{aligned} \tag{A.35}$$

Therefore, the probability that $R(\tau) \in (a,b)$ is $b-a$ which leads to that $R(\tau)$ is uniformly distributed in $(0,1)$. For a given τ, we define a uniformly distributed number r, where

$$\begin{aligned} r &= \exp(-A(t)k\tau), \quad \text{and} \\ \tau &= \frac{1}{A(t)k}\ln(\frac{1}{r}). \end{aligned} \tag{A.36}$$

This derivation can also be expanded to multi-reactions system. Suppose there are N reactions a chemical system. The time step τ is similar as Eq.A.36:

$$\tau = \frac{1}{a_0}\ln(\frac{1}{r}) \tag{A.37}$$

where a_0 is defined by all of the propensity functions in the system:

$$a_0 = \sum_{j=1}^{N} a_j \tag{A.38}$$

A.4 Derivation of stochastic mean

The master equation of Bicoid reaction-diffusion system is given by:

$$\begin{aligned}\frac{\partial}{\partial t}P(\mathbf{n},t) =& \underbrace{d\sum_{i=1}^{N-1}[(n_i+1)P(R_i^{\pm1}\mathbf{n},t)-n_iP(\mathbf{n},t)]}_{\texttt{Diffusion:}\,A\rightarrow P} + \underbrace{d\sum_{i=2}^{N}[(n_i+1)P(L_i^{\pm1}\mathbf{n},t)-n_iP(\mathbf{n},t)]}_{\texttt{Diffusion:}\,A\leftarrow P}\\ &+\tau_p^{-1}\sum_{i=1}^{N}[(n_i+1)P(K_i^{+1}\mathbf{n},t)-n_iP(\mathbf{n},t)]\\ &+S(t)P[(K_1^{-1}\mathbf{n},t)-P(\mathbf{n},t)],\end{aligned} \tag{A.39}$$

where $P(\mathbf{n},t)$ is joint probability of state vector $\mathbf{n}=[n_1,n_2,\ldots,n_i,\ldots,n_N]$ and $N=100$. $R_i^{\pm1}$, $L_i^{\pm1}$, K_i^{+1} and K_i^{-1} are state operators, which are defined by:

$$R_i^{\pm1}\mathbf{n} = [n_1,n_2,\ldots,n_i+1,n_{i+1}-1\ldots,n_N], \quad i=1,2,\ldots,N-1 \tag{A.40}$$

$$L_i^{\pm1}\mathbf{n} = [n_1,n_2,\ldots,n_{i-1}-1,n_i+1\ldots,n_N], \quad i=2,3,\ldots,N \tag{A.41}$$

$$K_i^{+1}\mathbf{n} = [n_1,n_2,\ldots,n_i+1,\ldots,n_N], \quad i=1,2,\ldots,N \tag{A.42}$$

$$K_i^{-1}\mathbf{n} = [n_1,n_2,\ldots,n_i-1,\ldots,n_N], \quad i=1,2,\ldots,N \tag{A.43}$$

In order to estimate parameters used in stochastic model, we have calculated vector $\overline{\mathbf{m}}$ by multiplying n_i and summing over the vector $\mathbf{n}$, $(i=1,2,\ldots,N)$ rather than averaging several Gillespie realizations of stochastic simulation.

$$\overline{\mathbf{m}} = [\overline{m}_1,\overline{m}_2,\ldots,\overline{m}_i,\ldots,\overline{m}_N], \quad i=1,2,\ldots,N \tag{A.44}$$

$$\overline{m}_i(t) = \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} n_iP(\mathbf{n},t) = \sum_{\mathbf{n}} n_iP(\mathbf{n},t), \quad i=1,2,\ldots,N \tag{A.45}$$

Where $\overline{m}_i(t)$ gives the mean number of Bicoid molecules at time t in i-th sub-volume. By multiplying Equation A.39 by n_j and sum over $\mathbf{n}$, the diffusion term (first line in A.39) becomes:

$$\begin{aligned}\frac{\partial}{\partial t}\sum_{\mathbf{n}} n_iP(\mathbf{n},t)\Big|_{A\rightarrow P} =& d\sum_{j=1}^{N-1}\underbrace{\left[\sum_{\mathbf{n}} n_i(n_j+1)P(R_j^{\pm1}\mathbf{n},t)-\sum_{\mathbf{n}} n_in_jP(\mathbf{n},t)\right]}_{\text{ⓐ}}\\ &+d\sum_{i=2}^{N}\underbrace{\left[\sum_{\mathbf{n}} n_i(n_j+1)P(L_j^{\pm1}\mathbf{n},t)-\sum_{\mathbf{n}} n_in_jP(\mathbf{n},t)\right]}_{\text{ⓑ}}\end{aligned} \tag{A.46}$$

We first consider i goes through A-P axis except anterior and posterior ends, in which $i = 2, ..., N-1$.

1. j=i

The right hand side term ⓐ in Equation A.46 with $j = i$ becomes:

$$ⓐ|_{j=i} = \sum_{\mathbf{n}} n_i(n_i+1)P(R_i^{\pm 1}\mathbf{n}, t) - \sum_{\mathbf{n}} n_i^2 P(\mathbf{n}, t). \tag{A.47}$$

In order to change $P(R_i^{\pm 1}\mathbf{n})$ to $P(\mathbf{n}, t)$, we define the transformation as follow:

$$R_i^{\pm 1}\mathbf{n} = \{n_1, n_2, \ldots, n_i + 1, n_{i+1} - 1 \ldots, n_N\}, \quad \text{to} \tag{A.48}$$

$$\mathbf{n} = \{n_1, n_2, \ldots, n_i, n_{i+1} \ldots, n_N\}, \quad i = 1, 2, \ldots, N-1 \tag{A.49}$$

Equation A.47 becomes:

$$\begin{aligned} ⓐ|_{j=i} &= \sum_{\mathbf{n}} (n_i - 1)n_i P(\mathbf{n}, t) - \sum_{\mathbf{n}} n_i^2 P(\mathbf{n}, t) \\ &= -\sum_{\mathbf{n}} n_i P(\mathbf{n}, t) \\ &= -\overline{m}_i(t) \end{aligned} \tag{A.50}$$

For diffusion term ⓑ with $j = i$, we have:

$$ⓑ = \sum_{\mathbf{n}} n_i(n_j+1)P(L_j^{\pm 1}\mathbf{n}, t) - \sum_{\mathbf{n}} n_i n_j P(\mathbf{n}, t) \tag{A.51}$$

where the transformation of $L_i^{\pm 1}$ is given by:

$$L_i^{\pm 1}\mathbf{n} = \{n_1, n_2, \ldots, n_{i-1} - 1, n_i + 1 \ldots, n_N\}, \quad \text{to} \tag{A.52}$$

$$\mathbf{n} = \{n_1, n_2, \ldots, n_i, n_{i+1} \ldots, n_N\}, \quad i = 2, 3, \ldots, N \tag{A.53}$$

and ⓑ becomes:

$$\begin{aligned} ⓑ|_{j=i} &= \sum_{\mathbf{n}} (n_i - 1)n_i P(\mathbf{n}, t) - \sum_{\mathbf{n}} n_i^2 P(\mathbf{n}, t) \\ &= -\sum_{\mathbf{n}} n_i P(\mathbf{n}, t) \\ &= -\overline{m}_i(t). \end{aligned} \tag{A.54}$$

2. j=i-1

When $j = i - 1$, we get

$$ⓐ|_{j=i-1} = \sum_{\mathbf{n}} n_i(n_{i-1}+1)P(R_{i-1}^{\pm 1}\mathbf{n}, t) - \sum_{\mathbf{n}} n_i n_{i-1} P(\mathbf{n}, t), \tag{A.55}$$

where

$$R_{i-1}^{\pm 1}\mathbf{n} = \{n_1, n_2, \ldots, n_{i-1}+1, n_i - 1 \ldots, n_N\}, \quad \text{to} \tag{A.56}$$

$$\mathbf{n} = \{n_1, n_2, \ldots, n_{i-1}, n_i \ldots, n_N\}, \quad i = 1, 2, \ldots, N-1 \tag{A.57}$$

and Equation A.55 becomes:

$$\begin{aligned} ⓐ|_{j=i-1} &= \sum_{\mathbf{n}} (n_i+1)n_{i-1}P(\mathbf{n}, t) - \sum_{\mathbf{n}} n_i n_{i-1} P(\mathbf{n}, t) \\ &= \sum_{\mathbf{n}} n_{i-1} P(\mathbf{n}, t) \\ &= \overline{m}_{i-1}(t). \end{aligned} \tag{A.58}$$

When $j = i - 1$, the ⓑ will become zero because:

$$\begin{aligned} ⓑ|_{j=i-1} &= \sum_{\mathbf{n}} n_i(n_{i-1}+1)P(L_{i-1}^{\pm 1}\mathbf{n}, t) - \sum_{\mathbf{n}} n_i n_{i-1} P(\mathbf{n}, t), \\ &= \sum_{\mathbf{n}} n_i n_{i-1} P(\mathbf{n}, t) - \sum_{\mathbf{n}} n_i n_{i-1} P(\mathbf{n}, t) \\ &= 0 \end{aligned} \tag{A.59}$$

3. j=i+1

When $j = i + 1$, ⓐ is zero as similar as Equation A.59. ⓑ is given by:

$$\begin{aligned} ⓑ|_{j=i+1} &= \sum_{\mathbf{n}} n_i(n_{i+1}+1)P(L_{i+1}^{\pm 1}\mathbf{n}, t) - \sum_{\mathbf{n}} n_i n_{i+1} P(\mathbf{n}, t), \\ &= \sum_{\mathbf{n}} (n_i+1)n_{i+1} P(\mathbf{n}, t) - \sum_{\mathbf{n}} n_i n_{i+1} P(\mathbf{n}, t) \\ &= \overline{m}_{i+1}(t) \end{aligned} \tag{A.60}$$

The corresponding ODEs of the master equation for degradation and production (second and third lines in Equation A.39) can be derived by the same way, which is also true when we consider the two ends of the compartments ($i = 1$ and $i = N$). Finally, we get

the ODEs, which is used in Chapter 3 Section 3.3 for parameter estimation, as follow:

$$\frac{\partial}{\partial t}\overline{m}_1 = d(\overline{m}_2 - \overline{m}_1) - \tau_p^{-1}\overline{m}_1 + S(t), \qquad i = 1 \tag{A.61}$$

$$\frac{\partial}{\partial t}\overline{m}_i = d(\overline{m}_{i+1} + \overline{m}_{i-1} - 2\overline{m}_i) - \tau_p^{-1}\overline{m}_i, \quad i = 2, 3, \ldots, N-1 \tag{A.62}$$

$$\frac{\partial}{\partial t}\overline{m}_N = d(\overline{m}_{N-1} - \overline{m}_N) - \tau_p^{-1}\overline{m}_N, \qquad i = N \tag{A.63}$$

A.5 Schur complement

The idea of Schur complement has been introduced since 1851. After 150 years, it still plays an important role in matrix analysis, statistics and many areas of mathematics. In this appendix, we provide the details of Schur complement closely following Gallier (2011)'s work in order to derive multivariate Gaussian distribution in Section 4.1, Chapter 4.

Let a $k \times k$ matrix be partitioned as a 2×2 block matrix:

$$\begin{bmatrix} \underbrace{A}_{m\times m} & \underbrace{B}_{m\times n} \\ \underbrace{C}_{n\times m} & \underbrace{D}_{n\times n} \end{bmatrix}. \tag{A.64}$$

where A, B, C and D are: $m \times m$, $m \times n$, $n \times m$ and $n \times n$ matrices, respectively. Notice that $k = m + n$.

Consider a linear system, where

$$Ax + By = c, \tag{A.65}$$

$$Cx + Dy = d. \tag{A.66}$$

Solving the system for y, we have

$$y = D^{-1}(d - Cx), \tag{A.67}$$

where D is invertible. Substituting it into Equation A.65, we get

$$Ax + B(D^{-1}(d - Cx)) = c, \tag{A.68}$$

and

$$(A - BD^{-1}C)x = c - BD^{-1}d. \tag{A.69}$$

By assuming that matrix $A - BD^{-1}C$ is invertible, the solution becomes:

$$x = (A - BD^{-1}C)^{-1}(c - BD^{-1}d), \tag{A.70}$$

and

$$y = D^{-1}(d - C(A - BD^{-1}C)^{-1}(c - BD^{-1}d)). \tag{A.71}$$

Rewriting the above equations, we have

$$x = (A - BD^{-1}C)^{-1}c - (A - BD^{-1}C)^{-1}BD^{-1}d, \tag{A.72}$$

$$y = -D^{-1}C(A - BD^{-1}C)^{-1}c + (D^{-1} + D{-}1C(A - BD^{-1}C)^{-1}BD^{-1})d. \tag{A.73}$$

Finally, the solution for the matrix $[AB, CD]^{-1}$ of the linear system is given by:

$$\begin{aligned}\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D^{-1} + D{-}1C(A - BD^{-1}C)^{-1}BD^{-1})d \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}. \end{aligned} \tag{A.74}$$

A.6 Derivations of Gaussian process covariance functions

Here we derive the analytical solutions for the covariance function $\mathbf{k}_{\mathbf{m},f}(t,t')$ in Equation 4.61 and $\mathbf{K}_{\mathbf{m},\mathbf{m}}(t,t')$ in Equation 4.60 referred to in the main thesis. In our linear dynamical system, the matrix $\mathbf{A}$, defined in thesis, is a non-diagonal matrix due to the protein diffusion term. We need eigendecomposition of the system matrix to arrive at this solution which is expressed in forms of the eigenvectors (Polderman and Willems, 1998).

Derivation of covariance $\mathbf{k}_{\mathbf{m},f}(t,t')$

$$\mathbf{k}_{\mathbf{m},f}(t,t') = \int_0^t \exp((t-u)\mathbf{A})\mathbf{s}k_{f,f}(u,t')du, \tag{A.75}$$

where $k_{f,f}$ is the squared exponential covariance function given by:

$$k_{f,f} = \exp\left(-\frac{(t-t')^2}{l^2}\right). \tag{A.76}$$

We rewrite Equation (A.75) as

$$\mathbf{k}_{\mathbf{m},f}(t,t') = \int_0^t \exp(\mathbf{A}t - \mathbf{A}u)\mathbf{s}\exp(-\frac{(t'-u)^2}{l^2})du. \tag{A.77}$$

Replacing -$\mathbf{A}$ by its eigendecomposition,

$$\exp(-\mathbf{A}) = \exp(\mathbf{V}\mathbf{P}\mathbf{V}^{-1}) = \mathbf{V}\exp(\mathbf{P})\mathbf{V}^{-1}, \tag{A.78}$$

where matrix $\mathbf{P}$ is a diagonal matrix with the eigenvalues of $-\mathbf{A}$ on the diagonal and $\mathbf{V}$ is the matrix of eigenvectors. Substitute Equation (A.78) into Equation (A.77) and collecting terms in a convenient way:

$$\begin{aligned}
\mathbf{k}_{\mathbf{m},f}(t,t') =& \mathbf{V}\exp(-\mathbf{P}t)\int_0^t \exp\left(\mathbf{P}u - \left(\frac{t'}{l}\right)^2\mathbf{I} + \frac{2t'u}{l^2}\mathbf{I} - \left(\frac{u}{l}\right)^2\mathbf{I}\right)du\mathbf{V}^{-1}\mathbf{s} \\
=& \mathbf{V}\exp(-\mathbf{P}t)\int_0^t \exp\left[-\left(\left(\frac{u}{l}\right)^2\mathbf{I} - \frac{2t'u}{l^2}\mathbf{I} + \left(\frac{t'}{l}\mathbf{I} + \frac{\mathbf{P}l}{2}\right)^2\right) + \left(\frac{t'}{l}\mathbf{I} + \frac{\mathbf{P}l}{2}\right)^2 - \left(\frac{t'}{l}\right)^2\right] \\
& du\mathbf{V}^{-1}\mathbf{s} \\
=& \mathbf{V}\exp\left(-\mathbf{P}t + \mathbf{P}t' + \left(\frac{\mathbf{P}l}{2}\right)^2\right)\int_0^t \exp\left(-\left(\frac{u}{l}\mathbf{I} - \frac{t'}{l}\mathbf{I} - \frac{\mathbf{P}l}{2}\right)^2\right)du\mathbf{V}^{-1}\mathbf{s}.
\end{aligned} \tag{A.79}$$

Here we recognize the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$, and can write the solution as:

$$\mathbf{k}_{\mathbf{m},f}(t,t') = \frac{l\sqrt{\pi}}{2}\mathbf{V} \exp\left(-\mathbf{P}(t-t') + \left(\frac{\mathbf{P}l}{2}\right)^2\right)\left(\text{erf}\left(\frac{t'}{l}\mathbf{I} + \frac{\mathbf{P}l}{2}\right) + \text{erf}\left(\frac{t-t'}{l}\mathbf{I} - \frac{\mathbf{P}l}{2}\right)\right)\mathbf{V}^{-1}\mathbf{s}. \tag{A.80}$$

Derivation of covariance $\mathbf{K}_{\mathbf{m},\mathbf{m}}(t,t')$

$$\mathbf{k}_{\mathbf{m},\mathbf{m}}(t,t') = \int_0^t \int_0^{t'} \exp((t-u)\mathbf{A})\mathbf{s}\left(\exp((t'-u')\mathbf{A})\mathbf{s}\right)^T k_{f,f}(u,u')dudu'. \tag{A.81}$$

With the same transformation as in Equation (A.78), we get:

$$\mathbf{k}_{\mathbf{m},\mathbf{m}}(t,t') = \mathbf{V}\int_0^t \int_0^{t'} \exp((u-t)\mathbf{P})\mathbf{V}^{-1}\mathbf{s}\mathbf{s}^T(\mathbf{V}^{-1})^T\left(\exp((u'-t')\mathbf{P})\right)^T k_{f,f}(u,u')dudu'\mathbf{V}^T. \tag{A.82}$$

Denoting $\mathbf{E} = \mathbf{V}^{-1}\mathbf{s}\mathbf{s}^T(\mathbf{V}^{-1})^T$ and rearranging terms,

$$\begin{aligned}\mathbf{k}_{\mathbf{m},\mathbf{m}}(t,t') =& \mathbf{V}\int_0^t \exp((u-t)\mathbf{P})\mathbf{E}\int_0^{t'}\left(\exp((u'-t')\mathbf{P})\right)^T \exp\left(-\frac{(u-u')^2}{l^2}\mathbf{I}\right)du'du\mathbf{V}^T \\ =& \frac{l\sqrt{\pi}}{2}\mathbf{V}\int_0^t \exp((u-t)\mathbf{P})\mathbf{E} \\ & \left\{\left[\text{erf}\left(\frac{u}{l}\mathbf{I}+\frac{\mathbf{P}l}{2}\right) + \text{erf}\left(\frac{t'-u}{l}\mathbf{I} - \frac{\mathbf{P}l}{2}\right)\right]\exp\left[(u-t')\mathbf{P} + \left(\frac{\mathbf{P}l}{2}\right)^2\right]\right\}du\mathbf{V}^T.\end{aligned} \tag{A.83}$$

In order make the derivation clear, we define two diagonal matrices $\mathbf{X}$, $\mathbf{Y}$ as:

$$\mathbf{X} = \exp((u-t)\mathbf{P}) \tag{A.84}$$

and

$$\mathbf{Y} = \left[\text{erf}\left(\frac{u}{l}\mathbf{I}+\frac{\mathbf{P}l}{2}\right) + \text{erf}\left(\frac{t'-u}{l}\mathbf{I} - \frac{\mathbf{P}l}{2}\right)\right]\exp\left[(u-t')\mathbf{P} + \left(\frac{\mathbf{P}l}{2}\right)^2\right]. \tag{A.85}$$

Equation (A.83) then becomes:

$$\mathbf{k}_{\mathbf{m},\mathbf{m}}(t,t') = \frac{l\sqrt{\pi}}{2}\mathbf{V}\int_0^t \mathbf{X}\mathbf{E}\mathbf{Y}du\mathbf{V}^T. \tag{A.86}$$

We proceed to integrate Equation (A.86) as follows. First rewriting elementwise

$$\mathbf{k_{m,m}}(t,t') = \frac{l\sqrt{\pi}}{2}\mathbf{V}\int_0^t \begin{bmatrix} \mathbf{E}(1,1)\mathbf{X}(1,1)\mathbf{Y}(1,1) & \mathbf{E}(1,2)\mathbf{X}(1,1)\mathbf{Y}(2,2) & \cdots & \mathbf{E}(1,N)\mathbf{X}(1,1)\mathbf{Y}(N,N) \\ \mathbf{E}(2,1)\mathbf{X}(2,2)\mathbf{Y}(1,1) & \mathbf{E}(2,2)\mathbf{X}(2,2)\mathbf{Y}(2,2) & \cdots & \mathbf{E}(2,N)\mathbf{X}(2,2)\mathbf{Y}(N,N) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}(N,1)\mathbf{X}(N,N)\mathbf{Y}(1,1) & \mathbf{E}(N,2)\mathbf{X}(N,N)\mathbf{Y}(2,2) & \cdots & \mathbf{E}(N,N)\mathbf{X}(N,N)\mathbf{Y}(N,N) \end{bmatrix} du\mathbf{V}^T \tag{A.87}$$

Collecting the eigenvalues of matrix -**A** in vector **p**, we rewrite Equation (A.83) as:

$$\mathbf{k_{m,m}}(t,t') = \frac{l\sqrt{\pi}}{2}\mathbf{V}\int_0^t \left\{\mathbf{E}.*\left[\exp((u-t)\mathbf{p}^T)\left(\left(\operatorname{erf}\left(\frac{u}{l}\mathbf{I}+\frac{\mathbf{p}l}{2}\right)+\operatorname{erf}\left(\frac{t'-u}{l}\mathbf{I}-\frac{\mathbf{p}l}{2}\right)\right).*\exp\left((u-t')\mathbf{p}+\left(\frac{\mathbf{p}l}{2}\right)^2\right)\right)\right]\right\} du\mathbf{V}^T. \tag{A.88}$$

This can be evaluated as:

$$\mathbf{k_{m,m}}(t,t') = \frac{\sqrt{\pi}l}{2}\mathbf{V}\left(\mathbf{E}.*\mathbf{F}\right)\mathbf{V}^T, \tag{A.89}$$

where

$$\begin{aligned}\mathbf{F}(i,j) =& \frac{1}{\mathbf{p}(i)+\mathbf{p}(j)} \\ &\left\{\exp(b1)\left[\operatorname{erf}\left(\frac{\mathbf{p}(j)}{2}l+\frac{t}{l}\right)-\operatorname{erf}\left(\frac{\mathbf{p}(j)}{2}l+\frac{t-t'}{l}\right)\right]\right. \\ &+\exp(b2)\left[\operatorname{erf}\left(\frac{\mathbf{p}(j)}{2}l-\frac{t'}{l}\right)-\operatorname{erf}\left(\frac{\mathbf{p}(j)}{2}l\right)\right] \\ &+\exp(b3)\left[\operatorname{erf}\left(\frac{\mathbf{p}(i)}{2}l-\frac{t}{l}\right)-\operatorname{erf}\left(\frac{\mathbf{p}(i)}{2}l\right)\right] \\ &\left.+\exp(b4)\left[\operatorname{erf}\left(\frac{\mathbf{p}(i)}{2}l+\frac{t'}{l}\right)-\operatorname{erf}\left(\frac{\mathbf{p}(i)}{2}l-\frac{t-t'}{l}\right)\right]\right\},\end{aligned}$$

and

$$b1 = \left(\frac{\mathbf{p}(j)}{2}l\right)^2+\mathbf{p}(j)t-\mathbf{p}(j)t';\quad b2 = \left(\frac{\mathbf{p}(j)}{2}l\right)^2-\mathbf{p}(i)t-\mathbf{p}(j)t';$$
$$b3 = \left(\frac{\mathbf{p}(i)}{2}l\right)^2-\mathbf{p}(i)t-\mathbf{p}(j)t';\quad b4 = \left(\frac{\mathbf{p}(i)}{2}l\right)^2-\mathbf{p}(i)t+\mathbf{p}(i)t'.$$

References

A. Abu-Arish, A. Porcher, A. Czerwonka, N. Dostatni, and C. Fradin. High mobility of Bicoid captured by fluorescence correlation spectroscopy: Implication for the rapid establishment of its gradient. *Biophysical Journal*, 99(4):L33–L35, 2010.

T. Aegerter-Wilmsen, C.M. Aegerter, and T. Bisseling. Model for the robust establishment of precise proportions in the early Drosophila embryo. *Journal of Theoretical Biology*, 234(1):13–19, 2005.

U. Alon. *An introduction to systems biology: Design principles of biological circuits.* Chapman and Hall/CRC, London, 2006.

M. Alvarez, D. Luengo, and N.D. Lawrence. Latent force models. In *International Conference on Artificial Intelligence and Statistics*, volume 5 of JMLR: W&CP 5, pages 9–16, 2009.

S.S. Andrews and D. Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical Biology*, 1:137–151, 2004.

M. Ashyraliyev, K. Siggens, H. Janssens, J. Blom, M. Akam, and J. Jaeger. Gene circuit analysis of the terminal gap gene huckebein. *PLoS Computational Biology*, 5:e1000548, 2009.

M. Barenco, D. Tomescu, D. Brewer, R. Callard, J. Stark, and M. Hubank. Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biology*, 7(3):R25, 2006.

M.J. Beal, F. Falciani, Z. Ghahramani, C. Rangel, and D.L. Wild. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, 21 (3):349–356, 2005.

S. Bergmann, O. Sandler, H. Sberro, S. Shnider, E. Schejter, B.Z. Shilo, and N. Barkai. Pre-Steady-State Decoding of the Bicoid Morphogen Gradient. *PLoS Biology*, 5(2): 965–991, 2007.

S. Bergmann, Z. Tamari, E. Schejter, B.Z. Shilo, and N. Barkai. Re-examining the stability of the Bicoid morphogen gradient. *Cell*, 132:15–17, 2008.

T. Berleth, M. Burri, G. Thoma, D. Bopp, S. Richstein, G. Frigerio, M. Noll, and C. Nüsslein-Volhard. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *EMBO Journal*, 7(6):1749–1756, 1988.

J.M. Bernardo, A.F.M. Smith, and M. Berliner. *Bayesian theory*, volume 62. Wiley, 1994.

G. Brönner and H. Jäckle. Control and function of terminal gap gene activity in the posterior pole region of the Drosophila embryo. *Mechanisms of Development*, 35(3): 205–211, 1991.

B. Calderhead and M. Girolami. Estimating Bayes factors via thermodynamic integration and population MCMC. *Computational Statistics & Data Analysis*, 53(12): 4028–4045, 2009.

S.K. Chan and G. Struhl. Sequence-specific RNA binding by Bicoid. *Nature*, 388(6643): 634, 1997.

D. Cheung, C. Miles, M. Kreitman, and J. Ma. Scaling of the Bicoid morphogen gradient by a volume-dependent production rate. *Development*, 138(13):2741–2749, 2011.

C.M. Child. Patterns and problems of development. 1941.

S.M. Cohen and G. Jürgens. Mediation of Drosophila head development by gap-like segmentation genes. *Nature*, 346:482–485, 1990.

R.L. Cooperstock and H.D. Lipshitz. Control of mRNA stability and translation during Drosophila development. *Cell and Developmental Biology*, 8:541–549, 1997.

M. Coppey, A.M. Berezhkovskii, Y. Kim, A.N. Boettiger, and S.Y. Shvartsman. Modeling the bicoid gradient: Diffusion and reversible nuclear trapping of a stable protein. *Developmental Biology*, 312(2):623–630, 2007.

M. Coppey, A.N. Boettiger, A.M. Berezhkovskii, and S.Y. Shvartsman. Nuclear trapping shapes the terminal gradient in the Drosophila embryo. *Current Biology*, 18(12):915–919, 2008.

N.A.C. Cressie. *Statistics for spatial data, revised edition*. John Wiley & Sons, 1993.

F. Crick. Diffusion in Embryogenesis. *Nature*, 225:420–422, 1970.

A. Dalcq. *Form and causality in early development*. Cambridge University Press, 1938.

S. Dalessi, A.M de Lachapelle, and S. Bergmann. Modeling morphogen gradient formation from arbitrary realistically shaped sources. *Journal of Theoretical Biology*, 294: 130–138, 2011.

M.A. Dewar, V. Kadirkamanathan, M. Opper, and G. Sanguinetti. Parameter estimation and inference for stochastic reaction-diffusion systems: Application to morphogenesis in D. melanogaster. *BMC Systems Biology*, 4(1):21, 2010.

R. Dilão and D. Muraro. mRNA diffusion explains protein gradients in Drosophila early development. *Journal of Theoretical Biology*, 264:847–853, 2010.

H. Driesch. *The science and philosophy of the organism*, volume 2. A. and C. Black, London, 1908.

W. Driever and C. Nüsslein-Volhard. A gradient of bicoid protein in Drosophila embryos. *Cell*, 54(1):83–93, 1988a.

W. Driever and C. Nüsslein-Volhard. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, 54(1):95–104, 1988b.

W. Driever and C. Nüsslein-Volhard. The Bicoid protein is a positive regulator of Hunchback transcription in the early Drosophila embryo. *Nature*, 337(6203):138–143, 1989.

W. Driever, V. Siegel, and C. Nüsslein-Volhard. Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen. *Development*, 109(4):811–820, 1990.

R.O. Duda and P.E. Hart. *Pattem classification and scene analysis*. Wiley, 1973.

S. Engblom, L. Ferm, A. Hellander, and P. Lötstedt. Simulation of stochastic reaction-diffusion processes on unstructured meshes. Technical report, Uppsala University, Sweden, 2008.

A. Ephrussi and D.S. Johnston. Seeing Is Believing: The Bicoid Morphogen Gradient Matures. *Cell*, 116(2):143–152, 2004.

R. Erban, J. Chapman, and P. Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *Arxiv preprint arXiv:0704.1908*, 2007.

R. Erban and S.J. Chapman. Stochastic modelling of reaction–diffusion processes: Algorithms for bimolecular reactions. *Physical Biology*, 6:046001, 2009.

E.L. Ferguson and K.V. Anderson. Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the Drosophila embryo. *Cell*, 71(3):451–461, 1992.

D. Ferrandon, I. Koch, E. Westhof, and C Nüsslein-Volhard. RNA-RNA interaction is required for the formation of specific bicoid mRNA 3' UTR-STAUFEN ribonucleoprotein particles. *EMBO Journal*, 16(7):1751–1758, 1997.

D.A. Fick. On liquid diffusion. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(63):30–39, 1855.

V.E. Foe, G.M. Odell, and B.A. Edger. *The development of Drosophila melanogaster*. Cold Spring Harbor Lab press, 1993.

Y. Fomekong-Nanfack, J.A. Kaandorp, and J. Blom. Efficient parameter estimation for spatio-temporal models of pattern formation: Case study of Drosophila melanogaster. *Bioinformatics*, 23(24):3356–3363, 2007.

Y. Fomekong-Nanfack, M. Postma, and J. Kaandorp. Inferring Drosophila gap gene regulatory network: A parameter sensitivity and perturbation analysis. *BMC Systems Biology*, 3(1):94, 2009.

G. Frigerio, M. Burri, D. Bopp, S. Baumgartner, and M. Noll. Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell*, 47(5):735–746, 1986.

H.G. Frohnhöfer and C. Nüsslein-Volhard. Organization of anterior pattern in the Drosophila embryo by the maternal gene bicoid. *Nature*, 324:120–125, 1986.

J. Gallier. Schur Complements and Applications. *Geometric Methods and Applications*, 38:431–437, 2011.

P. Gao, A. Honkela, M. Rattray, and N.D. Lawrence. Gaussian process modelling of latent chemical species: Applications to inferring transcription factor activities. *Bioinformatics*, 24(16):i70, 2008.

W. Gelbart, L. Bayraktaroglu, B. Bettencourt, and K. Campbell. The FlyBase Database of the Drosophila Genome Projects and community literature. *Nucleic Acids Research*, 31:172–175, 2003.

M.A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry*, 104:1876–1889, 2000.

D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361, 1977.

T. Gregor, W. Bialek, R.R. de Ruyter van Steveninck, D.W. Tank, and E.F. Wieschaus. Diffusion and scaling during early embryonic pattern formation. *Proceedings of the National Academy of Sciences*, 102(51):18402–18407, 2005.

T. Gregor, D.W. Tank, E.F. Wieschaus, and W. Bialek. Probing the limits to positional information. *Cell*, 130:153–164, 2007a.

T. Gregor, E.F. Wieschaus, A.P. McGregor, W. Bialek, and D.W. Tank. Stability and nuclear dynamics of the Bicoid morphogen gradient. *Cell*, 130:141–152, 2007b.

O. Grimm, M. Coppey, and E.F. Wieschaus. Modelling the Bicoid gradient. *Development*, 137(14):2253–2264, 2010.

U. Grossniklaus, K.M. Cadigan, and W.J. Gehring. Three maternal coordinate systems cooperate in the patterning of the Drosophila head. *Development*, 120(11):3155–3171, 1994.

J. Hattne, D. Fange, and J. Elf. Stochastic reaction-diffusion simulation with MesoRD. *Bioinformatics*, 21:2923–2924, 2005.

F. He, T.E. Saunders, Y. Wen, D. Cheung, R. Jiao, P. Rein ten Wolde, M. Howard, and J. Ma. Shaping a morphogen gradient for positional precision. *Biophysical Journal*, 99(3):697–707, 2010.

I. Hecht, W.J. Rappel, and H. Levine. Determining the scale of the Bicoid morphogen gradient. *Proceedings of the National Academy of Sciences*, 106(6):1710–1715, 2009.

M. Hoch, C. Schröder, E. Seifert, and H. Jäckle. cis-acting control elements for Krüppel expression in the Drosophila embryo. *EMBO Journal*, 9(8):2587–2595, 1990.

D.M. Holloway, L.G. Harrison, D. Kosman, C.E. Vanario-Alonso, and A.V. Spirov. Analysis of Pattern Precision Shows that Drosophila Segmentation Develops Substantial Independence from Gradients of Maternal Gene Products. *Developmental Dynamics*, 235(11):2949–2960, 2006.

A. Honkela, C. Girardot, E.H. Gustafson, Y.H. Liu, E.E.M. Furlong, N.D. Lawrence, and M. Rattray. Model-based method for transcription factor target identification with limited data. *Proceedings of the National Academy of Sciences*, 107(17):7793, 2010.

B. Houchmandzadeh, E. Wieschaus, and S. Leibler. Precise domain specification in the developing Drosophila embryo. *Physical Review E*, 72(6):061920, 2005.

B. Houchmandzadeh, E. Wieschaus, and Leibler.S. Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature*, 415:798–802, 2002.

D. Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(17): 2271, 2003.

J. Jaeger. The gap gene network. *Cellular and Molecular Life Sciences*, 68(2):243–274, 2011.

J. Jaeger, M. Blagov, D. Kosman, K.N. Kozlov, Manu, E. Myasnikova, S. Surkova, C.E. Vanario-Alonso, M. Samsonova, D.H. Sharp, and J. Reinitz. Dynamical analysis of regulatory interactions in the gap gene system of Drosophila melanogaster. *Genetics*, 167(4):1721–1737, 2004a.

J. Jaeger, S. Surkova, M. Blagov, H. Janssens, D. Kosman, K.N. Kozlov, Manu, E. Myasnikova, C.E. Vanario-Alonso, M. Samsonova, D.H. Sharp, and J. Reinitz. Dynamic

control of positional information in the early Drosophila embryo. *Nature*, 430(6997): 368–371, 2004b.

M.E. Kavousanakis, J.S. Kanodia, Y. Kim, I.G. Kevrekidis, and S.Y. Shvartsman. A compartmental model for the Bicoid gradient. *Developmental Biology*, 345(1):12–17, 2010.

J. Lam and J.M. Delosme. An efficient simulated annealing schedule: Derivation. Technical report, 1988a.

J. Lam and J.M. Delosme. An efficient simulated annealing schedule: Implementation and evaluation. Technical report, 1988b.

N.D. Lawrence, G. Sanguinetti, and M. Rattray. Modelling transcriptional regulation using Gaussian processes. In *Advances in Neural information Processing Systems*, volume 19, page 785. Citeseer, 2007.

S.C. Little, G. Tkačik, T.B. Kneeland, E.F. Wieschaus, T. Gregor, and C. Desplan. The Formation of the Bicoid Morphogen Gradient Requires Protein Movement from Anteriorly Localized mRNA. *PLoS Biology*, 9(3):1–47, 2011.

W. Liu and M. Niranjan. Matching Models to Data in Modelling Morphogen Diffusion. In *Proceedings of The Third International Workshop Machine Learning in Systems Biology*, pages 55–63, 2009.

W. Liu and M. Niranjan. The role of regulated mRNA stability in establishing Bicoid morphogen gradient in Drosophila embryonic development. *PLoS One*, 6(9):e24896, 2011.

W. Liu and M. Niranjan. Gaussian process modelling for bicoid mRNA regulation in spatio-temporal Bicoid profile. *Bioinformatics*, 28(3):366–372, 2012.

U. Löhr, H. R. Chung, M. Beller, and H. Jäckle. Bicoid - morphogen function revisited. *Fly*, 4(3):236–240, 2010.

D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, 28:3049–3082, 2009.

Manu., S. Surkova, A.V. Spirov, V.V. Gursky, H. Janssens, A.R. Kim, O. Radulescu, C.E. Vanario-Alonso, D.H. Sharp, M. Samsonova, and J. Reinitz. Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Computational Biology*, 5(3):e1000303, 2009a.

Manu., S. Surkova, A.V. Spirov, V.V. Gursky, H. Janssens, A.R. Kim, O. Radulescu, C.E. Vanario-Alonso, D.H. Sharp, M. Samsonova, and J. Reinitz. Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS Biology*, 7(3):e1000049, 2009b.

K.V. Mardia, C. Goodall, E.J. Redfern, and F.J. Alonso. The kriged Kalman filter. *Test*, 7(2):217–282, 1998.

T.H. Morgan. *Regeneration*. Macmillan, New York, 1901.

P.M Morse and H. Feshbach. Gauss's theorem. *Methods of Theoretical Physics, Part I*, pages 37–38, 1953.

E. Myasnikova, A. Samsonova, K. Kozlov, M. Samsonova, and J. Reinitz. Registration of the expression patterns of Drosophila segmentation genes by two independent methods. *Bioinformatics*, 17:3–12, 2001.

I.T. Nabney. *Netlab: Algorithms for pattern recognition*. Advances in Pattern Recognition. Springer, Berlin, 2001.

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

Y. Okabe-Oho, H. Murakami, S. Oho, and M. Sasai. Stable, precise, and reproducible patterning of Bicoid and Hunchback molecules in the early Drosophila embryo. *PLoS Computational Biology*, 5(8):e1000486, 2009.

D. Pe'er, A. Regev, G. Elidan, and N. Friedman. Inferring subnetworks from perturbed expression profiles. *Bioinformatics*, 17(suppl 1):S215–S224, 2001.

B.E. Perrin, L. Ralaivola, A. Mazurie, S. Bottani, and J. Mallet. Gene networks inference using dynamic Bayesian networks. *Bioinformatics*, 19(suppl 2):ii138–ii148, 2003.

M.W. Perry, A.N. Boettiger, and M. Levine. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proceedings of the National Academy of Sciences*, 108(33):13570–13575, 2011.

J.P. Petschek, N. Perrimon, and A.P. Mahowald. Region-specific defects in l(1)giant embryos of Drosophila melanogaster. *Developmental Biology*, 119(1):175–189, 1987.

A. Pisarev, E. Poustelnikova, M. Samsonova, and J. Reinitz. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic Acids Research*, 37(Database issue):D560–D566, 2009.

A. Porcher, A. Abu-Arish, S. Huart, B. Roelens, C. Fradin, and N. Dostatni. The time to measure positional information: Maternal Hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. *Development*, 137(16):2795–2804, 2010.

A. Porcher and N. Dostatni. The Bicoid morphogen system. *Current Biology*, 20(5): R249–R254, 2010.

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, 2004.

C.E. Rasmussen and C.K.I. Williams. *Gaussian processes for machine learning*. MIT Press, Cambridge, MA., 2006.

J. Reinitz, E. Mjolsness, and D.H. Sharp. Model for cooperative control of positional information in Drosophila by bicoid and maternal hunchback. *Journal of Experimental Zoology*, 271(1):47–56, 1995.

J. Reinitz and D.H. Sharp. Mechanism of eve stripe formation. *Mechanisms of Development*, 49:133–158, 1995.

R. Rivera-Pomar, X. Lu, N. Perrimon, H. Taubert, and H. Jäckle. Activation of posterior gap gene expression in the Drosophila blastoderm. *Nature*, 376(6537):253–256, 1995.

H. Roelink, J.A. Porter, C. Chiang, Y. Tanabe, D.T. Chang, P.A. Beachy, and T.M. Jessell. Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. *Cell*, 81(3):445–455, 1995.

K.W. Rogers and A.F. Schier. Morphogen Gradients: From Generation to Interpretation. *Annual Review of Cell and Developmental Biology*, 27(1):377–407, 2011.

T.P. Runarsson and X. Yao. Search biases in constrained evolutionary optimization. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 35(2):233–243, 2005.

K. Sachs, O. Perez, D. Pe'er, D.A. Lauffenburger, and G.P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721): 523–529, 2005.

K. Saha and D.V. Schaffer. Signal dynamics in Sonic hedgehog tissue patterning. *Development*, 133(5):889–900, 2006.

F.J. Salles, M.E. Lieberfarb, C. Wreden, J.P. Gergen, and S. Strickland. Coordinate initiation of Drosophila development by regulated polyadenylation of maternal messenger RNAs. *Science*, 266(5193):1996–1999, 1994.

K. Sander. Pattern formation in longitudinal halves of leaf hopper eggs (Homoptera) and some remarks on the definition of Embryonic regulation. *Development Genes and Evolution*, 167(4):336–352, 1971.

G. Sanguinetti, N.D. Lawrence, and M. Rattray. Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. *Bioinformatics*, 22(22): 2775, 2006a.

G. Sanguinetti, M. Rattray, and N.D. Lawrence. A probabilistic dynamical model for quantitative inference of the regulatory mechanism of transcription. *Bioinformatics*, 22(14):1753, 2006b.

G. Sanguinetti, A. Ruttor, M. Opper, and C. Archambeau. Switching regulatory models of cellular stress response. *Bioinformatics*, 25(10):1280–1286, 2009.

R.M. Schultz. Regulation of zygotic gene activation in the mouse. *BioEssays*, 15:531–538, 1993.

J.R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, 1994.

H. Spemann and H. Mangold. Introduction of embryonic primordia by implantation of organizers from a different species. *Roux's Arch. Entw. Mech*, 100(3):599–638, 1924.

A. Spirov, K. Fahmy, M. Schneider, E. Frei, M. Noll, and S. Baumgartner. Formation of the Bicoid morphogen gradient: An mRNA gradient dictates the protein gradient. *Development*, 136:605–614, 2009.

D. St Johnston, D. Beuchle, and C. Nüsslein-Volhard. Staufen, a gene required to localize maternal RNAs in the Drosophila egg. *Cell*, 66(1):51–63, 1991.

D. St Johnston, W. Driever, T. Berleth, S. Richstein, and C. Nüsslein-Volhard. Multiple steps in the localization of bicoid RNA to the anterior pole of the Drosophila oocyte. *Development (Cambridge, England)*, 107:13, 1989.

M. L. Stein. *Interpolation of Spatial Data: Some theory for kriging*. Springer Verlag, 1999.

G. Struhl, P. Johnston, and P.A. Lawrence. Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell*, 69(2):237–249, 1992.

G. Struhl, K. Struhl, and P.M. Macdonald. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*, 57:1259–1273, 1989.

H.F Stumpf. Mechanism by which cells estimate their location within the body. *Nature*, 212:430–431, 1966.

D. Summerbell, J.H. Lewis, and L. Wolpert. Positional information in chick limb morphogenesis. *Nature*, 244(5417):492–496, 1973.

P. Surdej and M. Jacobs-Lorena. Developmental Regulation of bicoid mRNA Stability is Mediated by the First 43 Nucleotides of the 3' Untranslated Region. *Molecular and Cellular Biology*, 18(5):2892–2900, 1998.

D Tautz. Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. *Nature*, 332(6161):281–284, 1988.

C. Tickle, D. Summerbell, and L. Wolpert. Positional signalling and specification of digits in chick limb morphogenesis. *Nature*, 254(5497):199–202, 1975.

A.M. Turing. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952.

S. Tweedie, M. Ashburner, K. Falls, P. Leyland, P. McQuilton, S. Marygold, G. Millburn, D. Osumi-Sutherland, A. Schroeder, R. Seal, and H. Zhang. FlyBase: Enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Research*, 37:D555–D559, 2009.

G. Von Dassow and G. Schubiger. How an actin network might cause fountain streaming and nuclear migration in the syncytial Drosophila embryo. *The Journal of Cell Biology*, 127:1637–1653, 1994.

O. Wartlick, A. Kicheva, and M. González-Gaitán. Morphogen gradient formation. *Cold Spring Harbor Perspectives in Biology*, 1(3):a001255, 2009.

D. Weigel, G. Jürgens, M. Klingler, and H. Jäckle. Two gap genes mediate maternal terminal pattern information in Drosophila. *Science*, 248(4954):495–498, 1990.

K. Weigmann, R. Klapper, T. Strasser, C. Rickert, G. Technau, H. Jäckle, W. Janning, and C. Klämbt. FlyMove – a new way to look at development of Drosophila. *Trends in Genetics*, 19:310–311, 2003.

D.J. Wilkinson. Parameter inference for stochastic kinetic models of bacterial gene regulation: A Bayesian approach to systems biology (with discussion). In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors, *Bayesian Statistics*, volume 9, pages 679–705. Oxford University Press, 2011.

L. Wolpert. Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology*, 25(1):1–47, 1969.

L. Wolpert. Positional information revisited. *Development*, 107(Supplement):3–12, 1989.

L. Wolpert. One hundred years of positional information. *Trends in Genetics*, 12(9): 359–364, 1996.

L. Wolpert. *Principles of development*. Oxford University Press, 2007.

Y.F. Wu, E. Myasnikova, and J. Reinitz. Master equation simulation analysis of immunostained Bicoid morphogen gradient. *BMC Systems Biology*, 1(1):52, 2007.

T. R. Xu, V. Vyshemirsky, A. Gormand, A. Krieqsheim, M. Girolami, G. S. Baillie, D. Ketley, A.J. Dunlop, G. Milligan, M. D. Houslay, and W. Kolch. Inferring Signaling Pathway Topologies from Multiple Perturbation Measurements of Specific Biochemical Species . *Science Signaling*, 3(113):Ra20, 2010.

G.K. Yasuda and G. Schubiger. Temporal regulation in the early embryo: Is MBT too good to be true? *Trends in Genetics*, 8:124–127, 1992.