THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Stochastic modelling and analysis of early mouse development

Sofia Tapani

Division of Mathematical Statistics
Department of Mathematical Sciences
Chalmers University of Technology and the University of Gothenburg
Göteborg, Sweden 2011

Stochastic modelling and analysis of early mouse development
Sofia Tapani
ISBN 978-91-7385-590-7

Doktorsavhandlingar vid Chalmers Tekniska Högskola
Ny serie nr 3271
ISSN 0346-718X

Division of Mathematical Statistics
Department of Mathematical Sciences
Chalmers University of Technology and the University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Cover: A bright-field image of a mouse egg.
Typeset with LaTeX.
Printed in Göteborg, Sweden 2011

Kärlek

Abstract

The aim of this thesis is to model and describe dynamical events for biological cells using statistical and mathematical tools. The thesis includes five papers that all relate to stochastic modelling of cells.

In order to understand the development and patterning of the early mammalian embryo, stochastic modelling has become a more important tool than ever. It allows for studying the processes that mediate the transition from pluripotency of the embryonic cells to their differentiation. It is still unclear whether the positions of cells determine their future fates. One alternative possibility is that cells are pre-specified at random positions and then sort according to a already set fate. Mouse embryonic cells are thought to be equivalent in their developmental properties until approaching the eight-cell stage. Some biological studies show, in comparison, that patterning can be present already at sperm entry and in the pronuclei migration.

We investigate in Paper I the dynamics of the pronuclei migration by analysing their trajectories and find that not only do the pronuclei follow a noise corrupted path towards the centre of the egg but they also have some attraction to each other which affects their dynamics. Continuing in Paper II and III, we use these results to model this behaviour with a coupled stochastic differential equation model. This enables us to simulate distributions that describe the meeting plane between pronuclei which in turn can be related to the orientation of the first cleavage of the egg. Our results show that adding randomness in sperm entry point is different from the randomness added through the environment of the egg. We are also able to show that data sets with normal eggs and eggs treated with an actin growth inhibitor give rise to considerably different model dynamics, suggesting that the treatment is affecting the migration in an invasive way. Altering the pronuclei dynamics can alter the polarity of the egg and may transfer into the later axis-formation process.

Invasiveness of experimental procedures is a difficult issue to handle. The alternative to invasive procedures is not appealing since it means that important developmental features may not be discovered because of individual variability and noise, leading to guesswork of the underlying mechanisms. The embryonic cells are easily affected by treatments performed to make the measuring, made by hand, easier or by the light exposure of the microscope. Treatments as such are used for example for producing flourescent proteins in membranes or slowing processes down. Paper IV and Paper V serve to analyse how light induced stress affects yeast cells and we employ a method for analysing the noisy non-stationary time series, which are a result of the yeast experiments, using wavelet decomposition.

Keywords: correlation, CUSUM, mouse, non-stationarity, pronucleus, stochastic differential equation, time series, wavelet decomposition, yeast

Acknowledgments

I would like to thank my supervisor Torbjörn Lundh and my co-supervisor Aila Särkkä for being by my side through the process of producing the thesis. You both complement each other in your guidance and have with your support let me explore the interdisciplinary field of stochastic modelling in biology. I am very grateful for your comments, ideas and reading of my manuscripts.

I would also like to thank all my co-authors, in particular Alice Kozakevicius, Kristofer Bodvard and I would like to especially thank Magdalena Zernicka-Goetz for welcoming me in her lab and for her invaluable time and comments. I would like to take the opportunity to thank everyone at Magdalena's lab at the Welcome trust/Gurdon institute in Cambridge for being so friendly and eager to collaborate. Samantha Morris, thank you for sharing your desk and your research with me. The visits in Cambridge have not just been of great benefit for my project but also rewarding on a personal level.

Many thanks goes to my wonderful friends and colleagues at the department. During tough times you are what still makes math fun. I have enjoyed discussing math and life, having fika and going to parties with all of you. Thank you!

Dear friends, I truly hope that you know how important you are to me. I know that writing my thesis has sometimes kept me distant, still you kept cheering for me and picking me up when I felt down. I am so happy that you are my friends, taking me out to nice coffee places, cooking food for me, having drinks with me, dancing with me, convincing me that it is a good day for a dive. I love you all.

The sea, thank you for changing, loving, hugging me and just being a home where I can rest.

Lastly, I wish to thank my family for their support and trust in what I will accomplish. Per, thank you for your never failing love and patience. Thank you for being mine.

Sofia Tapani
Göteborg, October 2011

List of papers

This thesis includes the following papers.

I. Tapani, S. (2011). Data analysis of pronuclei movements using CUSUM change-point analysis of correlation measures. *Preprint.*

II. Tapani, S., Udagawa, J., Plusa, B., Zernicka-Goetz, M., Lundh, T. (2009). Three dimensional mathematical modelling of pronuclei dynamics for the mouse. *Proceedings of the 10th European Congress of Image Analysis and Stereology.* Peer-reviewed.

III. Tapani, S., Zernicka-Goetz, M. (2011). Mathematical modelling of pronuclei migration for the mouse. *Preprint.*

IV. Bodvard, K., Wrangborg, D., Tapani, S., Logg, K., Sliwa, P., Blomberg, A., Kvarnström, M., Käll, M. (2011). Continuous light exposure causes cumulative stress that affects the localization oscillation dynamics of the transcription factor Msn2p. *Biochim. Biophys. Acta. 1813*, 358-366.

V. Tapani, S., Kozakevicius, A. (2011). Wavelet change-point analysis for non-stationary time series. *To appear in The Journal of Wavelet Theory and Applications.*

Contents

Chapter 1

Background

How random is a mouse? The prevailing opinion has been that it is completely random whether cells in the early mouse embryo become the fetus and/or become extraembryonic structures such as the placenta. In many non-mammalian species the development follows a fixed set of instructions. In comparison, mammalian cell fates are very flexible as the cells easily recover from quite extensive perturbations. On the other hand, the early mammalian embryo is not merely a blob of uniform cells. Cells do show some preferences of adopting certain fates according to their initial positions, which in turn will govern their development, and for instance which cells will form the front or the back of a mouse.

We describe these random, but still guided dynamics, with mathematical models. However, a difficulty is that the randomness hides the main guidelines of the dynamics. Also, the cells of the embryos are very sensitive to light exposure, which is necessary during imaging of the processes. We turn to statistical tools to try to tackle these difficulties and to resolve unknown features in the data. To further study how this can be carried out, we develop and apply denoising methods in data from a simpler model organism, namely yeast, which also reacts to light induced stress.

Cellular populations are almost always heterogenous when it comes to their function, (Snijder et al., 2009). The individual differences often overshadow the overall population behaviour, and since the cell is an extremely complex system, it becomes difficult to draw overall conclusions. The activity of cells can come across as completely different even if the cells belong to the same treatment group, ranging in behaviour from stationarity to non-stationarity and transient expressions, which can be hard to comprehend, (Diggle, 1999).

The simple reason for this might be differing measurement techniques but the cells can also be in different states by pure chance from the beginning, see (Schatten and Donovan, 2004). Bio-materials are subject to internal fluctuating processes and are never really in a true state of equilibrium. In understanding this, there is a need for large quantities of data. However, imaging cells is not an easy task as cells are affected by the illumination stress coming from the microscope, they become bleached, their processes are slowed down and altered, or the cells simply die.

On the other hand, really taking care of cells and treating them well to gain lots of uncontaminated data is not enough. Generating a large amount of data can also become a problem in itself, since manual measuring procedures are often time consuming and tedious for even a small data set. Better analysis methods for tracking and quantifying changes in cell populations such as automatic or semi-automatic image analysis methods should be used. However, to be effective on large populations of cells these methods should be developed further. Computational methods that can integrate mathematical modelling with quantitative experimental data can be of great use in understanding the noisy dynamics behind these complex mechanisms.

The evolution of a scientific theory can be seen as a long sequence of iteratively improved models for the function of nature. At each step in the iteration we add some newly discovered phenomena and translate it into the appropriate mathematical (or statistical) language and sometimes we also delete parts that have been proved wrong or unnecessary. The purpose of a model is not to perfectly recreate a particular behaviour but to produce a suitable image of the real world as we are able to perceive it at this particular moment in time. A model needs only to be a sufficient approximation to the real world problem that we investigate by simulations.

The general opinion has for a long time been that polarity and differentiation in the mammalian cells are not decided until after the first cleavage of the embryo and possibly not even until the 32-cell stage. In for example frogs and sea urchin embryos, the up and down, and right and left directions (the embryonic axes' directions) are already decided at the first cleavage. However, recent studies on mouse embryos have indicated that this property could also be the case for mammals. These experiments have been made in three dimensions instead of just two as in the earlier studies. Measurements on embryos studied in two dimensions could be biased due to geometrical constraints on the microscopy slide. Other reasons for biased results could also be the biological effect of the colour used for dying the embryos and chemicals used to slow down the fertilisation process for better laboratory results.

Part of the research of this thesis is done in collaboration with the Magdalena Zernicka-Goetz's Group at the Gurdon Institute in Cambridge. The group studies development of spatial patterning and determination of cell fate of mouse embryos. There are three major areas that the Magdalena Zernicka-Goetz group takes special interest in:

- How polarity in the egg is first established.
- Decisions that allow embryonic cells to shift their initially symmetric patterns into asymmetric ones.
- The influence of preimplantary decisions, for example early cell fate decisions, on the development of embryo patterning after implantation in the womb.

We utilise mathematical modelling together with statistical methods to try to shed some light on some parts of these intriguing questions.

A major problem in this area is the challenging data with its vast individual variability and noise. It can be tough to identify a common trend in the midst of all dissimilarity. To investigate how to handle noisy non-stationary data we have explored the dynamics of a simpler model system, the yeast cell. We investigated illumination stress reactions under different light intensities. Every cell imaging technique includes light sources, so these results can aid in understanding how the cells natural behaviour is altered. Earlier studies have not been able to resolve these effects. However, methods of stationary time series analysis have been the standard in this area so far but they were not useful in our analysis since individual variations smoothed out the results at a too large extent.

Chapter 2

Introduction

This chapter serves to introduce some concepts of developmental biology, cellular measurements as well as of stochastic modelling in biology. This chapter may be useful as a reference when reading the following chapters and to give a biological background for the papers included in the thesis.

First we give an introduction to the development of the mammalian embryo, starting from production of the sperm and the egg cell and ending at the blastocyst formation. Then, we will discuss the ongoing debate in the developmental society that is the basis for a part of the work carried out in this thesis. Last, a brief introduction of the mathematical tools applied in the thesis work is provided.

2.1 From gametes to blastocyst

The purpose of this chapter is to give a short review of the biological background needed. The content of this section is mainly based on (Gilbert, 2006) and (Wolpert et al., 1998). The general production of the reproductive cells is common for most mammals, such as humans and mice. Given here is a description of the complete procedure of the early development, from production of the gametes to the blastocyst formation. For the modelling carried out in this thesis and the papers included we use the background for the fertilisation process and the early embryogenesis.

2.1.1 Production of reproductive cells

In ordinary cell division called mitosis, the DNA duplicates itself by undergoing a number of complex processes and the single cell divides into two complete cells. The purpose of fertilisation is to fuse the genetic material from two individuals. The fusion does not take place in the fertilised egg but is first complete at the two-cell stage. Since the number of chromosomes in a somatic cell needs to be constant, the germ cells must be produced by reduction division, also called meiosis. Diploid precursors for the gametes have two copies of each chromosome, one maternal and one paternal. The number of copies is halved during meiosis so that the gametes contain only one copy of each chromosome and diploid sets of the chromosomes will be restored at fertilisation.

2.1.2 Oogenesis and spermatogenesis

Oogenesis is the production of eggs, also called oocytes. First, the germ cells that form the oocytes divide mitotically a few times and then they enter prophase of the first meiotic division. No further cell multiplication occurs after this division until further development of the egg is induced by sexual maturity of the female. After the ovulation, completion of the first meiotic division follows and the first polar body is formed. The egg is now blocked at the second metaphase of meiosis until fertilisation starts. After fertilisation, the second polar body will be formed and the division of the genetic material is now complete. At this point, the maternal genetic material concentrated in a pronucleus is present in the egg at the position next to the second polar body, see Figure 2.1.

Spermatogenesis is the production of the sperms. Germ cells that develop into sperm enter the embryonic testis and become arrested in the cell cycle. After birth, the cells start to divide mitotically, forming a population of stem cells, which divide meiotically and then differentiate into sperms, as depicted in Figure 2.2.

2.1.3 The fertilisation process

Fusion of the sperm and the egg is the start of the development and is named the fertilisation process. Its purpose is to transfer genetic material from parents to offspring. At the time when the sperm enters, the egg is arrested in metaphase of its second meiotic division and the formation of the female pronucleus is completed as the sperm enters the egg, see Figure 2.3. Cell membranes of

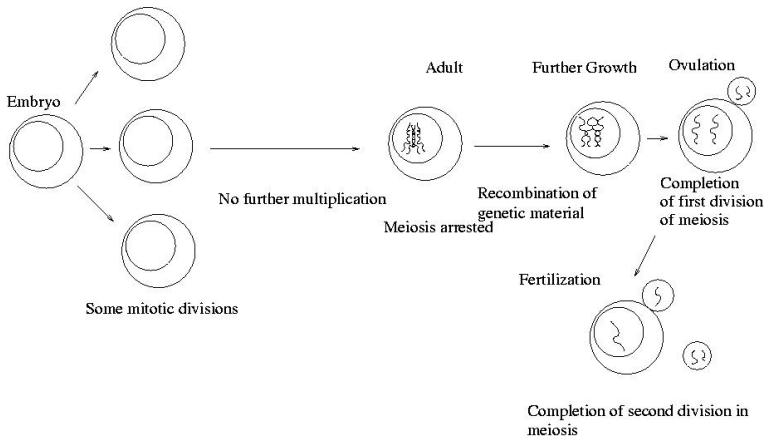


Figure 2.1: Production cycle of the egg cell, during a mammal's life from embryo to adult.

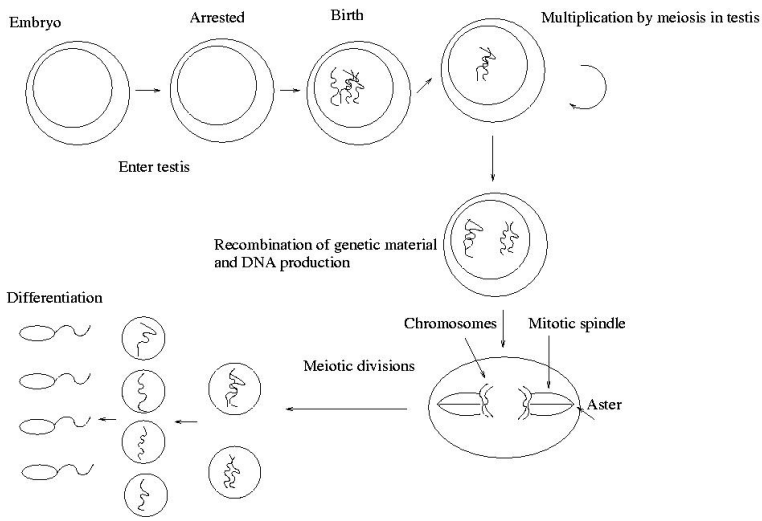


Figure 2.2: Production cycle of sperm, during a mammal's life from embryo to adult.

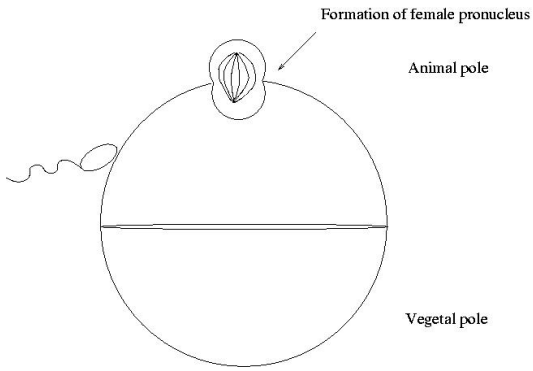


Figure 2.3: A sperm enters the egg (left) and the formation of the female pronucleus is completed. The animal pole is given by the position of the second polar body (this is also referred to as the upper hemisphere of the egg.).

the oocyte, and the sperm fuse and enters the cytoplasm of the egg, where it transforms into the male pronucleus. Simultaneously, the outer membrane of the egg changes allowing no other sperm to enter. In mammals, the fertilisation triggers the completion of the meiosis and thus the formation of the second polar body. Half of the maternal genetic material is left in the zygote (the fertilised egg) and the female pronucleus is now completely formed. In order to the pronuclei to fuse, the male and the female migrate towards each other guided by fibrous structures called microtubules. Microtubules are polymers involved in processes such as cytokinesis and mitosis. They can accomplish both pulling and pushing forces in the cell through employing polymerisation and motor proteins. Motor proteins are contained in the zygote cytoplasm and can move along the microtubules, thus enabling transport in the directions of the microtubule. When the pronuclei meet at some point in the egg, their nuclear envelopes start to break down, which is the start of the first mitosis. DNA reproduction follows separately in the male and female pronuclei and instead of producing a common nucleus in the zygote, a true diploid nucleus in mammals is first visible in the two-cell stage.

Fertilisation is a process rather than an instant event. The pronuclei move quite slowly and hence the speed of the fertilisation differs largely between species. For mammals, the pronuclei migration takes about 12 hours compared to less than one hour for the sea urchin. After the fusion of the genetic material and the first cell division, the cells continue to divide mitotically and form an embryo. The ensuing mitotic events form the start for the next process called embryogenesis.

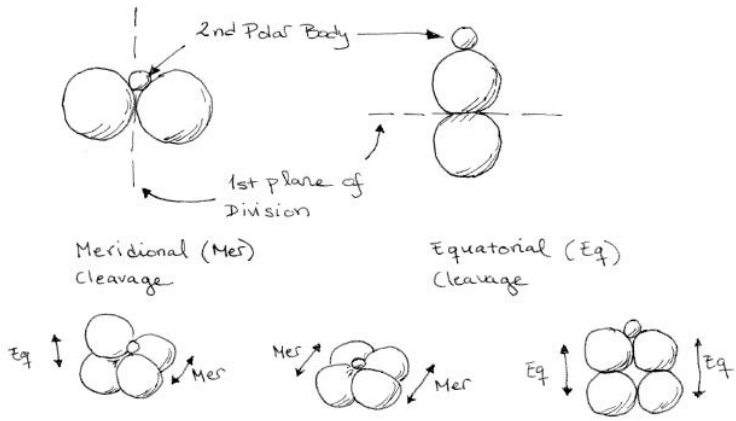


Figure 2.4: Mammalian early development is characterised by rotational division, meaning that the first two divisions are equatorial (Eq) and meridional (Mer) in relation to the second polar body for most early embryos. However there are alternatives, also shown in the figure.

2.1.4 Embryogenesis until blastocyst formation

Mammalian cleavage and patterning are very different from most other patterns of embryonic cell division (Gilbert, 2006). The embryonic cells are called blastomeres and form together a cell structure called the blastula.

The mitotic cleavage process following fertilisation is rotational (Forgacs and Newman, 2005) which means that the first cleavage is a normal cleavage but the direction in relation to the second polar body and sperm entry is still not known (Hiiragi and Solter, 2004; Plusa et al., 2005). For some examples see Figure 2.4. However, in the second round for most embryos, one blastomere will divide meridionally and the other equatorially.

When cells have gone through the third round of division, a process called compaction is initiated. Activation of cell adhesion proteins lead to tighter bonds between cells and the blastomeres form a close ball. Cells on the inside of the ball form gap junctions and exchange molecules and ions between them. At the 16-cell stage, called the morula, most outside cells will stay on the outside to become trophoblast cells (trophectoderm blastomeres). They form the trophectoderm which will go on to build extraembryonic structures such as the placenta, while the inside cells will keep a more pluripotent (can become many things) behaviour. Differentiation into trophoblasts (cells forming the

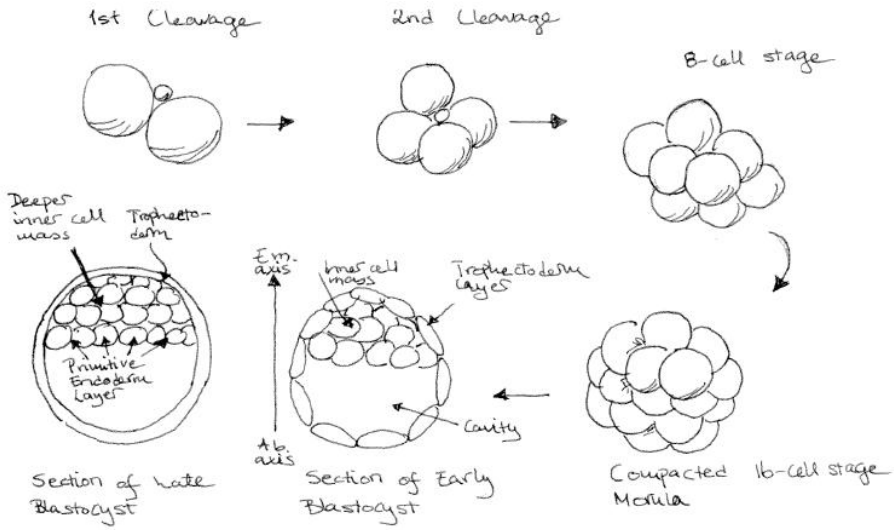


Figure 2.5: A cartoon of the mitotic divisions leading up to blastocyst formation and fixation of the embryonic-abembryonic axis.

outer layer of the blastocyst) is the first cell fate decision event (Wolpert et al., 1998), and once this decision is made, the inner cell mass and the trophoblasts express differing genes. See Figure 2.5 for a cartoon of this process.

The inner cell mass becomes positioned at one side of the ring of trophoblasts during a process called cavitation. A fluid filled volume is created at the other side through the trophoblasts secreting fluid into the inside of the morula. The resulting structure is called the blastocyst, see Figure 2.5. The direction of the dorsal-ventral (back-front) axis in mice correlates with the orientation of the embryonic-abembryonic axis in the blastocyst embryo (Gilbert, 2006). The inner cell mass that is in contact with the trophoblast will form the dorsal axis while the cells in contact with the blastocyst fluid are primitive endoderm cells and give rise to the ventral axis.

2.1.5 Randomness vs rigidity

Some claim that prepatterning is already present in the egg and some that the dorsal-ventral axis is not formed until the blastocyst stage, the latter meaning that the axis formation could even be affected by early events such as the entry

of the sperm. To resolve this disagreement, statistical methods and computer simulations can be utilised to set up different models for the two hypotheses and compare them to each other and their prediction of real data. A natural question then to be asked is: How do we formalise this in mathematical terms? In (Honda et al., 2008), a 3D vertex dynamics model for a cell aggregate is presented. It is stated that their model, driven by chance, can give rise to the seemingly ordered blastocyst shape of the embryo by just applying geometrical mechanistic properties to the aggregate. However, (Dietrich and Hiiragi, 2007) identified critical events leading to divergence of the trophectoderm, epiblast and primitive endoderm lineages in the embryo. They found out that Oct4, a transcription factor for pluripotency, is present in all cells up to the blastocyst stage. However, Oct4 gradually disappears from the trophectoderm cells after this. Other conclusions include that Cdx2 and Nanog transcription factors are randomly distributed in cells at the first stages, though a definite pattern is established during compaction, indicating early patterning through random processes and cell sorting. Outside cells become committed to trophectoderm at the blastocyst stage through expression of Cdx2. However, (Jedrusik et al., 2008) showed that Cdx2 can influence cell allocation even earlier. Increasing Cdx2 levels in individual blastomeres promotes symmetric divisions, see Figure 2.6, and as a consequence these cells contribute more to the trophectoderm. In (Bischoff et al., 2008) it was investigated through tracking of the cell lineages from the two-cell stage to the blastocyst stage, whether the tendency to divide symmetrically or asymmetrically showed any spatial or temporal pattern. The results showed that the majority of the individual blastomeres give rise to distinct blastocyst regions. They found out that the blastocyst cavity, defining the embryonic-abembryonic pole, forms were symmetric divisions predominate and that blastomere origin in relation to the animal-vegetal axis in the zygote can influence the pattern of symmetric/asymmetric divisions.

It is certainly not clear whether the positions of cells determine their future fate decisions or if cells are pre-specified at random positions and then through a cell sorting process position themselves correctly. Blastomeres of the early mouse embryo are generally considered to be equivalent in their developmental properties until the eight-cell stage. In (Piotrowska-Nitsche et al., 2005) it is shown that this is not always the case. A pattern of second cleavage divisions in which a meridional division is followed by an equatorial division allows for identification of blastomeres with differing fates and developmental properties. One blastomere that inherits some vegetal membrane (the side of the egg cell furthest away from the second polar body) will give rise to more trophectoderm and the sister blastomere donates more cells to the inner cell mass showing polarity at this early stage.

In (Meilhac et al., 2009) simple computer models for cell sorting movements and

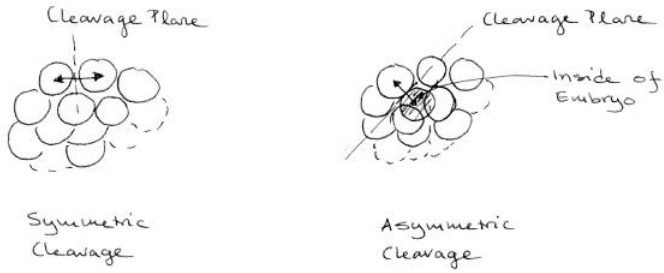
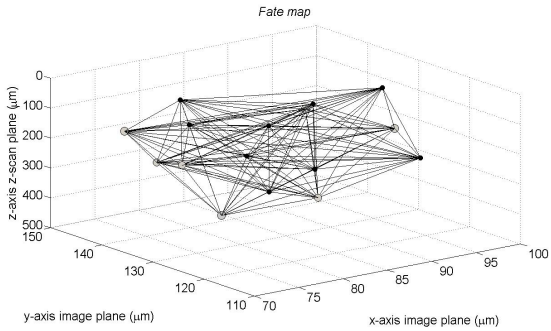


Figure 2.6: Example showing differences between symmetric and asymmetric divisions. Symmetric division takes place when the two daughter cells retain the position of the mother cell, i.e. an outer cell produces two outer cells. Asymmetric division means that the mother cell produces one inner and one outer daughter.

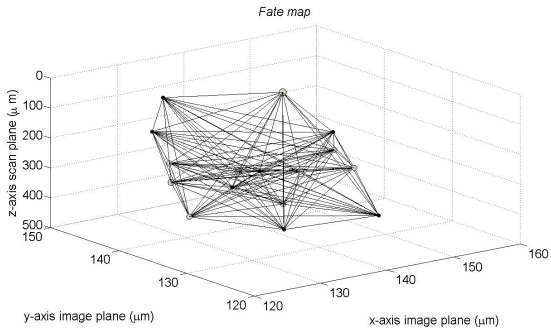
position dependent induction suggest that both these components are present during primitive endoderm formation. Furthermore, (Morris et al., 2010) investigated how the origin of embryonic pluripotent cells contributed to the inner cell mass and to primitive endoderm cells positioned at the top of the cavity in the blastocyst. The primitive endoderm cells form an important signalling centre for the dorsal-ventral axis formation. Cell lineages were traced back to the 8-16 cell stages and the question was if these primitive endoderm progenitors tend to cluster and share the same origin. Figures 2.7(a) and 2.7(b) show two fate maps of 16-cell embryo blastomeres plotted to show the geometry of an early 16-cell embryo. The idea is to study these maps to see if there is some patterning present.

2.2 Stochastic modelling tools

Mathematical biology is an interdisciplinary science, which is shown by the diversity of methods used in this thesis. We are applying a variety of procedures to describe and interpret some biological cellular processes. The main mathematical and statistical tools used in this thesis are stochastic differential equations, non-parametric cumulative sum methods and wavelets. I will give a short description of each method.



(a)



(b)

Figure 2.7: Fate maps of 16-cell embryo blastomeres. Grey blastomeres denote primitive endoderm progenitors. The size of each blastomere marker is proportional to the number of primitive endoderm cells the progenitor gives rise to. Edges are drawn to connect every cell to every other cell in the embryo in order to give a simplistic outline of the overall shape.

2.2.1 Dynamical systems and SDE:s

When modelling physical quantities that vary over time, such as the pronuclei migration for the mammalian embryo, the need for a model which relates the derivative of a function to the function itself arises. The motion of a pronucleus is described by its position and its velocity. Differential equations are equations that state such a relationship. Since we do not assume that the motion is completely deterministic we have chosen a stochastic model. A stochastic differential equation (SDE) is a differential equation where one or several terms are stochastic (Øksendal, 2003). The solution to an SDE is itself a stochastic process. SDE:s are commonly used to model fluctuating systems such as stock prices and thermal variations. This is useful when having an experimentally measured trajectory which does not behave exactly as predicted but appears to be contaminated by noise. A typical SDE can look like

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dB(t) \tag{2.1}$$

where B(t) is a Wiener process (Karatzas and Shreve, 1991) or a standard Brownian motion, defined below. The function $\mu(X(t), t)$ is the drift or mean motion of the model, and $\sigma(X(t), t)$ is the standard deviation function. Solving Equation 2.1 should be interpreted as a search for the function X(t) that solves the integral equation

$$X(t) = \int_0^t \mu(X(s), s)ds + \int_0^t \sigma(X(s), s)dB(s) \quad \text{for all } 0 < t < T$$

where the first integral is an ordinary Lebesgue integral, the second integral is a stochastic Itô integral (Karatzas and Shreve, 1991) and $T \in \mathbb{R}^+$.

A real-valued stochastic process B(t) is called Brownian motion if

(i) $B(0) = 0$ almost surely,

(ii) $B(t) - B(s)$ is normally distributed N(0,t-s) for all $0 \leq s \leq t$,

(iii) for all times $0 < t_1 < t_2 < \cdots < t_n$, the random variables $B(t_1), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ are independent (independent increments).

An example of a simulated Brownian trajectory can be seen in Figure 2.8.

In this thesis, we generate a set of three dimensional (since the murine zygote is of course a 3D system) sample paths $\hat{\mathbf{B}}(\omega) = (B^1(\omega), B^2(\omega), B^3(\omega))$ on an interval $[0, T]$. Each such sample path generates an approximate solution x_t to

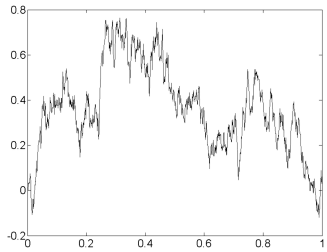


Figure 2.8: A realisation of the Brownian motion process, starting in 0.

our model SDE by using a truncated Taylor expansion of the real solution x on successive small subintervals of $[0, T]$, (Øksendal, 2003).

2.2.2 Non-parametric CUSUM

In three of the papers in the thesis we use a non-parametric statistical change-point algorithm employing cumulative sums (CUSUM). The technique is used to monitor change detection in statistical quality control. The particular version of the method applied in the thesis was presented in (Taylor,). Change-point analysis by calculating cumulative sums is used when over-viewing stochastic processes for changes in parameter values (Page, 1954; Brockwell and Davis, 1991). If a change of the mean value is to be detected we perform the steps presented in Algorithm 1.

First, the cumulative sums, S_{i+1}^b, $i = 1, \ldots, n-1$, as well as the corresponding maximum, $S_{\max}$, and minimum values, $S_{\min}$, of S_i^b are calculated . After calculating the test statistic, given by the reference difference R_{diff} in Algorithm 1, a bootstrap analysis (permutation test) is performed according to (Efron and Tibshirani, 1993) and (Manly, 1997) to find the significance of the candidate change-point. The permutation of the observed data, denoted by P_b is simply a random reordering done at each iteration $b > 0$, and P_0 denotes the original data without permutation. We determine how many of the permuted sample differences are less than the difference calculated from the original data, our R_{diff} value. The confidence level of the change-point is calculated as the proportion of bootstrap sample differences less than or equal to the difference of the data sample in B bootstrap repetitions, as suggested in (Taylor,).

Candidate points are identified by the bootstrap test at a desired confidence level. If a significant change has been detected, we take the position m associ-

Algorithm 1: Estimation of Change-Point Candidate

input : x_n : Input data; n: length of data;
B : Maximal bootstrap iterations; DCL : Desired confidence level

output: m : Change-point position

Bootstrap analysis
counter $\leftarrow 0$
for $b \leftarrow 0$ **to** B **do**
  $Y \leftarrow P_b(x_n)$
  $\bar{Y} \leftarrow \frac{1}{n}\sum_{i=0}^{n-1} Y_i$
  Cumulative sum
  $S_0^b = 0$
  for $i \leftarrow 1$ **to** $n-1$ **do**
    $S_i^b = S_{i-1}^b + (Y_i - \bar{Y})$
  Maximum, minimum and difference variations
  $S_{\max}(b) = \max_{i=0,\ldots,n-1} S_i$
  $S_{\min}(b) = \min_{i=0,\ldots,n-1} S_i$
  $S_{\text{diff}}(b) = S_{\max}(b) - S_{\min}(b)$
  $R_{\text{diff}} \leftarrow S_{\text{diff}}(0)$
  if $S_{diff}(b) < R_{diff}$ **then**
    counter $\leftarrow$ counter+1
Determination of significance
$p \leftarrow (\text{counter}/B)$
if $p > DCL$ **then**
  $|S_m| = \max_{i=0,\ldots,n} |S_i^0|$
  $m \leftarrow i_*$: position where $|S_m|$ occurs
if *No change point is found* **then**
  $m \leftarrow n$
return m

ated to $|S_m|$ as the estimator for the actual time the change has occurred.

An example of a CUSUM chart of the entire time series is shown in Figure 2.9. A possible change point is seen at time point 14. From these charts we can decide whether the observations tend to be above or below the mean value. A descending part shows a tendency for the observation to be below the mean and an ascending part a tendency to be above the mean.

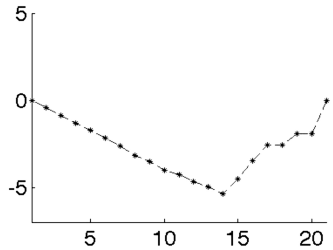


Figure 2.9: A CUSUM chart of a time series, note that a possible change has occured at the time point 14.

2.2.3 Wavelet filtering

Wavelets are characterised by scale and position, and are therefore useful in many signal processing problems and for analysing variations in data (Percival and Walden, 2000).

A wavelet is a small wave or oscillation that has a finite reach. It has limited duration and zero average, unlike in Fourier analysis, where sine waves are infinite and smooth (Bloomfield, 1976). Wavelets are often designed to have specific properties that suit the application. Wavelets are combined into families with shift, scale and sum methods, and to extract information from a signal the convolution operation is used. The application fields are many, such as image analysis, signal analysis and decompression of information, see for instance (Gonzalez and Woods, 2006). The scale aspects give us the opportunity to analyse data on different scales, for instance characterising the texture of a cellular or a material surface or detecting areas of quick variation in stock prices. The localisation aspects offer us tools to handle edge detection and short term phenomena like short changes in amplitude or variations of a signal. Wavelet analysis breaks up the signal into scaled and shifted versions of the mother wavelet. With wavelet analysis we are able to perform localised analysis of data, which is an advantage when handling non-stationary time series. Stationary methods would miss change-points, smooth out trends and not detect self-similarity.

Decomposition of a signal with wavelets is carried out at a fixed number of levels, so that the signal is split into an approximation (or a trend) and several details corresponding to the number of levels. We are able to resolve variations in the details with respect to the trend, which makes wavelets an effective denoising tool. In fact, when decomposing data in a multi-resolution way, data with different types of trends can be better handled and at the same time

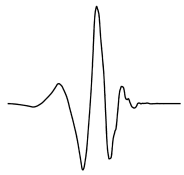


Figure 2.10: The Daubechie 4 mother wavelet.

short-term oscillatory behavior can be better analysed and determined.

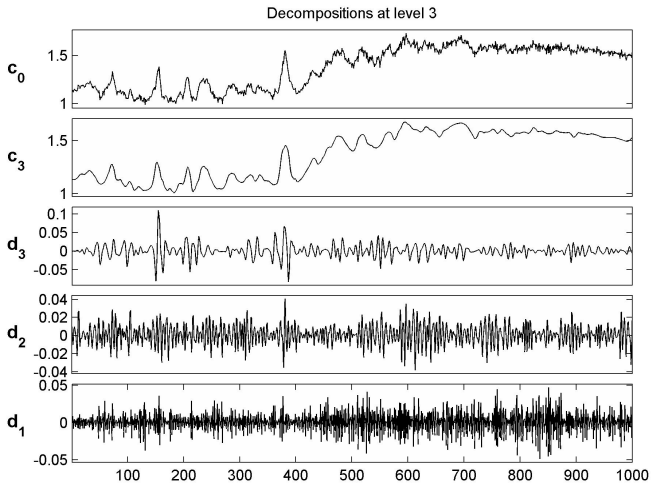
For the analysis in the thesis we use the so called Daubechies wavelet family see Figure 2.10, (Percival and Walden, 2000; Flandrin, 2000). These wavelet functions form an orthonormal basis for the vector space of square-integrable functions. Since the wavelets of the Daubechie family are orthonormal, the decomposition will result in independent details, which is useful since the statistical analysis is based on an independence assumption. Assume that we have a set of measurements, $\{x_t : t = 0, \ldots, n_0 - 1\}$, where the one-dimensional temporal process X_t is observed at n_0 consecutive equally spaced time points. Assume also that the initial amount of discrete data n_0 is a power of two, $n_0 = 2^{p_0}$, where p_0 is a positive integer.

The discrete wavelet transform of a signal $c_0 = x_t$ is given by passing the signal through both a low-pass filter g and a high-pass filter h by a scalar product operation recursively as

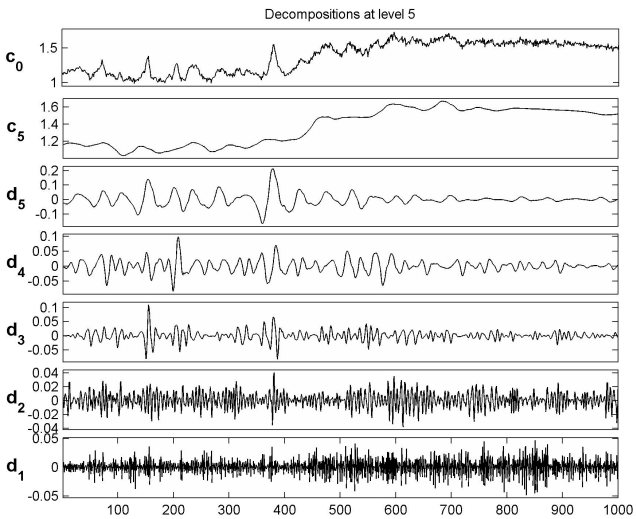
$$\begin{aligned} c_1[k] &= \sum_{i=0}^{D-1} h_i c_0[2k+i], \\ d_1[k] &= \sum_{i=0}^{D-1} g_i c_0[2k+i]. \end{aligned}$$

Here, D is the number of filter components, which depends on the chosen family, $c_1[k]$ and $d_1[k]$ are the kth components in the trend array of the decomposed signal and detail respectively. h_i and g_i are filter components of the high-pass and the low-pass filter, given by the chosen wavelet family. In the decomposition, half of the frequencies are removed so that the array has half the length of the original signal. $c_1[k]$ contains the low frequency information of the signal while the higher frequencies are resolved in the detail $d_1[k]$. By increasing the number of levels we are able to resolve finer and finer frequencies and the trend will become smoother. For a comparison of different choices of the number of

levels, see Figures 2.11(a) and 2.11(b).



(a) Three levels.



(b) Five levels.

Figure 2.11: A wavelet decomposition with dB4 wavelet family. c_0 denotes the signal, c_5 is the approximation and d is the corresponding high frequency details in the decomposition.

Chapter 3

Summary of papers

Five papers are included in this thesis, the first three papers cover mathematical and statistical methods for analysing and modelling the early events of the developing mammalian egg cell. The remaining two papers present the statistical methodology which was originally used in Paper I but developed further to analyse how yeast cells respond to light induced stress. Given in this chapter are brief summaries of these five papers and contributions of the author of the thesis are stated.

Paper I: Data analysis of pronuclei movements using CUSUM change-point analysis of correlation measures.

The purpose of the first paper is to present an initial analysis of data on pronuclei migration in the murine zygote. Manual observations were the basis of the hypothesis that the pronuclei not only move towards the centre of the egg but that they also have some attraction to each other. To investigate this possible attraction between pronuclei, we analyse data of the fertilisation process in 3D by different measures of spatial correlations. Our data consist of measurements from stacked bright field images of the fertilisation process. At each time point we can extract 3D coordinates for the two pronuclei.

Three different measures were analysed: Axis-wise correlation, spherical correlation and a projection measure. The projection measure is easier to interpret since the dimensionality is decreased and it is geometrically more intuitive. It is based on the projection of each pronuclei trajectory onto the axis in the direction towards the other pronucleus. This measures only the movements that are made in this direction and as such can be classified as possible attraction. Next we analyse the correlation of these projections for the pronuclei. We observed that when applying the correlation measures to the data, the migration dynamic was subject to phase changes over time. The first phase-change can be identified to relate to when pronuclei "hook-up" and attract each other, and the second when they have already attached themselves and start moving towards the centre of the egg. To find the significance of these events we used a CUSUM change-point analysis. We found this behaviour to be significant for all tested eggs and we also found differences between non-treated and drug-treated eggs.

The author of the thesis is the sole author of Paper I.

Paper II: Three dimensional mathematical modelling of pronuclei migration for the mouse

The question addressed in this paper is what happens between the sperm entering the egg and the fusion of the two pronuclei. Orientation of the opposing pronuclei probably plays a decisive role in the polarity of the developing embryo. Observations from Paper I led us to believe that not only is the migration directed towards the centre but also towards the other pronucleus. In this paper we introduce a basic model of the migration dynamics based on two forces of attraction, one towards the centre of the egg and one towards the other pronucleus. We use a stochastic differential equation model to capture the random fluctuations of the pronuclei paths.

Data from non-treated eggs are compared with data of eggs treated with Cytochalasin B, which is known to be an inhibitor of actin filament growth and to slow down the migration. We find significant changes in the dynamics by simulations from our models with parameters estimated from the two data sets.

The paper is co-authored with Jun Udagawa, Berenika Plusa, Magdalena Zernicka-Goetz and Torbjörn Lundh. The author of this thesis has contributed to the paper as main author of the paper and by major involvement in the modelling, simulation work and in the analysis of the results.

Paper III:
Mathematical modelling of pronuclei migration for the mouse

This paper is a continuation of Paper II, presenting new data and introducing a new model for pronuclei migration. Movements are directed towards the centre but also towards the other pronucleus. We present two models for the migration dynamics, including the basic one already shown in Paper II. The refined version, divides the centring force into two components of pushing and pulling mechanisms for accomplishing a movement in this direction.

Data of non-treated eggs are again compared with data of eggs treated with Cytochalasin B. We find significant changes to the dynamics by simulations from our models with parameters reestimated from the new data set. However, we do find some differing results from Paper II in the parameter results which we also are able to give an explanation for. Mainly the difference between the new data set and the old data set used in Paper II, is the higher resolution in the z-scan axis which means that we are able to resolve the migration with higher confidence in the results.

We also show that adding stochasticity dynamically during the migration leads to completely different patterns than using a deterministic path with noise added to the initial starting points.

The paper is co-authored with Magdalena Zernicka-Goetz and Torbjörn Lundh. The author of this thesis has contributed to the paper as main author and has had major involvement in the modelling, data acquisition, simulation work and analysis of the results.

Paper IV:
Continuous light exposure causes cumulative stress that affects the localization oscillation dynamics of the transcription factor Msn2p

Paper IV is a biologically oriented paper examining light induced stress in yeast cells. The exposure of cells to light is a potentially powerful stress factor during in vivo optical microscopy studies and is hence fundamental in many biological applications and studies. In yeast, the general transcription factor Msn2p translocates from the cytoplasm to the nucleus in response to illumination. Previously stationary analysis methods were used, but proved to be unsuccessful to model the behaviour. In this paper it we investigate how continuous illumination affects the localisation pattern of Msn2p-GFP in budding yeast. The localisation pattern was analysed using a novel approach that combines wavelet decomposition and change-point analysis. It was found out that the Msn2p nucleocytoplasmic localisation trajectories for individual cells exhibit up to three distinct and successive states: Msn2p localises to the cytoplasm, Msn2p rapidly shuttles between the cytoplasm and the nucleus, Msn2p localises to the nucleus. Many cells pass through all states consecutively at high light intensities, while at lower light intensities most cells only reach the first two states. This behaviour strongly indicates that continuous light exposure gradually increases the stress level over time, presumably through continuous accumulation.

The paper is co-authored with Kristofer Bodvard, David Wrangborg, Katarina Logg, Piotr Sliwa, Anders Blomberg, Mats Kvarnström and Mikael Käll. The author of the thesis has developed the statistical method utilised, made major contributions to the statistical analysis and wavelet method used, contributed to the writing process.

Paper V:
Wavelet change-point analysis for non-stationary time series

Here, we present the methodology behind the analysis in Paper IV. The yeast data are used as a validation example of a non-stationary time series for the developed algorithm.

An orthonormal wavelet transform together with a CUSUM change-point analysis is used to precisely localise positions where the non-stationary time series changes its features. We are able to decompose the time series in different scales due to the multi-resolution feature of the wavelet transform. This enables us to investigate stationary regions of the time series.

Because of the properties of the wavelet transform, it serves as a method to identify change of events and data properties with much more confidence and hence should be of interest to other applications than biological ones. Used as a preprocessing step it can be seen as a trend identifier and remover (if necessary), and signals can be analysed at multiple levels simultaneously. This is due to the multilevel frame work offered by the wavelet transform.

The paper is co-authored with Alice Kozakevicius. The author of this thesis has contributed to the paper as the main author and by a major involvement in developing the methodology, modelling, simulation work and analysis of the results.

Chapter 4

Future work

This chapter is a summary of collected ideas to pursue in the future for the different projects presented in this thesis. In the first two sections we have given some ideas on how to improve the work presented so far. The last two sections describe in more detail two ongoing projects.

4.1 Modelling the fertilisation process

The next challenge is to model the cytoskeleton on a finer scale and especially look into the mechanics of actin working on the microtubules (Footer et al., 2007). It has recently been shown that fertilisation locally reorganises the actin cytoskeleton around the site of sperm entry (Ajduk et al., 2011). This leads to the formation of the fertilisation cone and rhythmic contractions of the actin cytoskeleton throughout the phase of pronuclear movement. We believe that understanding these mechanisms will be the key to explaining the movements of the pronuclei in future models.

We would prefer to obtain a model that is more in consistence with the developmental biology literature. The noise should be more thoroughly modelled, as we did not take several different sources of noise into account in our models. An improved noise model should take into consideration the dynamics of the microtubule growth cycle, thermal vibrations of the egg cortex, the binding and unbinding of motor proteins along the microtubule and the errors from the measuring techniques.

Our intention is to make use of better analysis methods such as semiautomatic or automatic image analysis methods for tracking the pronuclei in the image sequences. The manual measuring procedure is not an easy task and should be performed by an experienced person with good knowledge of the migration process. Automatic tracking by image analysis should be done in order to minimize sources of human error as much as possible and to enable more efficient analysis of pronuclei trajectories. This would make data more easily available, more accurate and less subjective.

4.2 Wavelet decomposition

For further analysis, with the yeast application in mind, we would like to expand the method used in order to create a classifier for the different stress-states that the cells go through. We believe that the method has more potential than shown since the information at different detail decomposition levels have not yet been fully explored. Furthermore, extensive work is still carried out manually in this project and it would be a great improvement to automate it fully.

The method is not limited to the particular application presented in Paper IV. It can be used as a preprocessing step to almost all types of non-stationary time series. Due to the properties of the wavelet transform, it serves as a method to identify change of events and data properties with much more confidence than before. The algorithm consists of a multilevel framework and can be seen as a natural trend identifier and remover. Therefore, we would like to examine how the method performs for other applications, such as finance, electrical engineering and medicine. Specifically it would be interesting to analyse the performance of the method for detection of periods of changing volatility together with changing trends for stock prices. Another possibility would be to analyse time series of electricity prices and observe how certain events in nature would affect the variation of the prices.

We also intend to further investigate the mathematical properties of our wavelet framework to see if we can utilise the method in a more effective way. We would also like to compare how our method performs with other available filtering methods for non-stationary time series.

4.3 Testing for clustering of progenitor cells in the early embryo

As described in the introduction of this thesis, the primitive endoderm cells form an important signalling centre for the future dorsal-ventral axis formation. One hypothesis is that the progenitors of these cells tend to share origin in the early 16-cell embryo. Another hypothesis is that the origins are completely random at this stage. In order to measure the amount of clustering of the primitive endoderm progenitors we need to take several factors into account. Ideally, factors like individual cell size, cell contacts and other geometric constraints should be considered. However, for large data materials (large sets of embryos), it is cumbersome to measure all these factors and difficult to account for all types of clustering, see examples of two types of different clusters in Figure 4.1. Instead, we use a statistical method that takes into account the specific geometry of each embryo at the 16-cell stage and assigns a test statistic to the embryo. The value of the test statistic will then be compared to the value in the case of complete spatial randomness of positions of primitive endoderm progenitors. The null hypothesis we choose is that the primitive endoderm progenitors are equally likely to be assigned to any cell position in the 16-cell embryo, where the spatial cell positions are given by the original embryo from the data. The geometry of the embryo is kept and each blastomere is plotted at its original spatial position. All cells have the same probability of giving rise to any number of primitive endoderm cells given by the original data embryo. The alternative hypothesis is that the primitive endoderm cells tend to have the same origin in the 16-cell embryo. How do we interpret what it means to have the same origin? We have chosen a test statistic which measures the degree of clustering in terms of the primitive endoderm cells being inclined to be closer to each other than to other groups of cells. To test the hypothesis of the primitive endoderm progenitors arising from anywhere in the embryo we have taken into account the specific geometry of the embryo by keeping the cell positions fixed. Distances between cells at the 16-cell stage giving rise to primitive endoderm at the blastocyst were weighted according to the number of primitive endoderm cells derived. We suggest using the following statistic

$$S = \sum_i \frac{1}{N_i} \sum_j d_{ij}, \tag{4.1}$$

where d_{ij} is the distance between progenitors i and j, and N_i is the number of primitive endoderm cells that progenitor i gives rise to. We use this formulation to assure that cells producing more primitive endoderm are recognised as of greater importance. S is the sum of the average distances between the

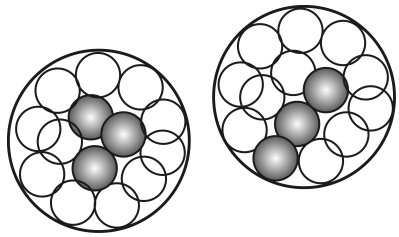


Figure 4.1: Two schematic examples of different clusters (grey) of primitive endoderm cells. The cluster to the left is more spatially concentrated than the one on the right. Both clusters will however be detected by the test statistic.

progenitors; the sum is over all primitive endoderm cells. Each distance between the cells contributing to the primitive endoderm population is weighed inversely to the number of primitive endoderm cells it gives rise to. The test is conditioned on $N_i > 0$ given by the data. These distances are then summed over the progenitors and compared to the approximate randomisation (Efron and Tibshirani, 1993) distribution of Equation (4.1). Therefore we randomise the positions, for example 1000 times. For each run of the randomisation we calculate the same statistic to obtain the randomisation distribution. A low value of the observed statistic compared to the statistics from this distribution will indicate that the primitive endoderm cells significantly tend to come from the same area in the 16-cell embryo; this is then interpreted as clustering. It should be noted that the randomisation distribution does not exclude all types of clustering. However, since clustering can occur at random in the embryo, this is a desired trait. Our null hypothesis represents complete spatial randomness and is to be compared to a non-random pattern indicating clustering. The p-value is derived from Monte Carlo simulations by

$$p = \frac{\sharp\{\text{samples} \leq \text{data}\}}{\sharp\{\text{samples}\}}.$$

Unpublished results indicate that there is significant clustering in a majority of the 16-cell stage embryos tested. Further, we would like to obtain a more extensive data set to verify our preliminary results and to investigate the properties of the test statistic in more detail, and to study whether the clustering test statistic could be developed further to be more intelligent and find more specific types of clusters.

4.4 A dynamical cell game

One way to investigate whether the origins are purely random is to use some type of permutation test as described in Section 4.3. Another possibility would be to build a probabilistic model, which includes the different states that cells go through during early development. A model of this kind can aid in understanding if cell fates are governed by their position or prepatterned in the early embryo. In the current stage the model contains only cell division and migration, however features such as apoptosis (cell death) would be straight forward to incorporate. We will in the following sections present suggestions on how to model events such as migration and mitotic divisions, as well as finally describe the full model at its current state. The ideas are described in 2D but the final future model will be in 3D.

The Grid and the Cellular units

We model the development of the early mammalian embryo by an aggregate of cells following a few simple rules. The cells can alter their gene expression or state, divide, migrate or die. Each cell will be affected only by its local neighbourhood as a cellular automata.

An embryonic cell is assumed to have a preferred circular shape. However, as cells move or divide, the early embryo or cell aggregate will change shape and the cells may be compressed or elongated. To allow for this we intersect the circle aggregate with a Voronoi grid determined by the centre of the circles as generator points (Okabe et al., 2000).

Let $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ be the locations at the cell centres and $\mathbf{x}(t)$ be the location of an arbitrary point in $\mathbb{R}^2$, at time t. Assuming that no two points coincide we define the aggregate as the dynamical Voronoi diagram defined by the set $\mathcal{V} = \{V(\mathbf{x}_1, t), \ldots, V(\mathbf{x}_n, t)\}$, where

$$V(\mathbf{x}_i, t) = \{(\mathbf{x}, t) \mid \|\mathbf{x}(t) - \mathbf{x}_i(t)\| \leq \|\mathbf{x}(t) - \mathbf{x}_j(t)\|, j \neq i\},$$

and $\mathbf{x} \in \mathbb{R}^2, t \in \mathbb{R}^+$. For a fixed point in time the dynamical Voronoi diagram is simply the ordinary Voronoi diagram. The set $\mathcal{V}$ is intersected with the set of circles

$$\mathcal{S}(t) = \bigcup_{i=1}^{n} b(\mathbf{x}_i(t), r_{x_i}),$$

where $b(\mathbf{x}_i(t), r_{x_i})$ is a circle with centre in $\mathbf{x}_i(t)$ and radius r_{x_i}. Each unit cell

contains a mark which remembers its ancestral mother from the two cell stage and possibly also gene expression.

Migration

In this section we present a flexible migration model which allows for modelling attractional movement in the aggregate between either cells of the same type or cells of different types or both. By changing parameter values we can obtain different test scenarios (for instance the two scenarios mentioned earlier) and compare to biological data values. Migration is modelled in 2D space by using the von Mises distribution (Fisher et al., 1987), which is the circular analogue to the normal distribution on a line. The distribution has a mean direction given by its location parameter α and its shape by the concentration parameter κ. The larger the value of κ, the more concentrated is the mode of the distribution towards the direction α. The probability density function is given by

$$f(\theta) = C \exp[\kappa \cos(\theta - \alpha)], \ (0 \leq \theta \leq 2\pi)$$

where the constant

$$C = \frac{1}{2\pi I_0(\kappa)},$$

and where the function I_0 is the modified Bessel function of zeroth order. Location and concentration parameters of the migration are given by the local neighbourhood of the specific cell position which is to be updated. The neighbourhood of a cell consists of Voronoi cells with centres within a certain distance from the centre point of the specific cell. The main direction is calculated according to how many neighbours of the same type the cell has and the resultant between these distances. A resultant is calculated for the vectors of the cells of the first type and one for the other cells. The length of these resultants gives the concentration parameters for the mixed von Mises distribution governing the migration of the specific cell. The direction of the resultants will give the location parameter. This attraction will lead to a mixture of two von Mises distributions, one for within type interaction and one for the between type, for the direction of movement for a specific cell, ϕ. This corresponds to attraction (or rejection if appropriate) to either sister cells or cells from a different mother. The direction is specified by an angle around the origin of the system and given by the following expression

$$\phi = \gamma f(\alpha_1, \kappa_1) + (1 - \gamma) f(\alpha_2, \kappa_2). \tag{4.2}$$

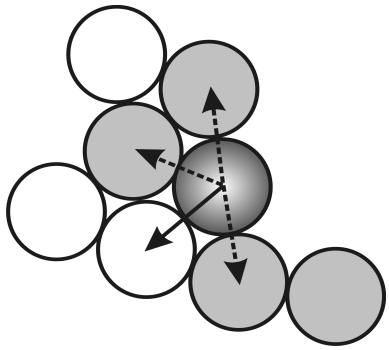


Figure 4.2: Cells with vector arrows pointing to them are in the local neighbourhood of the shadowed cell. One resultant is calculated for the vectors of the grey cells and one for the white cells. The length of these resultants gives the concentration parameters for the mixed von Mises distribution governing the migration of the shadowed cell. The direction of the resultants will give the location parameters.

Subscripts 1 and 2 denote neighbour cells of the same and opposing type . The parameter $\gamma \in [0, 1]$ tells how much of the movement is due to the cells of the same type and $1-\gamma$ how much is due to the cells of the different type. A larger γ leads to a stronger interaction with cells of the same type. Once the angular direction of movement is decided, the size of the movement is randomised by a uniform random variable on $[0, c]$ where c is a constant. The mean of the uniform distribution is set to a feasible value of the biological system. Parameters for the two von Mises distributions in Equation (4.2) are calculated by taking the positional vectors of each cell in the neighbourhood of the cell in question and calculating the resultant between these normalised by the number of cells in it, see Figure 4.2. A uniformly scattered neighbourhood of one cell type will lead to a shorter resultant and hence less of a directed attraction. More concentrated neighbourhood will instead give a longer resultant and more attraction to that direction. This can mimick a behaviour that some cells tend to stick to certain areas. The parameter κ will then be proportional to the length of the specific resultant and α gives its direction.

Division

A cell division is modelled to take place after an exponentially distributed time with intensity parameter $1/\lambda$. Cell area is preserved which means that the radius r_{x_i} of the daughter cells is calculated according to area preserva-

tion with equal proportion in both daughters. Compare with (Baddeley and Vedel Jensen, 2005).

The direction of the division plane of the mother cell depends on the local neighbourhood. If the mother is an inner cell that already has made her first cell fate decision (Gilbert, 2006), the daughters will also stay as inner cells and the division plane will be circular uniformly distributed. This feature will mimic the pluripotency of the inner cells. On the other hand, if the mother is an outer cell the division is directed. Outer cells tend to divide symmetrically with a higher probability, so that the two daughters also are outer cells. Symmetrical division becomes more common with time and the probability increases further into consecutive rounds of division. Hence, we will have two cases: either the mother cell divides symmetrically in the line of the outer cell sheet and forms two outer daughter cells with probability p_{out}, or the mother divides asymmetrically and forms an inner and an outer daughter with probability $1 - p_{\text{out}}$.

Initialisation

We are planning to have two options for the initial configuration of cells. The first alternative is to simulate the start from a single zygotic cell which undergoes several stages of mitosis. In this case cell divisions are organised as described previously. The other option is to start with a random configuration of a fixed number of cells (>1). The cells are randomly ordered into an aggregate where they all have spatial connection. When building this aggregate, a cell is placed in 2D area and the second cell is then placed uniformly at random at the boundary of the first circular cell. Next, we choose one of these already positioned cells and place the third cell in the same way as the second and avoiding overlapping with the existing cells. This procedure is repeated, until all cells have been randomly positioned in the area of simulation ensuring that all cells have at least one neighbour connection.

Simulations are initiated from the zygote and mitotic divisions continue until the 32-64 cell stages. Up to the 8 cell stage, cells are only allowed to divide, and after this stage each round of mitosis is followed by a period of dynamical position changes of cells according to the migration rules. At the 8-cell stage the outer cells are also striving (by the migration rules described earlier) towards a more even distribution around the inner cell mass to simulate the process of compaction.

Illustration

We have carried out a set of simulations from our current model to illustrate the process. First, a zygotic cell is placed into the area of simulation, Figure 4.3(a). At the second step this zygotic cell is divided after an exponentially distributed time into two daughters of equal size, see Figure 4.3(b). The division plane is set at random, however this can easily be incorporated with another distribution of the orientation of the first cleavage.

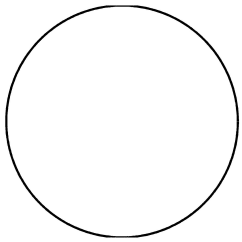
During the second cleavage we introduce the Voronoi cells. The circular radius is kept as a mark but the cells have shapes according to their Voronoi cell. Note that cross and plus sign markers in Figure 4.3(c) indicate descendants from each of the two daughters at the 2-cell stage. This simulation was finished at the 64-cell stage, seen in Figure 4.3(f).

Outlook

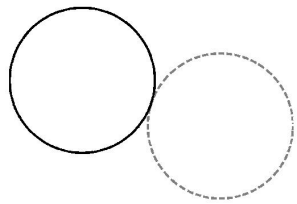
As mentioned earlier, the idea is to be able to compare different ideas of randomness for the early development of the mouse and therefore the model has to be flexible enough to cover several scenarios. We have sketched a model setup for a cellular game where cells can be allowed to move and to divide in a completely random way or be governed by more strict rules of orientations. This is a good start to be able to model several types of systems and then test them against each other to see which hypothesis is most probable. Our aim is to introduce a model that can be used to simulate observed features of a mammalian cell aggregate, and be applied in the future to deduce which important features may play decisive roles in governing the fate of mammalian regulative cells and the formation of the dorsal-ventral axis.

The model we have so far is only a beginning, and there are several issues that have to be considered. The first task will be to translate the two dimensional dynamics into a three dimensional cell aggregate. Even though von Mises distributions can be changed to Fisher distributions (the analogue on the sphere) and the dynamical Voronoi grid can easily be generalised into 3D, taking into account movement and dependencies will be computationally and graphically challenging.

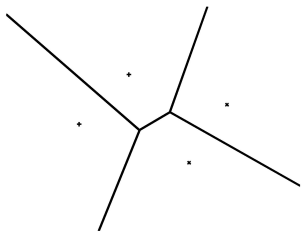
So far, only cell division and migration have been included in the model, but it should be easy to add alterations of gene expression by changing a mark of each cell. Gene expression changes can be simulated by either geometrical position or random allocation in the model.



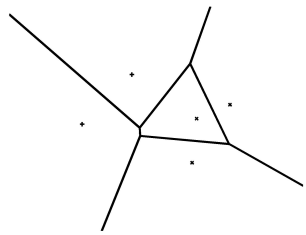
(a) The zygote, placed in the simulation area randomly.



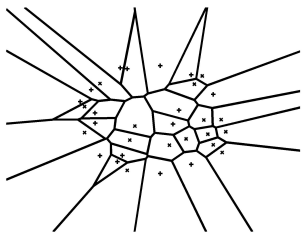
(b) The 2-cell stage, where orientation of first cleavage is random.



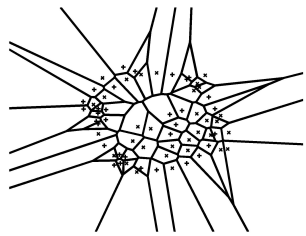
(c) The 4-cell stage, where the Voronoi grid is introduced. Each cell is still marked by its ancestral 2-cell stage mother and by a circular radius according to area preservation.



(d) Cells divide after an exponentially distributed time. The right most two cells are the new daughter cells after division.



(e) The 32-cell stage, open Voronoi cells are defined as outer cells.



(f) The compacted 64-cell stage. Note the compaction of the aggregate occurring due to the migration rules of the model.

Figure 4.3: A simulation of the dynamic cell game from a zygotic cell to the 64-cell stage. Descendants from the first two daughter cells (b) are marked with different markers.

Further, a physical model for the formation of the cavity would have to be introduced. Combining this with a model including changes of gene expression, sorting and apoptosis would give us a way to simulate the entire process of early mammalian development with different degrees of stochasticity. The model together with biological data may then help us to understand whether it is only a matter of chance which cells will become the fetus and which become the placenta, or whether there is some prepatterning of the cell positions and fates.

Bibliography

Ajduk, A., Ilozue, T., Windsor, S., Yu, Y., Seres, K., Bomphrey, R., Tom, B., Swann, K., Thomas, A., Graham, C., and Zernicka-Goetz, M. (2011). Rhythmic actomyosin-driven contractions induced by sperm entry predict mammalian embryo viability. *Nature Communications*, 2(417):1–10.

Baddeley, A. and Vedel Jensen, E. B. (2005). *Stereology for statisticians*. Chapman & Hall, Boca Raton, Florida.

Bischoff, M., Parfitt, D.-E. E., and Zernicka-Goetz, M. (2008). Formation of the embryonic-abembryonic axis of the mouse blastocyst: relationships between orientation of early cleavage divisions and pattern of symmetric/asymmetric divisions. *Development*, 135(5):953–962.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. Wiley, New York, 1st edition.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd edition.

Dietrich, J.-E. E. and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. *Development*, 134(23):4219–4231.

Diggle, P. J. (1999). *Time Series: A Biostatistical Introduction*. Clarendon Press, Boca Raton, Florida, 1st edition.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Fisher, N., Lewis, T., and Embleton, B. (1987). *Statistical Analysis of Sphercial Data*. Cambridge Univ. Press, New York.

Flandrin, F. (2000). *Time-frequency/time scale analysis*. Academic Press, San Diego, 1st edition.

Footer, M., Kerssemakers, J., Theriot, J., and M., D. (2007). Direct measurement of force generation by actin filament polymerization using an optical trap. *Proc. Natl. Acad. Sci. U. S. A.*, (104):2181–2186.

Forgacs, G. and Newman, S. (2005). *Biological Physics of the Developing Embryo.* Cambridge Univ. Press, Cambridge.

Gilbert, S. (2006). *Developmental Biology.* Sinauer Associates Inc., Sunderland, 8th edition.

Gonzalez, R. C. and Woods, R. E. (2006). *Digital Image Processing (3rd Edition).* Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Hiiragi, T. and Solter, D. (2004). First cleavage plane of the mouse is not predetermined but defined by the topology of the two apposing pronuclei. *Nature*, 430(15):360–364.

Honda, H., Motosugi, N., Nagai, T., Tanemura, M., and Hiiragi, T. (2008). Computer simulation of emerging asymmetry in the mouse blastocyst. *Development (Cambridge, England)*, 135(8):1407–1414.

Jedrusik, A., Parfitt, D.-E., Guo, G., Skamagki, M., Grabarek, J. B., Johnson, M. H., Robson, P., and Zernicka-Goetz, M. (2008). Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes & Development*, 22(19):2692–2706.

Karatzas, I. and Shreve, S. (1991). *Brownian motion and stochastic calculus.* Springer.

Manly, B. F. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology.* Chapman & Hall, Boca Raton, Florida, 2nd edition.

Meilhac, S. M., Adams, R. J., Morris, S. A., Danckaert, A., Le Garrec, J.-F. F., and Zernicka-Goetz, M. (2009). Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. *Developmental biology*, 331(2):210–221.

Morris, S., Teo, R., Li, H., Robson, P., Glover, D., and Zernicka-Goetz, M. (2010). Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc. Natl. Acad. Sci. U. S. A.*, (107):6364–6369.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams.* Probability and Statistics. Wiley, NYC, 2nd edition.

Øksendal, B. (2003). *Stochastic differential equations: an introduction with applications*. Springer.

Page, E. (1954). Continuous Inspection Scheme. *Biometrika*, (41):100–115.

Percival, D. and Walden, A. (2000). *Wavelet methods for time series analysis*. Cambridge University Press, Cambridge, 1st edition.

Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S., and Zernicka-Goetz, M. (2005). Four-cell stage mouse blastomeres have different developmental properties. *Development (Cambridge, England)*, 132(3):479–490.

Plusa, B., Hadjantonakis, A., Gray, D., Piotrowska-Nitsche, K., Jedrusik, A., Papaioannou, V., Glover, D., and Zernicka-Goetz, M. (2005). The first cleavage of the mouse zygote predicts the blastocyst axis. *Nature*, 434(17):391–395.

Schatten, G. and Donovan, P. (2004). Embryology: Plane talk. *Nature*, 430:301–302.

Snijder, B., Sacher, R., Rämö, P., Damm, E., Liberali, P., and Pelkmans, L. (2009). Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*, 461(7263):520–523.

Taylor, W. A. Change-Point Analysis: A Powerful New Tool For Detecting Changes. `http://www.variation.com/cpa/tech/changepoint.html`, 20110929.

Wolpert, L., Beddington, R., Brockes, J., Jessell, T., Lawrence, P., and Meyerowitz, E. (1998). *Principles of Development*. Oxford Univ. Press, New York, 1st edition.

Paper I

Paper II

Paper III

Paper IV

Reprinted from *Biochim. et Biophys. Acta (BBA) - Molecular Cell Research*, **2011**, *1813(2)*, 358-366, with permission from Elsevier.

Paper V