
# MATH554: Main Dissertation

*Mathematically modelling the topology of proliferating epithelial tissue.*

By Thomas Escalante

Supervisor: Dr. Bakhti Vasiev

Department of Mathematical Sciences

# Contents

Statement of Originality

This dissertation was written by me, in my own words, except for quotations from published and unpublished sources which are clearly indicated and acknowledged as such. I am conscious that the incorporation of material from other works or a paraphrase of such material without acknowledgement will be treated as plagiarism, according to the University Academic Integrity Policy. The source of any picture, map or other illustration is also indicated, as is the source, published or unpublished, of any material not resulting from my own research.

*Summary*

Epithelial tissues can be represented as unicellular layers which have distinctive topological characteristics (Patel, et al., 2009). The topological features of an epithelial tissue can be represented as a planar network where cell boarders form segments of straight lines (edges) and mathematicians can use polygons to represent individual cells when observed from the surface of a tissue. Hence the cross section of a cell in a monolayer epithelial sheet can be modelled mathematically as a k sided polygon (Xin and Karunarathna, 2018). Each cell has neighbours with other cells in their local neighbourhood, physically joined together by a junction. When tissues proliferate, cells in the tissue divide resulting in the generation of two daughter cells, each possessing a different number of edges. Therefore, the number of cells in a tissue double for every generation of tissue proliferation which means the distribution of k sided polygons in the tissue changes with time. When mathematicians study the topological distribution of a tissue, they measure the distribution of k sided cells with their neighbours, referred to as cell neighbour numbers (CNN) and we represent the distribution of various polygonal cells as a histogram (CEDH). Results from previous studies of proliferating epithelia show CEDHs from diverse organisms apparently stabilise to an equilibrium distribution at some point in time, where the majority are hexagonal. The values of k usually vary from four to nine in natural epithelia, where cells with less than four or more than nine edges are not commonly recorded. There are some instances where three and ten sided cells do appear, but their occurrences are negligible. Results from experiments with cells from various types of tissue drew the conclusion that the stabilised distribution which is reflected by the shape of the histograms seems to be universal across various organisms. This is an interesting concept to understand because due to the sheer diversity of living organisms it would be expected that, because of the mechanisms of cell division varying between species, the

distribution of daughter cells should naturally vary. However, this does not seem to be the case and is therefore an interesting phenomenon for scientific researchers to investigate.

Models constructed upon mathematical concepts have been developed to gain a better understanding of the universal shape of the CEDHs. All models have focused their assumptions based on the observations documented by Lewis, who investigated the topological behaviour of proliferating epithelia from the epidermis of cucumbers. Lewis focuses his investigations of regular hexagonal cells of cucumber epidermis with unit length. He investigates, using Geometrical calculations and actual cucumber epidermis, how divisions influence the size and area of the dividing mother cells in addition to observing the shapes of daughter cells. Two main conclusions become of interest. Firstly, he shows that most resting cells in tissue of cucumber are hexagonal, while calculating the distribution of other shapes. Secondly the division of various polygons results in the formation of two corresponding daughter cells that impact the number of edges of neighbouring cells. Dividing mother cells that are not six-sided seem to divide more often, ensuring the hexagonal cell distribution is maintained. Evidence shows that tissues maintain this hexagonal topology during tissue morphogenesis within their geometry of cell packing but the mechanical forces at the junctional level of cells results in an altered distribution of neighbouring cells after cell division, where a documented mathematical study of this is the vertex dynamics model (Farhadifar, 2007). The GPNP model (Gibson, et al., 2006) hypothesises that cellular proliferation is the only mechanism which impacts the polygonal cell distributions. Therefore, the model does not consider any mechanical influences from forces at the junctional level, nor does it consider any spatial correlations between cell edges. They also investigate the topology of tissues from a group of diverse organisms (fruit fly, frog, and marine invertebrate) to validate their mathematical model. The GPNP model reproduces experimental results well since the shape of the histograms seem to follow similar patterns from histograms produced in previous experiments. However, there was one discrepancy when comparing the results with experimental data: no four-sided cells were observed during proliferation. Nevertheless, the stable equilibrium distribution obtained from the mathematical model seemed to fit well with the actual CEDH obtained from their experiments of the different organisms. This led to Sandersius' attempts of improving the GPNP model by modifying the original assumptions. He was unsuccessful at making the GPNP model more biologically accurate. During the process of tissue growth, the number of edges within the population remains constant because, according to Lewis, tissues maintain their hexagonal cell shapes, but division of a k sided polygon can be affected by the division of neighbouring cells in the local neighbourhood, and therefore Gibson's' second documented investigations implemented a computational algorithm which randomly selects the orientation of the cleavage plane along which cell division takes place, using three separate methods; Uniform, Binomial and equal split.

There were a few aims of this project, were the first was to determine a probability density function which can be fitted to the cucumber distribution examined by (Lewis, 1928). Secondly, we aimed to extend our analytical model which simulates epithelial tissue growth, initially represented by master equations, further by considering an extended range of dividing mother cell edges. We then improved our model by incorporating the exponential relationship which exists between the number of cell edges and the probability of mother cell

division occurring into the matrices, which results in a more accurate replication of the stabilised topology of epithelial tissue. Finally, a cellular automata model is constructed which computationally incorporates the events of cell division according to a set of rules which reflects the various scenarios. The results of the cellular automata model confirmed the results we obtained from the continuous analytical models based upon master equations. There is some biological background of cells and their functions in addition to tissues and their topology, which gives a solid foundation to understanding the concept of topology. Differential equations are used to construct continuous models of different scenarios of cell division. We aim to replicate the evolving distribution of CEDH in epithelial tissue cells with time. The computational software package MATLAB has been used to produce bar charts and to solve the ordinary differential equations used to model the different scenarios of cellular division. Our results seem to reproduce the results obtained through experiments well only when the exponential relationship is incorporated into the models. For all models the generated CEDs were mostly inconsistent with experimental observations when cells were assumed to divide uniformly and binomially. However, our results for all models seemed to be consistent with experimental observations when cells were assumed to divide equally.

# Chapter 1. Background to dynamics of tissue topology

In this chapter we will consider key biological concepts that drive the development of tissue. We will then observe the experimental findings of Lewis who used cucumber epidermis to understand the topology of natural epithelia, described by their cell edge distribution. Ultimately it is this distribution of cucumber cells which is recognised by mathematicians as the universal distribution across epithelial tissues and is used in this investigation to compare our model distributions with. Then we will review some mathematical models that have incorporated biological factors, which aimed to simulate the development of epithelial tissue and to confirm the results of previous experiments upon natural epithelia.

## Section 1.1 Biological background

### Section 1.1.1 Epithelial Tissue

In multicellular organisms, the epithelia behave as a barrier between the body and the external environment which the organism is in direct contact with, protecting the body against various physical, chemical, and microbial threats (Vrana, et al., 2013). For example, the mucosal tissues have epithelia that defend against harmful pathogens by organisms breaking down the microbe via Phagocytosis. All epithelial cells form sheets of uniformly polarized cells (Rodriguez-Boulan and Macara, 2014) with the apical surface facing the external environment and the basal surface facing the basement membrane. In multicellular organisms they are spatially organized, where neighbouring cells are each bound via the direct physical interactions between cells (junctions), and this defines the limits between the apical and basal membrane (Le Bras and Le Borgne, 2014). The junctions are connected tightly in vertebrates, whereas the junctions are connected separately in invertebrates. The various interactions between cells of a tissue are what defines its topology. As previously mentioned already the cell junction between cells and their neighbours is important when considering the topology of a tissue since these junctions are what determines the geometry of cell packing. For example, plant cells associate with their neighbours not only by interactions between their cell walls but also with specialized junctions between their plasma membranes (Cooper, 2000). After cell division occurs their organisation is not impacted, due to there being no sliding between cells (Sahlin and Jönsson, 2010). Hence, the epidermal layer in plants gives scientists a convenient way of gaining information when investigating epithelial tissue topology.

Cells of certain tissue work together cooperatively to perform a specific role. There are four main types of tissue that form the organs of animals. These are muscular, epithelial, nervous, and connective. Epithelial tissue covers the surfaces of organs including the skin, the trachea, the reproductive tract, and the digestive tracts' inner lining. It has roles in absorbing water and nutrients, removing waste and secreting enzymes and hormones. Muscle cells are

responsible for the contraction of muscle that enable animals to move, which is highly useful if an animal needs to escape from a predator. In vertebrates there are three different types of muscle cells. Skeletal muscle is responsible for producing voluntary movements such as running, while cardiac and smooth muscle are responsible for producing involuntary movements, such as contraction of the heart. The epidermis of the skin is a stratified epithelial sheet where the cells are aligned in such a way that forms physical barriers on the body surface (Yokouchi, et al., 2016). Results from investigations relying upon microscopic methods show that cells in the monolayer epithelial sheets of different organisms resemble a polygonal cross section, varying in the number of edges and therefore the cell neighbour distributions can be visualised (figure 1.1).

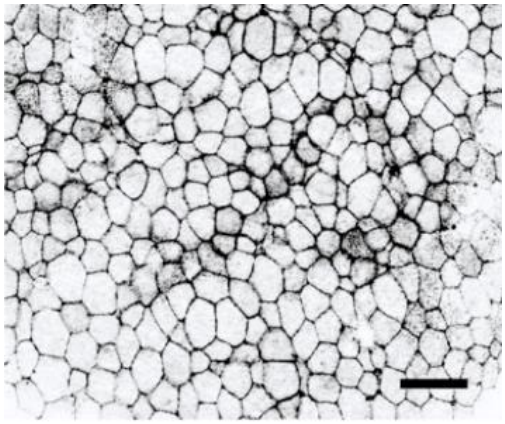


***Figure 1.1*** *Microscopic image illustrating the apical surface of a monolayer epithelial sheet taken from epiblast of early chick embryo (Sandersius, et al., 2011). Edges between neighbouring cells of the tissue are represented by black lines, where the number of edges of each cell clearly varies across the tissue.*

## Section 1.1.2 Cell division

During tissue development, monolayer epithelia physically expand due to increased cell size and divide which results in an increase in the number of cells, and therefore each generation of cell division results in topological changes. The complex hexagonal structure of epithelial tissues, according to Lewis, is maintained by this process of cellular division, but mechanical forces at the junctional level alters cell packing geometry and so every generation of cell division redistributes the cell neighbour numbers where (Farhadifar et al., 2007) has implemented a vertex dynamical model to investigate this mathematically.

To understand why the division of a cell results in a change of these topological distributions, we need to investigate the physical process of cell division. In unicellular organisms, reproduction is the mechanism of cell division, whereas in multicellular organisms it is the growth and maintenance of tissues and hence is vital to the development of multicellular organisms. Multicellular development is governed by cellular differentiation which is the process of regulating gene expression and cell signalling that is vital for communication (Sahlin and Jönsson, 2010). As tissues grow these cellular communications are affected directly

because of cellular proliferation and is responsible for maintaining a balance in the different numbers of polygons in epithelial cells which is vital for survival. Hence cell division underpins the evolution of CED. Tissue morphogenesis occurs when cell division causes mechanical properties of the cells to change. As epithelial tissue morphogenesis progresses the cells change their shape and divide, it can be extruded from the tissue and change the topology of neighbours (Lecuit and Lenne, 2007).

The cell cycle involves the complex process of producing two genetically identical daughter cells from a dividing mother cell. Mitosis involves the division of the nucleus of the dividing mother cell and the cytoplasm via cytokinesis, enabling 50% of the copied DNA of the parent cell to be distributed to each of the two daughter cells. It is Mitosis which underpins the cell cycle. When cells divide in the plane, epithelial cells typically round up, constrict in the middle and divide symmetrically to produce two genetically identical daughter cells which form new apical junctions with their neighbouring cells (Ragkousi and Gibson, 2014). Figure 1.2 illustrates the various stages involved in the cell cycle of a dividing cell, which summarises the main stages as:

- Stage 1 – Cells are arranged in a polygonal arrangement in the monolayer epithelium before proliferation occurs.
- Stage 2 – As the cells round up the nucleus of the dividing cell moves apically.
- Stage 3 – The orientation of the mitotic spindles along the cleavage plane of a mother cell is dependent on the topology of the neighbouring cells in the local vicinity.
- Stage 4 – Following the separation of chromosomes and cytokinesis, daughter nuclei move basally, and daughter cells form new junctions while elongating along the apicobasal axis.

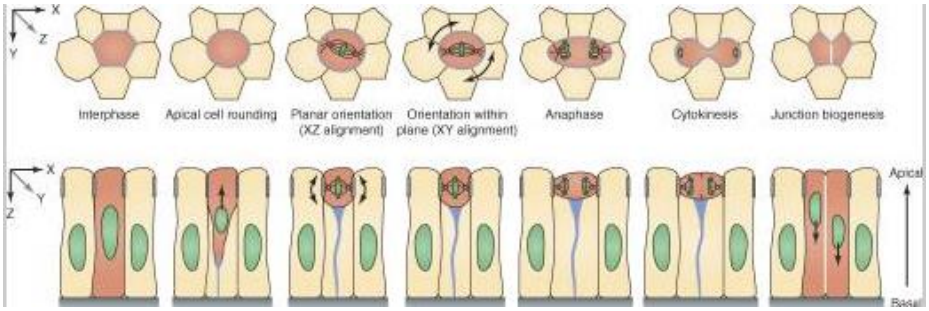



***Figure 1.2*** *Top row shows the apical cross section of a dividing mother cell. The bottom row shows the longitudinal view of the same dividing mother cell (Ragkousi and Gibson, 2014).*

## Section 1.1.3 Lewis' investigations

Lewis investigated the topological characteristics of cucumber epithelia using various mathematical measurements. Lewis aimed to explain the geometrical properties of cucumber cells when they undergo cell division, in addition to observing the geometrical properties of

the pairs of daughter cells when the mother cell divides, according to different ways the division plane is positioned. In the simple monolayer epithelial sheet of cucumber tissue, cells can be represented as polygons. Initially Lewis focuses on the geometrical properties of hexagons with edges of unit length and then later uses experimental methods to extract cells from the cucumber epidermis and aims to verify that his initial hypotheses, postulated using mathematical methods, were in fact valid in natural epithelia of cucumber tissue. Initially he constructs diagrams to illustrate cell division, where cells can divide with equal probability in either of three planes *a*, *b,* or *c* (figure 1.3, A). All the figures have been taken from the workings documented in (Lewis, 1926), (Lewis, 1928). The initial observations caused Lewis to postulate that if a single regular hexagon where to divide in the plane *a, b* or *c* it will produce a pair of daughter cells each possessing five edges and will cause two neighbouring cells to be heptagonal due to the addition of one edge (figure 1.3, B). If the six neighbouring cells should also divide in the planes *a, b* or *c* there would be altogether eight new hexagonal and six new pentagonal cells. This led Lewis to the first important hypothesis that divisions of the hexagons produce an equal number of pentagons and heptagons, were these return to six-sided cells as cell division continues. Lewis claims, from the early stages of tissue growth to the developed stages, the evolution of the fractions of hexagonal cells naturally balances the total loss and gain of edges in the population after each division of a mother cell, where the diagrams show that the average number of is therefore six (figure 1.3, C).

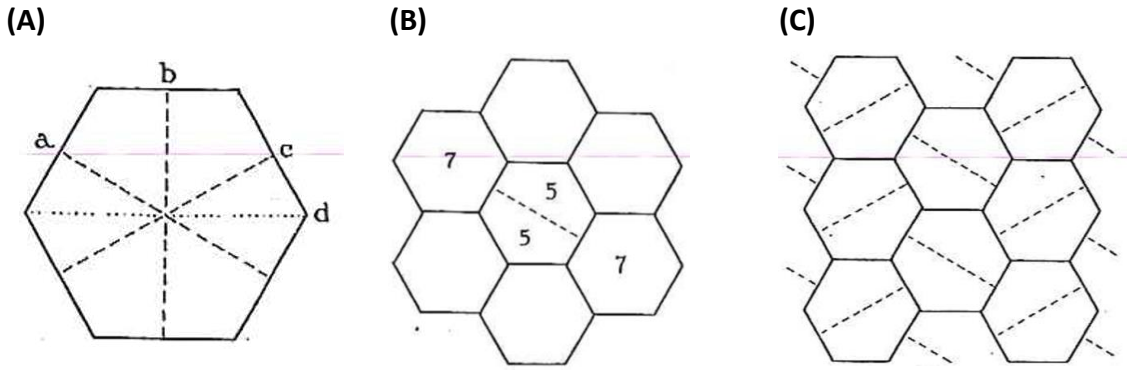



***Figure 1.3*** *Figure (A) shows three possible orientations of division, along three vertical planes a, b, and c. Figure (B) illustrates the division of a heptagonal mother cell producing equal numbers of heptagons and pentagons where neighbouring cells gain one extra edge. Figure (C) illustrates further cell divisions restoring the dividing cells back to their original hexagonal shape, illustrating the constant balancing of the removal and addition of edges during tissue proliferation.*

Lewis shows that if various cells were to divide within various division planes, there is balance in the number of edges of these dividing cells and neighbouring cells in the local neighbourhood. Biologically, this observation corresponds to the importance of tissues maintaining their cell sizes to maintain constant dynamical equilibrium of nutrients and gasses leaving and entering cells by diffusion and other complex processes of transportation. They then decided to use their initial calculations of the areas of various polygons of unit length to compare with the geometry of cells extracted from the epithelial tissue. Lewis extracted a tissue from a cucumber composed of tetrakaidekahedral cells. Their geometric comparisons

of these many cells led to the conclusion that the average number of edges is exactly six, verifying previous observations made by Wetzel with natural epithelia. Lewis then shows the geometrical arrangement of hexagonal cells as shown in figure 1.4, reflecting the pattern of cells in the pigment layer of the retina. They demonstrate that increasing the area of the central cell will result in the neighbouring hexagonal cells becoming irregular. The central hexagon then divides, the resulting daughter cells are two pentagons while simultaneously making two adjacent cells heptagonal. By using their calculated areas of polygons Lewis concludes that cells in the centre of the array increase their areas to maintain the topology of all the cells.

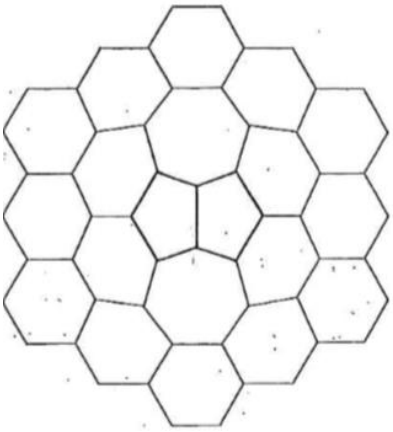


***Figure 1.4*** *Lewis' attempt to show geometrically the balance of edges brought about by the division of the central hexagon into two pentagons, with the formation of two adjacent neighbouring heptagons, in the pigment layer of the retina originally observed by Wetzel.*

Up until now a few topological observations can be stated from Lewis' geometrical investigations. If a mother cell divides along a division plane in the way represented in figure 1.4, this results in the removal of one cell from the population while this simultaneously results in the addition of two daughter cells. Therefore, there is a net addition of one cell to the population. Since the average cell is hexagonal, this means that the total number of edges in the network has increased by six. Lewis observes that the sum of the daughter cell edges will be four more than the number of edges of the dividing mother cell. This is regardless of the shape of the polygon which represents the dividing mother cell, and he shows this mathematically (figure 1.5). During cell division of a six-sided mother cell the two edges, $a$ and $b$, are removed from the population (figure 1.5, A). Hence there is a net loss of four edges from the population. Furthermore, due to the generation of two daughter cells the following five edges are added to the population: $a_1$, $a_2$, $b_1$, $b_2$ and $c$ (figure 1.5, B). However, each of the daughter cells possess these edges and since there are two daughter cells this number must be doubled and then added to the total population, so there is a net gain of ten edges to the population. Therefore, the net addition of the total number of edges is six. In conclusion, the division of a mother cell, on average, results in the addition of a hexagonal cell to the population.

**(A)** **(B)**


***Figure 1.5*** *(A) and (B) showing that whenever a mother cell divides the total number of new sides created is six.*

## Section 1.2 Mathematical Background

### Section 1.2.1 Topology of epithelial tissue

Topology is commonly understood by Mathematicians as the physical study of the properties that remain unchanged of an object in space if it was to be continuously deformed through motions, such as through twisting (Carlson, 2007). Topology has been used in the study of many physical structures in Biology, in particular the study of Morphogenesis. Morphogenesis investigates how biological structures evolve, where particular interest in experimental research for nearly over a century has been on understanding embryogenesis (Thom, 1969). Topology and Geometry are two important measurements to consider when understanding the mechanisms which stimulate the growth and development of epithelial tissue (Carter, et al., 2017). Mathematically, when we are considering the changes in the shape and size of a cell due to cellular division, we are studying the changes in the geometry of the cells. The Topology of a simple monolayer epithelium involves observing how these cells are connected within a planar network by studying the number of neighbouring cells for each polygonal cell in the tissue, corresponding to their defined geometries. Therefore, the topology of the epithelium is measuring how the distribution of cell neighbour numbers (CNN) changes over time due to cell division. When mother cells divide this directly affects the geometry of daughter cells in addition to the distribution of cell neighbour numbers within their local neighbourhood (Delannay and Le Caër, 1994). Changes in tissue topology refer to changes in the geometry of cells that are said to lose or gain edges brought about by the division of mother cells and the generation of two daughter cells (Yingzi, et al., 2012).

Epithelial tissues are commonly represented by unicellular layers and have quite distinctive topological features. Cells, as seen from the tissue surface can be represented by polygons whose number of edges varies between four and nine (Lewis, 1926). This number is equal to the total number of nearest cells that the cell is neighbours with. Investigations into the changes that occur in the topological physical structure of mother cells undergoing cellular

division have allowed advancements to be made in understanding complex cellular processes. In the previous section it was shown that epithelial cells from these cucumbers exhibit a distinctive cell–edge distribution, with the majority being hexagonal. Upon experiments on embryonic epithelia tissues, from a range of organisms, it has been shown that the statistics of CED are universal in tissues where cellular proliferation is the primary cell activity (Sandersius, et al., 2011). Cellular proliferation is the process by which cellular division events increases the number of cells in the tissue, and is a highly complex, tightly controlled process of the cell cycle (Yang and Krafts, 2014). Experiments upon animal and plant tissue confirm that the distributions of cellular polygons are significantly similar when different epithelia are investigated (Patel, et al., 2009).

When a mother cell divides into two identical daughter cells via a division line it is apparent that the number of sides of its two neighbours increases by one. It was shown in the previous section that the selection of the two affected neighbours comes from the orientation of the cleavage plane which seems to be influenced by the topology of cells in the local neighbourhood. Hence the division line crosses two edges of the dividing mother cell and so over time the number of edges of these cells, in addition to their fractions in the tissue evolves. Results recorded in the literature indicates that there is a point in time where the distribution of CED stabilises, indicating that further proliferation will no longer impact the distribution of CNN. A couple of mathematical models that aimed to describe the epithelial topology will now be discussed.

### Section 1.2.2 GPNP Model, Sandersius' modifications, and cleavage patterns

Gibson uses a continuous-time, non-spatial Markov model to demonstrate that the distribution of polygonal cell types in epithelia is a result of cell proliferation and aims to show that the different cell-edge distributions observed in epidermal tissue of cucumber in Lewis' experiments is universal across different organisms. They show that cell neighbour numbers were highest for hexagonal cells and successfully replicated the CEDs observed in such experiments. Proliferating epithelia rarely show the honeycomb pattern in their cell arrangement, often forming irregular polygon arrays because of cell division. The GPNP mathematical model was constructed, based on the following assumptions, which seem to be consistent with experimental evidence:

- **Assumption 1** - Cells have a polygonal shape with a minimum of *4* sides.
- **Assumption 2** – Random orientation of division plane for each cell in tissue.
- **Assumption 3** – Synchronous cell division in discrete generations.
- **Assumption 4** – There are no spatial correlations between the sides of cells.

However though, a few of the other assumptions do not seem to be consistent in biological systems, so the model is not completely successful in explaining the patterns of cell-edge distributions in proliferating epithelia. Cells with four edges do not appear in their observations which is not consistent with the polygon distribution observed in proliferating epithelia of diverse organisms. They predict, using a markov chain model that a stable

equilibrium distribution of polygons should emerge in proliferating epithelia, irrespective of the initial conditions. They use a discrete Markov chain to capture the stochastic nature of cell division (see MATH522). Sandersius and his co-workers investigated the CED generated from proliferating chick embryo which forms a monolayer epithelial like sheet during their early stages of development. They improve the non-spatial markov chain model proposed by Gibson by making the assumptions more biologically accurate. They do this by constructing three models referred as to as Neuman models (see MATH552).

The following histogram produced using MATLAB represents the polygonal distribution of cells generated from the GPNP model and the three separate Neuman models along with Lewis' experimental observations:

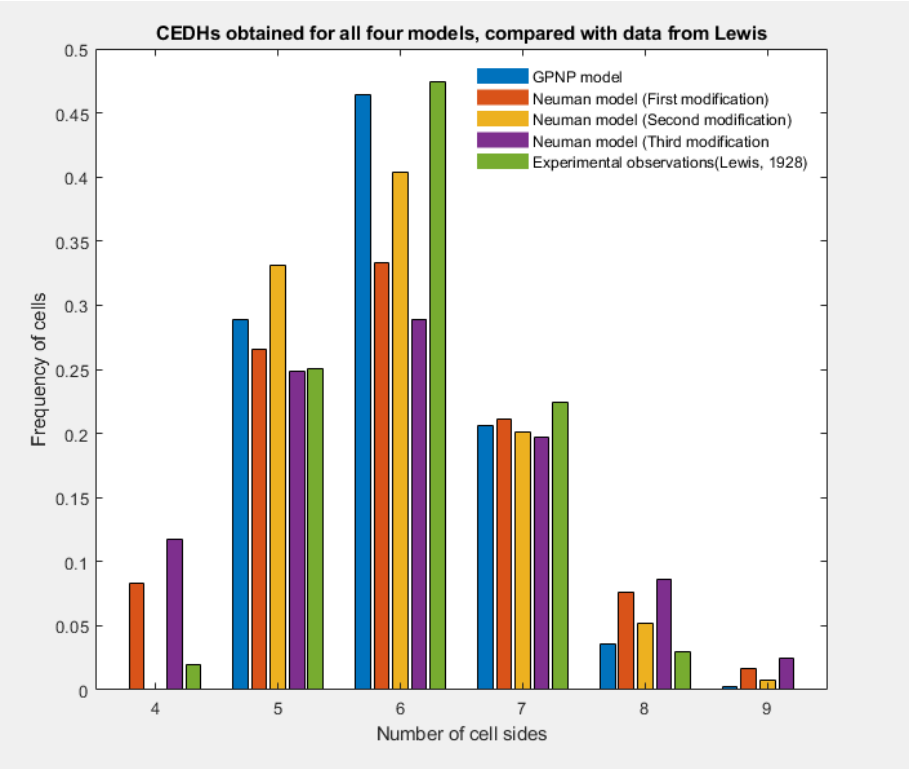



***Figure 1.6*** *Comparison of CEDs generated from mathematical models, compared with cucumber distribution.*

The GPNP model clearly replicates the experimental observations well, with the only inconsistency being the model does not detect the presence of 4-sided cells in the proliferating tissue. Sandersius' improvements to overcome this issue unfortunately failed at reproducing the statistics of the universal shape. The third model seems to overpredict 4, 8 and 9 sided cells the most. However, this model performed the most effectively at predicting the frequency of 5 sided cells. The second model does not predict the presence of 4 sided cells, the same behaviour of the GPNP model. The first model is the least effective at predicting the topology of hexagonal cells.

Gibson's' experiments of modelling cleavage patterns focus on observing the topological changes of proliferating epithelia by implementing a 2-D computational model, aiming to explain how various methods of cell division impacts the distribution of epithelial cell shapes

between different organisms. Gibson represents the monolayer epithelia as a planar network (figure 2.1) and investigates how the orientation of cleavage planes are chosen during various generations, under the hypothesis that the number of neighbouring cells and the shapes of proliferating mother cells both influence these orientations. These were first observed in Lewis' experiments of cucumber epidermis. A few conclusions made from Lewis' experiments and the GPNP model led to the natural decision for Gibson and his co-workers to try and explain how these orientations are chosen and the reason for such orientations. Gibson is interested in the number of neighbouring cells per cell in the network, which can be visually observed by the number of edges shared by cells. As already discussed in Lewis' experiments, every time a mother cell divides, the topology of this planar network is changed by the addition of new vertices, edges, and faces. The "Cleavage plane regulation model" (CMP) is implemented to describe how a dividing cell selects the two edges that the cleavage plane bisects, which occurs uniformly at random (Figure 1.7). Based upon their previous experimental observation they define the following assumptions:

- **Assumption 1** – Cell division is the only event that influences the network. The rearrangements of junctions due to cellular mechanical processes between neighbouring cells are not considered.
- **Assumption 2** – For every division cycle implemented by the CPM, all cells in the network undergo division once.
- **Assumption 3** – Two vertices and one edge are added to the network to reflect the generation of two identical daughter cells for each division of a mother cell.
- **Assumption 4** – Three sided cells are not formed due to their negligible fractions observed in the literature, so the cleavage plane crosses any two non-adjacent cells of dividing mother cell.

The epithelial topology is influenced by CPM in two different ways. The first is measuring how much of an impact the neighbouring cells have in selecting the cleavage plane orientation, computationally denoted as "Edge 1". According to Lewis, cells with less than six edges have a greater influence on the orientation of the division plane to restore their hexagonal cell geometry and therefore neighbouring cells with a smaller number of edges tend to have a greater influence on mechanical cell processes during tissue morphogenesis (Farhadifar, et al., 2007). The second component of CPM is the symmetrical pattern of distributed neighbours to the other two daughter cells, denoted as "Edge 2". Gibson implements cell division through the CPM algorithm by generating a new edge connecting "Edge 1" and "Edge 2". Gibson introduced four different methods the "Edge 1" could be chosen randomly using the algorithm, depicted in the top diagram of figure 1.8:

*Random*: While not considering the local topology, the alignment of the mitotic spindle proceeds where "Edge 1" is chosen uniformly at random from all cell edges in the population,

reflecting the model that assumes neighbouring cells do not influence the choice of cleavage plane orientation.

*Smallest neighbour*: The mitotic spindle aligns in such a position that division of a mother cell during the next cycle will add an extra edge to the neighbouring cell ("Edge 1") with the smallest number of edges.

*Largest neighbour*: This is the same as the previous scenario, with the only difference being the "Edge 1" is the neighbouring cell with the largest number of edges.

*Orthogonal*: Cleavage plane is rotated by 90 degrees every time the mother cell undergoes a division cycle. "Edge 1" is selected to the neighbouring edge from the previous cycle of division.

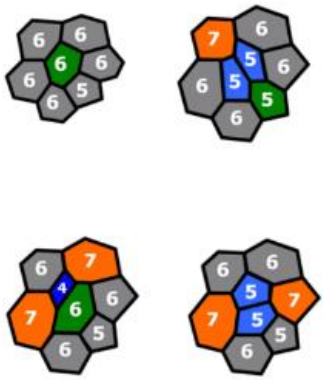


***Figure 1.7*** *CPM for a hexagonal cell illustrating topological changes due to cycles of division along with the generated pairs of daughter cells. Its two adjacent neighbouring cells gain an extra edge (orange cells), (Gibson, et al., 2009).*

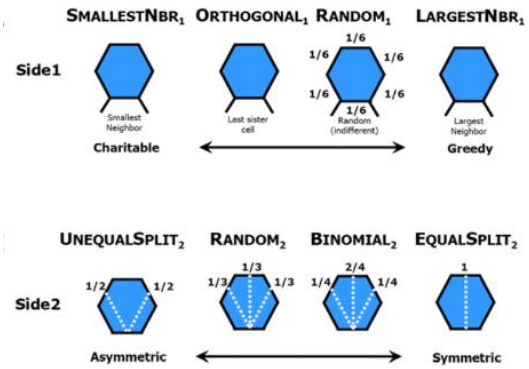



***Figure 1.8*** *Top diagram illustrating the four methods of selection of the first edge using CPM. The selection of the second edge is shown in bottom diagram. Possible cleavage planes are shown as dashed white lines while the fractions correspond to the probabilities of choosing a cleavage plane (Gibson, et al., 2009).*

Gibson now considers "Edge 2" of the cleavage plane randomly selected, which impacts the symmetrical distribution of neighbours to the daughter cells generated for each cycle of cell division. They introduced four different methods for this random selected, depicted in the bottom diagram of figure 1.8:

*Strategy 1 – Random*: For non-adjacent cells to the first edge, the second edge is chosen uniformly at random to avoid the generation of any three-sided daughter cells, where both symmetric and asymmetric cleavage planes are selected with the same probability.

*Strategy 2 – Equal Split*: "Edge 2" is selected in such a manner that the generated daughter cells are as geometrically equal as possible. For an $i$ - sided mother cell that divides, both daughter cells will therefore be $i$ - sided.

*Strategy 3 – Binomially orientated*: "Edge 2" is selected in such a manner that follows the binomial expansion. Symmetric daughter cells appear more often than asymmetric daughter cells.

*Strategy 3 – Unequal split*: This is the same as equal split, but the generated daughter cells will be as unequal as possible. For an $i$ - sided mother cell that divides, one of the daughter cells will be $i$ – sided while the other daughter cell will be ($i$-1) sided.

The stabilised equilibrium distribution of polygonal cells using their computational models are shown below in figure 1.9.

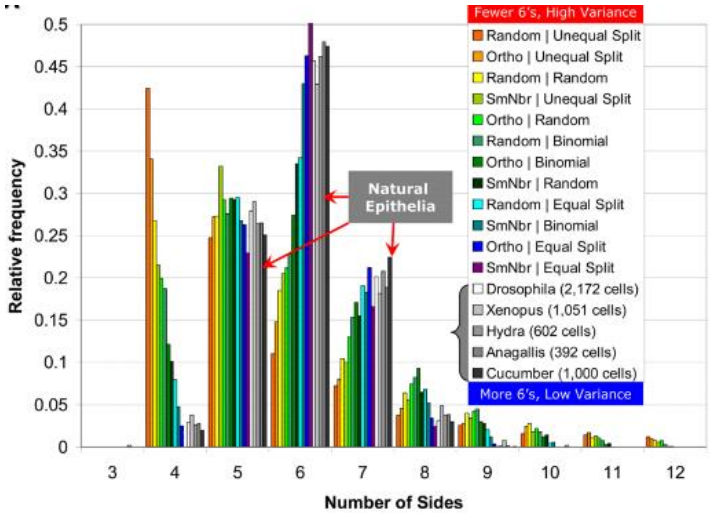



***Figure 1.9*** *Topological distributions for all simulated CPMs, compared with experimental recordings of natural epithelia (Gibson, et al., 2009).*

According to literature, every CPM should stabilise to an average topology of six sided cells due to the topology of cells in the local neighbourhood. By looking at figure 1.9 there seems to be differences in the ranges of the topology of each distribution generated from the three scenarios tested. Gibson shows that the corresponding level of symmetry of each scenario of cell division, which is defined by the symmetry of the orientation of the cleavage plane, is correlated with the variance of the distribution generated. The distribution for the equal split scenario agrees well with universal shape of natural epithelia. Due to the symmetrical nature of the equal split scenario, the range in the topological distribution is reduced. This is to be expected since the generated pair of daughter cells are equal in size and so the number of generated hexagons should be large in comparison to the number of hexagons generated from the other two models. The variance of polygonal cell distributions generated from this scenario reflects the variance of the polygonal distribution in natural epithelia and therefore is a strong model to describe the statistics of epithelial cell topology. Gibson shows that for all "Edge 1" methods tested, the percentage of hexagons in the population increases with increasingly symmetric "Edge 2" CPM. CPMs generated from the binomial and uniform

models do not stabilise to hexagonal cell shape. This is because of the unsymmetric properties of these two methods of division.

They compared the results on the simulated CPMs with the stabilised distribution from natural proliferating epithelia in biologically diverse organisms included plant tissue. They conclude that the topology of natural epithelia displays low variance in the topology of polygonal cells, implying that various organisms may have distinct mechanisms to supress shape variability during proliferation and to maintain the hexagonal cell topology. To conclude this section, proliferation in different organisms clearly have different mechanisms to maintain the topology of their tissues and is now better understood, which will enable us to implement this topological framework as the basis for our mathematical model.

# Chapter 2. Statistical Analysis of epithelial tissue topology

In the following chapter we aim to determine if there is a probability density function that can describe the universal statistics of the topology observed in natural proliferating tissue by fitting three distributions to the raw data: Poisson, normal and log normal. We aim to determine the closest fit, which can enable us to describe the topology of a proliferating tissue more accurately over a defined period.

## Section 2.1 Introduction

Cells, when represented as polygons, from the epidermis form physical connections between neighbouring cells in their neighbourhood, where the number of k sided cells connecting with neighbours varies during the process of tissue proliferation. The distribution of polygonal cell neighbour numbers (CNN) can be plotted as bar charts, where their shapes stabilise to a universal pattern between diverse organisms. The evolved distribution of CNN is recognised as one of the most used methods to describe tissue topology (Sandersius, et al., 2011). The earliest topological recordings were from Lewis in the 1920s where he investigated the effects of cell division on the shapes and sizes of cucumber epidermal cells which were represented as regular hexagons (figure 1.6).

We initially present the raw data obtained from proliferating natural epithelial tissues in *Drosophila*, *Xenopus* and *Hydra* (Gibson, et al., 2006) along with the raw data obtained from cucumber epidermis (Lewis, 1928) and display their topological shape in the form of bar charts. We then draw visual conclusions from looking at the recorded patterns of topology for these tissues. Following on from this we use Chi-squared statistical tests to measure the deviation of each of the experimental distributions from the calculated average distribution. To conclude we fit three probability distribution functions to the raw data and present these using curves.

| **Number of Cell edges** (i) | **Observed Drosophila** $(O_D i)$ | **Observed Hydra** $(O_H i)$ | **Observed Xenopus** $(O_X i)$ | **Observed Cucumber** $(O_L i)$ | **Total observed number of cells per number of edges** |
|---|---|---|---|---|---|
| 3 | 0 | 0 | 2 | 0 | 2 |
| 4 | 64 | 16 | 40 | 20 | 140 |
| 5 | 606 | 159 | 305 | 251 | 1321 |
| 6 | 993 | 278 | 451 | 474 | 2196 |
| 7 | 437 | 125 | 191 | 224 | 977 |
| 8 | 69 | 23 | 52 | 30 | 174 |
| 9 | 3 | 1 | 8 | 1 | 13 |
| 10 | 0 | 0 | 2 | 0 | 2 |
| Total Observations | 2172 | 602 | 1051 | 1000 | Sum of Total observations = 4825 |

***Table 1*** *Table displaying observed cell topology obtained within experiments of natural epithelia (Gibson, et al., 2006) and cucumber distribution (Lewis, 1928) along with the total number of cells per number of cell edges.*

The following table displays fractions of the various experimental distributions which allow us to compare these with the fractions of the cucumber distributions. The table also includes a column displaying the total distributions for each number of cell edges.

| **Number of Cell edges** (i) | **Drosophila distribution** $(F_D i)$ | **Hydra distribution** $(F_H i)$ | **Xenopus distribution** $(F_X i)$ | **Cucumber Distribution** $(F_L i)$ | **Total observed number of cells per number of edges** |
|---|---|---|---|---|---|
| 3 | 0 | 0 | 0.0019 | 0 | 0.000414 |
| 4 | 0.0295 | 0.0266 | 0.0381 | 0.02 | 0.029016 |
| 5 | 0.2790 | 0.2641 | 0.2902 | 0.251 | 0.273782 |
| 6 | 0.4572 | 0.4618 | 0.4291 | 0.474 | 0.455129 |
| 7 | 0.2012 | 0.2076 | 0.1817 | 0.224 | 0.202487 |
| 8 | 0.0318 | 0.0382 | 0.0495 | 0.03 | 0.036062 |
| 9 | 0.0014 | 0.0017 | 0.0076 | 0.001 | 0.002694 |
| 10 | 0 | 0 | 0.0019 | 0 | 0.000414 |
| Sum | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |

***Table 2*** *Table displaying fractions of cell topology obtained within experiments of natural epithelia (Gibson, et al., 2006) and cucumber distribution (Lewis, 1928) along with the total observed number of cells per number of cell edges.*

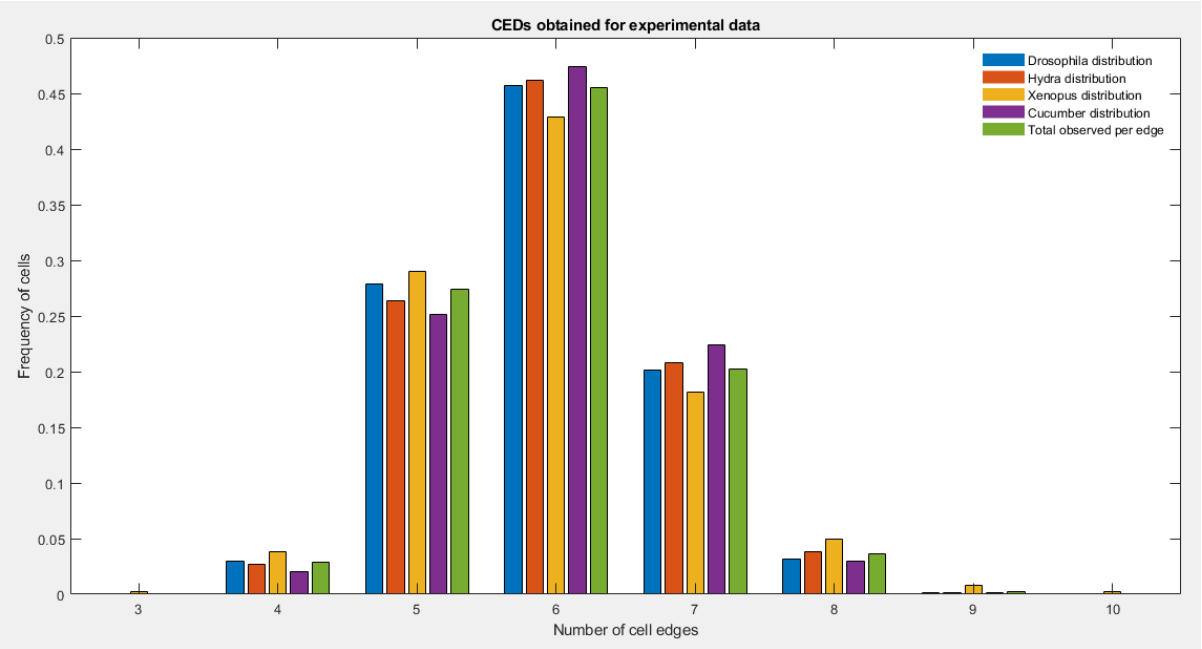



***Figure 2.1*** *Bar chart illustrating CEDs obtained from various experimental observations (Gibson, et al., 2006) and (Lewis, 1928), compared with the total observations per number of cell edges.*

The bar chart presented in figure 2.1 which illustrates the cell-edge distributions of proliferating epithelia in *Hydra*, *Xenopus*, *Drosophila*, and cucumber are remarkably similar. The consistent nature of all experimental distributions is the tissues maintain a hexagonal cell

topology where cells with six nearest neighbours appear the most. A significant fraction of cells with five and seven nearest neighbours were also detected. However, the result for Xenopus indicates that a small fraction of cells with three edges and ten edges appear which is inconsistent with the universal shape, where no three or ten sided cells appear to exist in recorded experiments. The bars for cucumber topology and the bars for the total topology of Gibson's' experiments are similar in height where their overall shape of the average cell-edge distribution is right skewed. This indicates that the average evolved distributions of all proliferating epithelia stabilise to an average topology consistent with the universal shape of tissue topology.

## Section 2.2 Comparing experimental data with each other

The Chi-Squared test measures how well a particular model can be fitted to actual observed data. The test statistic compares the size of any discrepancies between the expected results and the actual results, given the size of the sample and the number of variables in the relationship. We calculate degrees of freedom to determine if a certain null hypothesis ($H_0$) can be rejected based on the total number of variables and samples there are within an experiment. The negation of the assumed null hypothesis is known as the alternative hypothesis ($H_1$). As with any test statistic, the larger the sample the more reliable the results. For the $X^2$ test statistic:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

We will perform a chi-squared test to determine which of the fitted experimental distributions are the most like the cucumber distribution:

- $O_i$ is the observed data from Lewis.
- $E_i$ is the expected observations from the fitted experimental distributions.

The test statistic follows approximately a chi-square distribution with $(k - c)$ degrees of freedom where $k$ is the number of variables, and $c$ is the number of parameters:

$$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$$

- Test statistic: $(8 - c) = (8 - 1) = 7$

Therefore, the null hypothesis that the data are from a population described by experimental distribution is accepted if the following holds:

$$\text{Accept } H_0 \text{ if } X^2 > {X^2}_{1-\alpha,k-c}$$

Where ${X^2}_{1-\alpha,k-c}$ is the chi-squared critical value with $(k - c)$ degrees of freedom and significance level α, where we will use a significance level of $5\%$: $\alpha = 0.05$.

| **Number of Cell edges** (i) | **Observed Drosophila** $(O_D i)$ | **Expected Drosophila** $(E_D i)$ | **Observed Hydra** $(O_H i)$ | **Expected Hydra** $(E_H i)$ | **Observed Xenopus** $(O_X i)$ | **Expected Xenopus** $(E_X i)$ |
|---|---|---|---|---|---|---|
| 3 | 0 | 0 | 0 | 0 | 2 | 0 |
| 4 | 64 | 43.44 | 16 | 12.04 | 40 | 21.02 |
| 5 | 606 | 545.172 | 159 | 151.102 | 305 | 263.801 |
| 6 | 993 | 1029.528 | 278 | 285.348 | 451 | 498.174 |
| 7 | 437 | 486.528 | 125 | 134.848 | 191 | 235.424 |
| 8 | 69 | 65.16 | 23 | 18.06 | 52 | 31.53 |
| 9 | 3 | 2.172 | 1 | 0.602 | 8 | 1.051 |
| 10 | 0 | 0 | 0 | 0 | 2 | 0 |
| Total Observations | 2172 | 2172 | 602 | 602 | 1051 | 1051 |

***Table 3*** *Table displaying the observed and expected data for all experimental observations which can be used to perform statistical tests with the aim of concluding which experimental distribution fits Lewis' observations the best.*

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

- Null hypothesis $H_0$: Lewis's distributions are correct at fitting the experimental distributions.
- Alternative hypothesis $H_1$: Models do not fit the observed data.
- Significance level of $5\%$: $\alpha = 0.05$
- Significance level of $1\%$: $\alpha = 0.01$
- Degrees of freedom: $(k - 1) = 7$
- Critical value under null hypothesis: ${X_5}^2 = 14.067$
- Critical value under null hypothesis: ${X_1}^2 = 18.475$

Test statistic for Drosophila: $x^2 = \frac{(64-43.44)^2}{43.44} + \frac{(606-545.172)^2}{545.172} + \frac{(993-1029.528)^2}{1029.528} + \frac{(437-486.528)^2}{486.528} + \frac{(69-65.16)^2}{65.16} + \frac{(3-2.172)^2}{2.172} = 23.3978\ (4.dp)$

Test statistic for Hydra: $x^2 = \frac{(16-12.04)^2}{12.04} + \frac{(159-151.102)^2}{151.102} + \frac{(278-285.348)^2}{285.348} + \frac{(125-134.848)^2}{134.848} + \frac{(23-18.06)^2}{18.06} + \frac{(1-0.602)^2}{0.602} = 4.2381\ (4.dp)$

Test statistic for Xenopus: $x^2 = \frac{(40-21.02)^2}{21.02} + \frac{(305-263.801)^2}{263.801} + \frac{(451-498.174)^2}{498.174} + \frac{(191-235.424)^2}{235.424} + \frac{(52-31.53)^2}{31.53} = 49.7116\ (4.dp)$

Hence for a random variable $X^2$ which follows the ${X_5}^2$ distribution, then $P(X^2 > 14.067) = 0.05$. If the experimental distributions do fit Lewis' observations it can be shown that our test statistic $X^2$ will follow the ${X_5}^2$ distribution, so the $X^2$ value should unlikely be greater than 14.067 since there is only a 5% chance of this occurring. If we have a value of $X^2$ greater than 14.067 then we would conclude that the experimental distributions obtained by Gibson is statistically unlikely to produce such a large value, so there would be evidence at the 5% level that these distributions do not fit Lewis' topological distribution well. In fact, the $X^2$ value for our Hydra data is 4.2381 < 14.067. Hence there is no evidence at the 5% level and even at the 1% level to reject the null hypothesis and we can be $99\%$ confident that the Hydra distribution can be used to fit the CED obtained for 4-9 sided cells from Lewis. Since the calculated test statistics for both Drosophila and Xenopus are greater than the critical value, we reject the null hypothesis at the 5% and 1% level of significance and accept the alternative hypothesis with 99% confidence. The Hydra distribution can therefore be treated as the same distribution for cucumber epidermis.

To further our analysis into these distributions we will compare the distributions of Drosophila and Xenopus to determine if these can be treated as the same distributions, where the necessary data has been presented in the following table:

| Number of Cell edges (i) | Observed Drosophila $(O_D i)$ | Expected Drosophila $(E_D i)$ |
|---|---|---|
| 3 | 0 | 4.1268 |
| 4 | 64 | 82.7532 |
| 5 | 606 | 630.3144 |
| 6 | 993 | 932.0052 |
| 7 | 437 | 394.6524 |
| 8 | 69 | 107.514 |
| 9 | 3 | 16.5072 |
| 10 | 0 | 4.1268 |
| Total Observations | 2172 | 2172 |

***Table 4*** *Table displaying the observed and expected data for Drosophila observation with the aim of concluding whether the drosophila and Xenopus distributions are statistically different.*

- Null hypothesis $H_0$: Distributions for Xenopus and Drosophila are statistically similar and so can be treated as the same distribution.
- Alternative hypothesis $H_1$: Distributions are not statistically similar.
- Significance level of $5\%$: $\alpha = 0.05$
- Significance level of $1\%$: $\alpha = 0.01$
- Degrees of freedom: $(k-1) = 7$
- Critical value under null hypothesis: ${X_5}^2 = 14.067$
- Critical value under null hypothesis: ${X_1}^2 = 18.475$

$$x^2 = \frac{(64-82.7532)^2}{82.7532} + \frac{(606-630.3144)^2}{630.3144} + \frac{(993-932.0052)^2}{932.0052} + \frac{(437-394.6524)^2}{394.6524} + \frac{(69-107.514)^2}{107.514} + \frac{(3-16.5072)^2}{16.5072} = 38.5726\ (4.dp)$$

Since the calculated test statistic is greater than the critical value, we reject the null hypothesis at the 5% and 1% level of significance and accept the alternative hypothesis with 99% confidence. Hence the Drosophila distribution is not statistically like the Xenopus distribution and therefore cannot be treated as the same distribution.

To conclude this section, we have found that all three distributions display a CED shape which is like cucumber epidermis. By performing statistical tests, we have determined that the distribution of Hydra displays the CED shape of cucumber epidermis the most precisely, whereas Drosophila and Xenopus do not display the CED shape of cucumber epidermis enough to conclude they are good at reflecting this shape. We also see that the test statistic for Xenopus is more than twice as large as the test statistic for Drosophila. Hence the Xenopus distribution seems to reflect the shape of cucumber distribution significantly worse than Drosophila and as a final step we decided to test this difference. It was found that the distributions were statistically different and hence Xenopus is statistically poor while Hydra is statistically good at replicating the shape of the cucumber distribution.

## Section 2.3 Comparing statistical distributions with experimental data.

The aim of this section is to determine if tissue topology can be described using a statistical distribution. We will compare three statistical distributions; Poisson, Normal and Log-Normal which have been fitted to the data collected from Lewis' investigation into the topology of cucumber epidermis (Lewis, 1928). Microsoft excel has been used to generate curves which enables visual comparisons of the distributions of the experimental data with the three statistical distributions. The method of least squares is described initially which measures how accurately the statistical distributions fit along the curve generated from the cucumber distribution. Then we use the chi squared test statistic to determine if the shape of the best of these three fitted distributions is statistically like the distribution of cucumber.

Consider $x$ to be the number of cell edges of the dividing mother cell, which will appear on the x-axis. Then the frequency of events will be described by the y-axis and λ is the average number of events per interval (rate parameter). We will allow the values of $x$ to be 0, 1, 2, 3, 4 and 5. This corresponds to cells with several edges which range from 4 to 9. We will allow the mean to be 2; λ = 2. Microsoft excel solver is used to display two curves. The first will be the experimental observations generated from Lewis (cucumber distribution), and the second will be the data after it has been fitted with each of the three statistical distributions. We will apply the method of least squares to generate a nonlinear curve which will reflect the line of best fit for our data. It is assumed that the curve has the minimal sum of errors from a set of data (Molugaram & Rao, 2017). If we have the following data points: ($X_1$, $Y_1$), ($X_2$, $Y_2$), ($X_3$, $Y_3$), …, ($X_N$, $Y_N$) where $X$ is the independent variable and $Y$ is the dependent variable, we can allow the nonlinear curve generated from the method of least squares to be $f(x)$. The error from each individual data point in the given data set can be represented as $e_i$ and is calculated by the following:

- $e_1 = Y_1 - f(X_1)$
- $e_2 = Y_2 - f(X_2)$
- $e_3 = Y_3 - f(X_3)$… up to the $N$th data point

Hence the curve that is generated reflects the minimised sum of the squared errors:

$$\sum_{i=1}^{N}(e_i)^2 = \sum_{i=1}^{N}[Y_i - f(X_i)]^2$$

*Poisson Distribution*

The first probability function that will be considered is the curve generated from the Poisson distribution. This discrete distribution is used to model the number of events occurring within a given time, given the average number of times the event occurs over that defined period. If $X$ is a discrete random variable which represents the number of events $x$ observed over a given time for any non-negative integer $x$, the formula for this mass function is given as:

$$p(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \; \forall \, x \in \mathbb{Z} \geq 0 \tag{1.3}$$

A few parameters have been included in the formula, which can be defined as:

- λ is the shape parameter which indicates the average number of events in the given time interval. It is the expected value of $x$: $E[X] = \lambda$. This is also equal to the variance.
- $x$ is the number of occurrences.
- $x!$ is the factorial of $x$.
- $e$ is Euler's number (2.71828…)

For each of the three statistical distributions we display the observed polygonal fractions from the cucumber epidermis along with the fitted distributions, including a column displaying the sum of squared errors (tables 5, 6 and 7). The fitted curves are presented in figures 2.2, 2.3 and 2.4.

| X | Fractions for cucumber epidermis ($Yobs$) | Fitted Poisson distribution ($Ycalc$) | Error squared $(Yobs - Ycalc)^2$ |
|---|---|---|---|
| 4 | 0.02 | 0.120972 | 0.010195 |
| 5 | 0.251 | 0.255516 | 0.000020 |
| 6 | 0.474 | 0.269851 | 0.041677 |
| 7 | 0.224 | 0.189993 | 0.001156 |
| 8 | 0.03 | 0.100326 | 0.004946 |
| 9 | 0.001 | 0.042381 | 0.004946 |
| Sum: | 1.0 | 0.979 | 0.063 |

***Table 5*** *Table showing the fractions of cell shapes obtained from Lewis, where the Poisson distribution has been fitted to the data.*

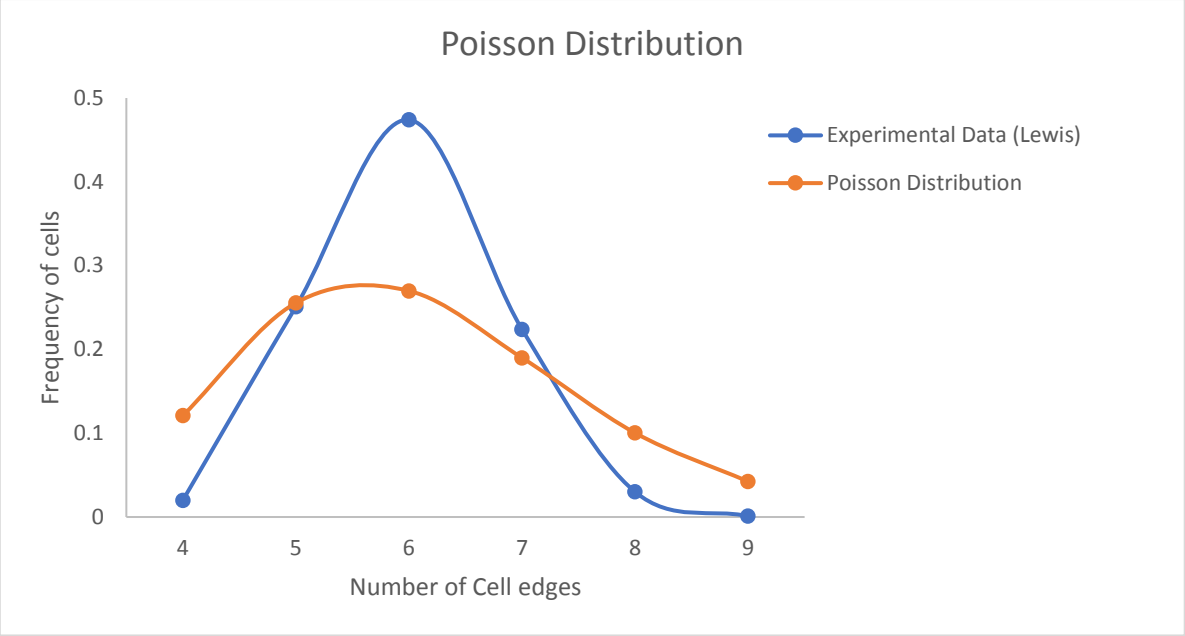



***Figure 2.2*** *Two separate plotted curves showing the experimental observations recorded by Lewis (Blue) along with the fitted Poisson distribution (Orange). The method of least squares is used to plot the nonlinear curve using Microsoft excel solver.*

Running the solver programme to define the minimum value for the sum of squared errors by changing the value of λ generated a value of:

$$\lambda = 2.112199$$

Due to there being an error value of $0.063$, this corresponds to clear, significant, deviations in the shape of the fitted Poisson distribution away from the experimental observations. Therefore, we can conclude that the Poisson distribution is not useful at replicating the statistics of the topology of proliferating epithelial tissue.

*Normal Distribution*

The second probability function that will be considered is the curve generated from the normal distribution. This is one of the most common distributions that are used in nature to describe random variables whose distributions are not known, such as the height or weight of individuals. These random variables can be expressed as the sum of many random variables and the Central Limit theorem states that this sum will be approximately normal. In the same manner as for the Poisson distribution, the method of least squares is applied to fit the recorded experimental data and we will let $x$ be the number of cell edges. The random variable is parametrised by three statistical measurements:

- The mean of the data = $\mu$ = 0
- Standard deviation = σ
- Variance = $\sigma^2$ = 1

The shape of the histogram is described as a bell curve where its width depends on the standard deviation. If $X$ is a continuous random normal variable, then we can say that $X \sim N(\mu, \sigma^2)$ and the P.D.F of this variable is described by the normal curve $f(x, \mu, \sigma)$ which has the following formula:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

| X | Fractions for cucumber distribution ($Yobs$) | Fitted Normal distribution ($Ycalc$) | Error squared $(Yobs - Ycalc)^2$ |
|---|---|---|---|
| 4 | 0.02 | 0.06076290 | 0.001662 |
| 5 | 0.251 | 0.25510946 | 0.0000169 |
| 6 | 0.474 | 0.40369971 | 0.004942 |
| 7 | 0.224 | 0.24078758 | 0.000282 |
| 8 | 0.03 | 0.05413193 | 0.000582 |
| 9 | 0.001 | 0.00458687 | 0.0000293 |
| Sum: | 1.0 | 0.958 | 0.006 |

***Table 6*** *Table showing the fractions of cell shapes obtained from cucumber distribution, where the Normal distribution has been fitted to the data.*

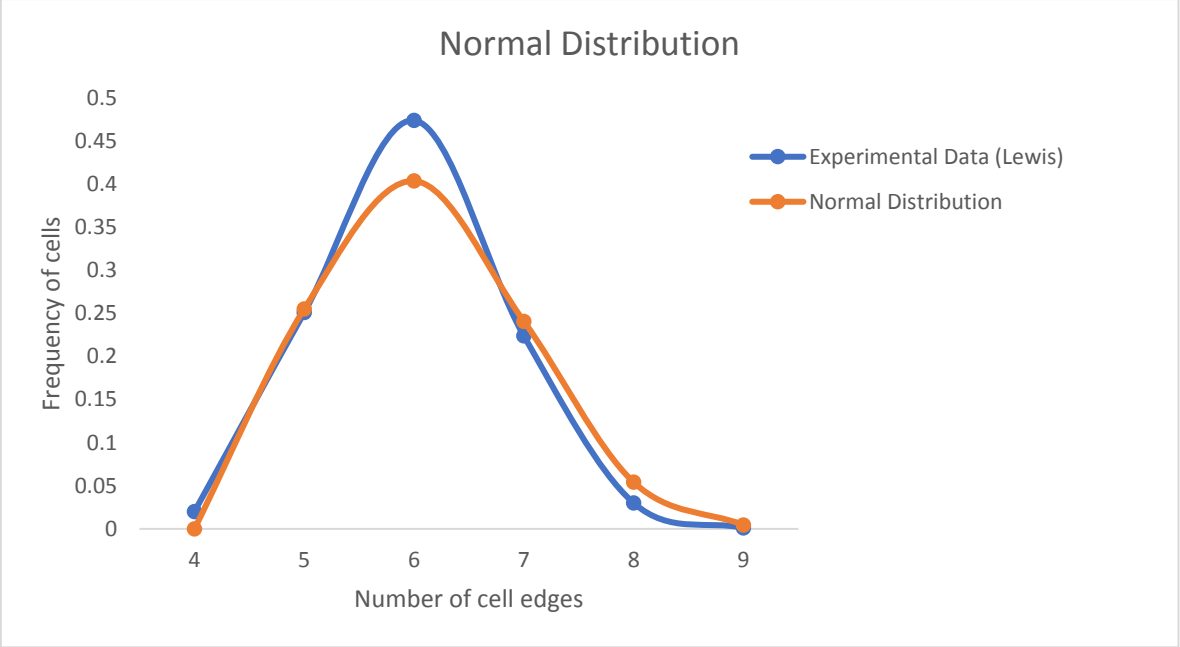



***Figure 2.3*** *Two separate plotted curves showing the topology of cucumber epidermis (Blue) along with the fitted Normal distribution (Orange). The method of least squares is used to plot the nonlinear curve using Microsoft excel solver.*

Running the solver programme to define the minimum value for the sum of squared errors by changing the value of $\mu$ and $\sigma$ generates values of:

$$\mu = 1.97039\ , \sigma = 1.01236$$

Due to there being a smaller error value of $0.006$ in comparison with the error value generated from the fitted Poisson distribution, this corresponds to less obvious deviations in the shape of the fitted Normal distribution away from the experimental observations. Therefore, we can conclude that the Normal distribution performs more accurately at replicating the statistics of the topology of proliferating epithelial tissue. However though, the symmetrical nature of the normal distribution is inconsistent with the unsymmetrical shape of the experimental histogram. In addition to this we know that $x \in \mathbb{R}$ for the Normal distribution whereas $x > 0$ in the experimental data since $x$ represents the various cell shapes that were observed.

*Log-Normal distribution*

For the final statistical model, we will consider the log-normal fitted histogram. The random variable $X$ has a logarithm which follows a normal distribution. For the standard normal variable $X$ it has a mean $\mu = 0$ and a variance $\sigma^2 = 1$ and can be defined as the following:

$$X = e^{\mu+\sigma Z}$$

$$\therefore logX = \log\left(e^{\mu+\sigma Z}\right)$$

$$\therefore logX = \mu + \sigma z$$

Since $Z$ is normal, $\mu + \sigma z$ is also normal which means that the logarithm of $X$ is normally distributed.

The log-normal distribution, where $x > 0$ satisfies the following equation:

$$f(x, \mu, \sigma) = \frac{1}{x \cdot \sigma\sqrt{2\pi}} e^{-\frac{(ln\, x-\mu)^2}{2\sigma^2}}$$

| **X** | **Fractions for Cucumber distribution ($Yobs$)** | **Fitted Log-Normal distribution ($Ycalc$)** | **Error squared $(Yobs - Ycalc)^2$** |
|---|---|---|---|
| 4 | 0.02 | 0.021062 | 0.00000113 |
| 5 | 0.251 | 0.204428 | 0.002169 |
| 6 | 0.474 | 0.465543 | 0.0000715 |
| 7 | 0.224 | 0.242236 | 0.000333 |
| 8 | 0.03 | 0.047396 | 0.000303 |
| 9 | 0.001 | 0.037601 | 0.00134 |
| Sum: | 1.0 | 1.018 | 0.004 |

***Table 7*** *Table showing the fractions of cell shapes obtained from cucumber distribution, where the Log-Normal distribution has been fitted to the data.*

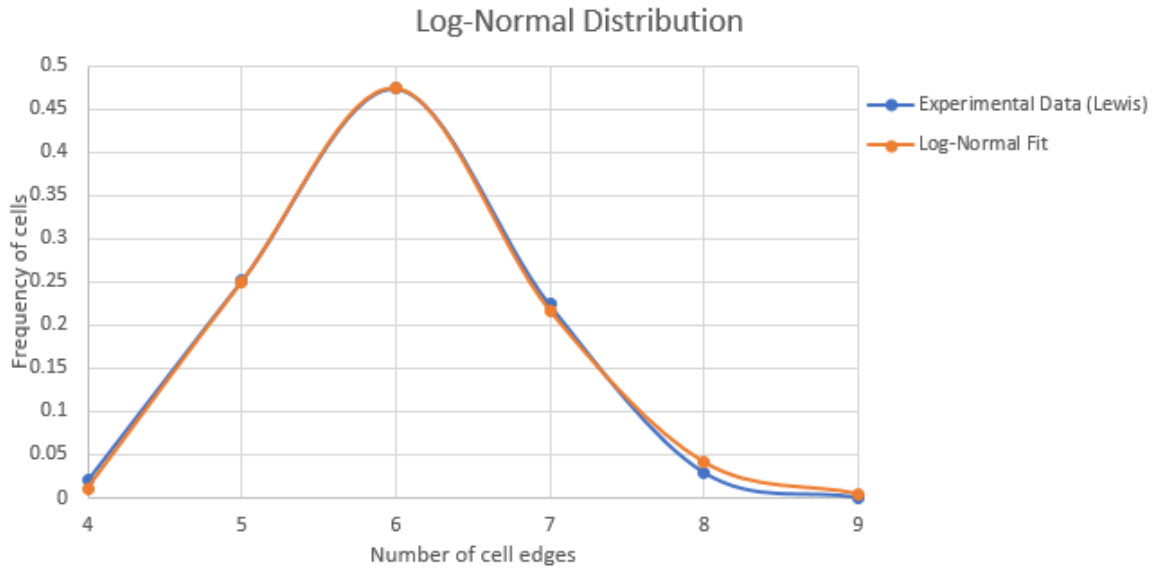



***Figure 2.4*** *Two separate plotted curves showing the topology of cucumber epidermis (Blue) along with the fitted Log-Normal distribution (Orange). The method of least squares is used to plot the nonlinear curve using Microsoft excel solver.*

Running the solver programme to define the minimum value for the sum of squared errors by changing the value of $\mu$ and $\sigma$ generates values of:

$$\mu = 1.789103 \, , \sigma = 0.139965$$

This model generates the smallest error value of $0.004$ in comparison with the error values generated from the fitted Poisson distribution and the Normal distribution, where the shape of the fitted Log-Normal distribution is highly like the shape of the experimental observations. Therefore, we can conclude that the Log- Normal distribution performs the most effectively at accurately predicting the statistics of the topology of proliferating epithelial tissue. However though, the log-normal distribution is a continuous function whereas the experimental data is discontinuous. Since the data observed from Lewis' experiments describes the distribution of cell-neighbour numbers, this is not a continuous variable. Therefore, this may account for the variations between the histograms generated from the three statistical models. Nevertheless, we know that the third statistical model performs the most effectively and therefore we can use the Chi-squared test to calculate how accurate this model fits the experimental data

| Number of Cell edges | Observed Cucumber distribution $(O_L i)$ | Log-Normal fit | Expected Cucumber distribution $(E_L i)$ |
|---|---|---|---|
| 4 | 20 | 0.011332 | 11.332 |
| 5 | 251 | 0.250102 | 250.102 |
| 6 | 474 | 0.474964 | 474.964 |
| 7 | 224 | 0.217392 | 217.392 |
| 8 | 30 | 0.041440 | 41.440 |
| 9 | 1 | 0.004512 | 4.512 |

***Table 8*** *Table displaying the observed and expected data for the probability distributions fitted to the experimental distribution of cucumber epidermis. These can then be used to verify whether the log-normal curve does fit Lewis' observations the best.*

- Null hypothesis $H_0$: Log-Normal distribution does fit the observed data.
- Alternative hypothesis $H_1$: Log-Normal distribution does not fit the observed data.
- Significance level of $5\%$: $\alpha = 0.05$
- Significance level of 1%: $\alpha = 0.01$
- Degrees of freedom: $(k-1) = 5$
- Critical value under null hypothesis: ${X_5}^2$ =11.07
- Critical value under null hypothesis: ${X_1}^2$ =15.09

Test statistic $x^2 = \frac{(20-11.332)^2}{11.332} + \frac{(251-250.102)^2}{250.102} + \frac{(474-474.964)^2}{474.964} + \frac{(224-217.392)^2}{217.392} + \frac{(30-41.440)^2}{41.440} + \frac{(1-4.512)^2}{4.512} = 12.73\ (4.sf)$

Since the calculated test statistic is greater than the critical value at the 5% level but less than that at the 1% level, we accept the null hypothesis and reject the alternative hypothesis at the 5% level. We can be $95\%$ confident that the Log-Normal distribution can be used to describe the fitted CED obtained for 4-9 sided cells. To conclude this chapter, we have discovered that the distributions of Drosophila, Xenopus and Hydra are like the distribution of cucumber. This confirms that natural epithelia do in fact display a topology which is universal across various tissue types. We then discovered that the log-normal distribution is the best probability density function out of the three to describe the universal shape of cucumber and hence it can be predicted that proliferating natural epithelia will display a topological distribution which can be described by a log-normal curve.

# Chapter 3. Continuous mathematical model for topology

In the preliminary report we defined mathematical equations which describe the changes made to the distribution of various polygonal cells during proliferation, such as the changes made to the number of edges of neighbouring cells due to the division of mother cells. We also explained the three scenarios of cell division based upon the investigations of (Gibson, et al., 2009). We then calculated a system of linear equations based upon the probabilities of various polygonal cells dividing, then applied Linear Algebra methods to solve these. We assumed that there are no cells in the population with more than nine edges or cells with less than four edges. In this chapter we initially alter our model slightly by allowing three sided cells to exist and summarise how extending the range of dividing mother cells shifts the shape of the stabilised CEDs. Finally, the exponential relationship between the number of edges and the probability of cell division is incorporated into the model assumptions, making the model equations non-linear, with the aim of making our analytical model based upon master equations more biologically accurate.

## Section 3.1 Master equations describing dynamics of epithelial tissue topology

As a reminder we will define the terms used in the system of equations. If we were to allow three-sided mother cells to exist in the population this would mean that the division line would have to cross any two edges, including any adjacent edges. Therefore, the generated pairs of daughter cells for each possible random division event will be different in comparison with that for the four-to-nine-sided continuous model. For simplicity the only matrices that will be presented here will be that for the daughter cells. If we consider an $i$ - sided mother cell in the proliferating tissue, then according to previous literature cell division will result in the generation of two geometrically identical daughter cells with their number of edges totalling to $i + 4$. If the two adjacent neighbouring cells are $j$ - sided then the division of this mother cell will add an extra edge to the shape of these two cells, each possessing $j + 1$ edges. As a reminder we will define the terms used in the linear system of equations:

- The total number of all cell shapes at time any point in time, $N(t)$.
- The number of $i$ - sided cells in the population of proliferating tissue, $N_i(t)$.
- $M_i$: The balance from the removal of $i$ - sided mother cells and addition of two $i$ – sided daughter cells.
- $N_i$: Considers the addition of edges of neighbouring cells after each division.
- The fraction of $i$ - sided cells in the population of proliferating tissue, $p_i(t) = \frac{N_i(t)}{N(t)}$

For each of the considered scenarios of cell division, we will have a system of linear ordinary differential equations describing the evolution of the polygonal cells in the tissue, previously stated to be $\dot{p}_i$. The coefficient of the matrix $\bar{P}$ is a combination of matrices which describe the topological changes due to tissue growth and can be described by the following expression:

$$\begin{bmatrix}\dot{p}_3\\ \dot{p}_4\\ \dot{p}_5\\ \dot{p}_6\\ \dot{p}_7\\ \dot{p}_8\\ \dot{p}_9\end{bmatrix} = (M + D + N - I) \cdot \begin{bmatrix}p_3\\ p_4\\ p_5\\ p_6\\ p_7\\ p_8\\ p_9\end{bmatrix} \tag{3.1}$$

Here we have made use of the Identity matrix, I, and have made use of other matrices which can be expressed in the following matrices:

$$N = \begin{bmatrix}-2 & 0 & 0 & 0 & 0 & 0\\ 2 & -2 & 0 & 0 & 0 & 0\\ 0 & 2 & -2 & 0 & 0 & 0\\ 0 & 0 & 2 & -2 & 0 & 0\\ 0 & 0 & 0 & 2 & -2 & 0\\ 0 & 0 & 0 & 0 & 2 & -2\\ 0 & 0 & 0 & 0 & 0 & 2\end{bmatrix} \text{ and } M = -I. \tag{3.2}$$

Thus, we can express equation (3.1) as $\dot{\boldsymbol{p}} = A\bar{P}$ where $A = D + N - 2I$

*UNIFORMLY ORIENTATED*: The division line crosses any two edges of the mother cell, where for this scenario they all occur with equal probability.

$$D = \begin{bmatrix}1 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & 0\\ 1 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & \frac{1}{3}\\ 0 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & \frac{1}{3}\\ 0 & 0 & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & \frac{1}{3}\\ 0 & 0 & 0 & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & \frac{1}{3}\\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{7} & \frac{1}{3}\\ 0 & 0 & 0 & 0 & 0 & \frac{2}{7} & \frac{1}{3}\end{bmatrix}$$

$$\therefore A = D + N - 2I = \begin{bmatrix} -3 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & 0 \\ 3 & -\frac{10}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & \frac{1}{3} \\ 0 & \frac{8}{3} & -\frac{7}{2} & \frac{2}{5} & \frac{1}{3} & \frac{2}{7} & \frac{1}{3} \\ 0 & 0 & \frac{5}{2} & -\frac{18}{5} & \frac{1}{3} & \frac{2}{7} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{12}{5} & -\frac{11}{3} & \frac{2}{7} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{7}{3} & -\frac{26}{7} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{16}{7} & -\frac{5}{3} \end{bmatrix}$$

*BINOMIALLY ORIENTED*: The probabilities for different orientations of a cleavage line are non-uniform and given by a binomial probability.

$$D = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{32} & 0 \\ 1 & 1 & \frac{3}{4} & \frac{1}{2} & \frac{5}{16} & \frac{3}{16} & \frac{1}{16} \\ 0 & \frac{1}{2} & \frac{3}{4} & \frac{3}{4} & \frac{5}{8} & \frac{15}{32} & \frac{5}{16} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{5}{8} & \frac{5}{8} & \frac{5}{8} \\ 0 & 0 & 0 & \frac{1}{8} & \frac{5}{16} & \frac{15}{32} & \frac{5}{8} \\ 0 & 0 & 0 & 0 & \frac{1}{16} & \frac{3}{16} & \frac{5}{16} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{32} & \frac{1}{16} \end{bmatrix}$$

$$\therefore A = D + N - 2I = \begin{bmatrix} -3 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{32} & 0 \\ 3 & -3 & \frac{3}{4} & \frac{1}{2} & \frac{5}{16} & \frac{3}{16} & \frac{1}{16} \\ 0 & \frac{5}{2} & -\frac{13}{4} & \frac{3}{4} & \frac{5}{8} & \frac{15}{32} & \frac{5}{16} \\ 0 & 0 & \frac{9}{4} & -\frac{7}{2} & \frac{5}{8} & \frac{5}{8} & \frac{5}{8} \\ 0 & 0 & 0 & \frac{17}{8} & -\frac{59}{16} & \frac{15}{32} & \frac{5}{8} \\ 0 & 0 & 0 & 0 & \frac{33}{16} & -\frac{61}{16} & \frac{5}{16} \\ 0 & 0 & 0 & 0 & 0 & \frac{65}{32} & -\frac{31}{16} \end{bmatrix}$$

*EQUAL SPLIT*: The division line connects only two opposing edges of the mother cell.

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

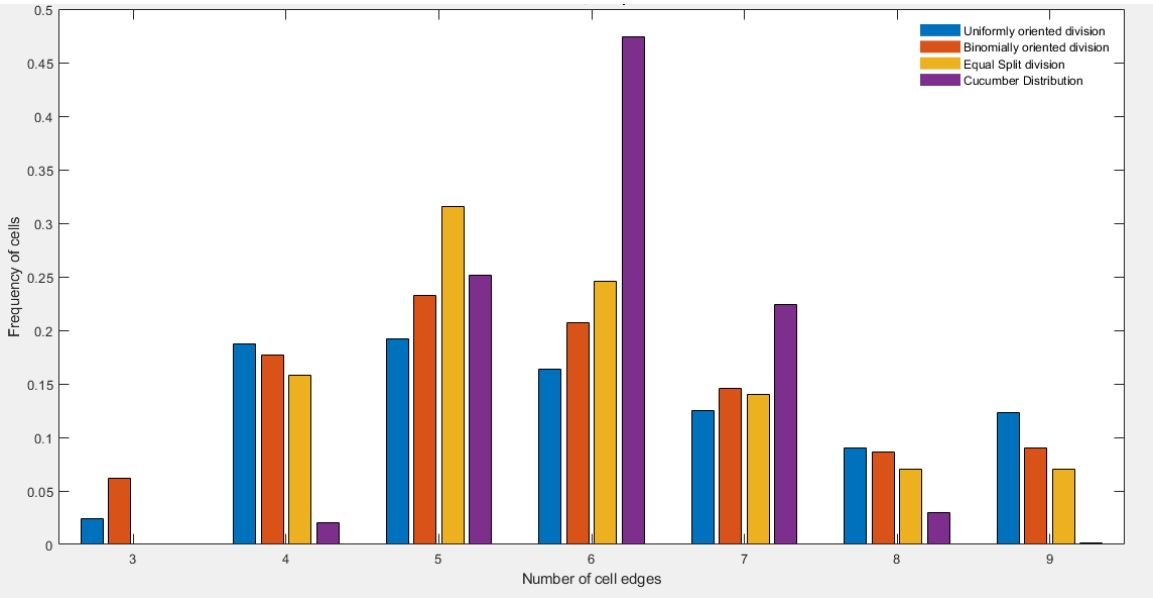
$$\therefore A = D + N - 2I = \begin{bmatrix} -3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & -3 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & -4 & 1 & 2 & 1 \\ 0 & 0 & 0 & 2 & -4 & 0 & 1 \\ 0 & 0 & 0 & 0 & 2 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix}$$

For each of the considered scenarios of cell division, we now have a system of linear ordinary differential equations describing the evolution of the polygonal cells in the tissue of the from $\dot{\boldsymbol{p}} = A\boldsymbol{p}$. We now solve the systems of equations in the same way for the previous four-to-nine-sided model, where the entries of the matrix $A$ are the coefficients of the probabilities of a mother cell dividing into a pair of daughter cells with at least one of these being $i$-sided

## Section 3.2 Comparing the generated CED using various linear models.

The various systems of linear equations were inputted into MATLAB which replicates the evolving polygonal distributions of a proliferating virtual tissue. Boundary conditions were then inputted to reflect the topology of the tissue before development occurred. We assumed initially that the cells in the tissue were hexagonal. The following are the simulated bar charts for each of the stabilised CEDs for various ranges of mother cell edges after running the program for a few generations.

**(A)**




**(B)**

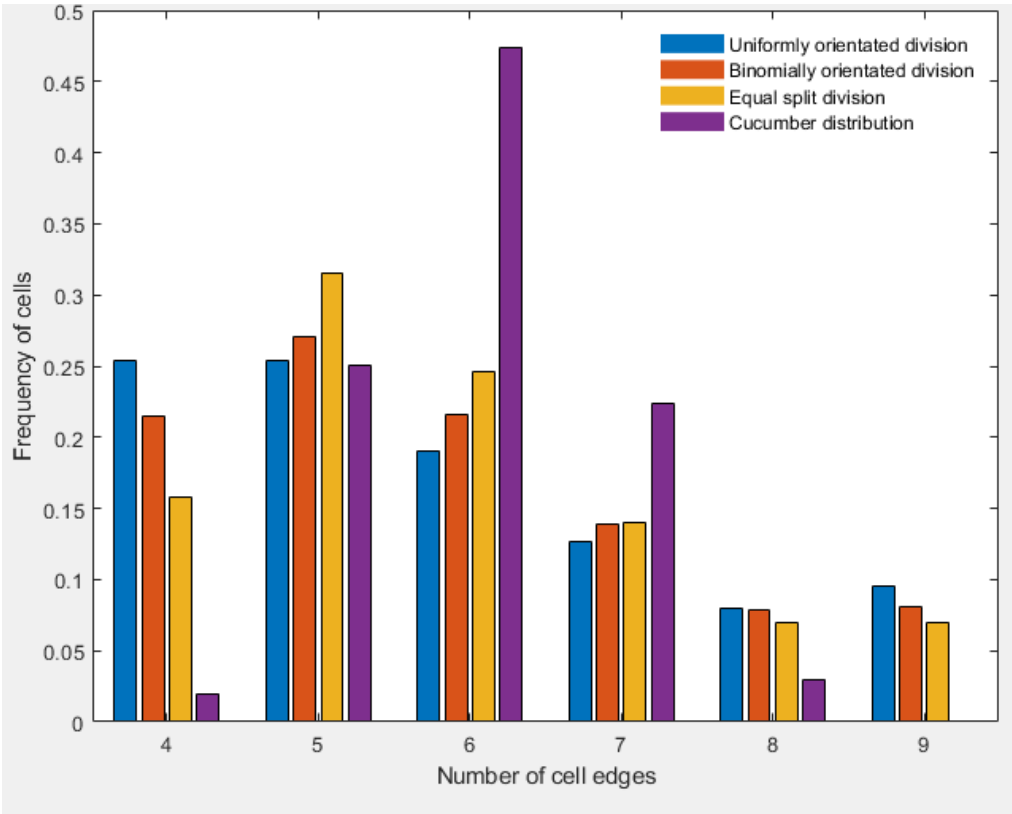



***Figure 3.2*** *Illustrating the accuracy of simulated stabilised CEDs for Uniform, Binomial and Equal split scenarios of cellular division, respectively, for various ranges of mother cell edges using analytical models. (A) presents the stabilised results using the 3-to-9-sided model and (B) illustrates the stabilised results using the 4-to-9-sided model.*

If we compare the evolved distribution generated from the three to nine-sided model (A) with that from the four to nine-sided model (B), there are significant differences between the shapes generated when mother cells divide according to the uniform scenario and the Binomial scenario. If we compare the evolved distribution generated from the three to nine-sided model with that from the four to nine-sided model, there is no difference between the shapes generated when mother cells divide according to the equal split scenario. The number of discrepancies between all models seems to increase as the range of mother cell edges increases. By allowing three-sided mother cells to divide the three to nine-sided model still predicts that most polygonal cells for all scenarios in proliferating tissue are pentagonal whereas experimental observations show that most polygonal cells in organisms are hexagonal. The uniform orientated model generates a CED which significantly shifts the universal histogram, overpredicting a population size of four-sided cells and underpredicting a population size of six sided cells.

## Section 3.3 Including probability of division on sidedness

Within the documented experiments of cucumber epidermis, (Lewis, 1928) calculated the respective ratios of dividing mother cells to resting mother cells for all polygons. If we consider the proportion of each of these ratios, we notice that dividing mother cells with a larger number of edges tend to have a greater chance of dividing which seems to be exponentially described. Hence mother cells that have more edges tend to undergo cellular division more frequently in comparison to the rate of division of mother cells with a smaller number of edges. Since the area of a cell increases with its number of edges, this implies that larger mother cells divide more often than smaller mother cells. Biologically, this relationship could be explained due to larger cells having a greater volume and a smaller surface area to volume ratio which reduces the maintenance of the dynamical equilibrium of vital nutrients and gases. Hence, these larger cells are more likely to divide to restore this equilibrium. Furthermore, mother cells are required to reach full maturity before division can occur and those cells in the population that are older have an increased number of edges over time due to the division of neighbouring cells.

| **Number of Cell edges** $(\mathrm{i})$ | **4** | **5** | **6** | **7** | **8** | **9** | **10** | |
|---|---|---|---|---|---|---|---|---|
| **1000 resting cells** | *20* | *251* | *474* | *224* | *30* | *1* | *0* | *1000* |
| **1000 dividing cells** | *0* | *16* | *255* | *478* | *224* | *26* | *1* | *1000* |
| **Ratio of dividing cells to resting cells** $(\mathbf{X_i})$ | *0* | *0.0637* | *0.538* | *2.1339* | *7.4667* | *26* | | |
| $\boldsymbol{\sigma_i = \frac{X_i}{\sum_{i=4}^{9} X_i}}$ | *0* | *0.0018* | *0.0149* | *0.0589* | *0.2062* | *0.7182* | | |

***Table 9*** *Lewis` calculations illustrating the distribution, with respect to their number of edges, of polygons for 1000 resting and dividing cells. Cells with a greater number of edges are more likely to divide.*

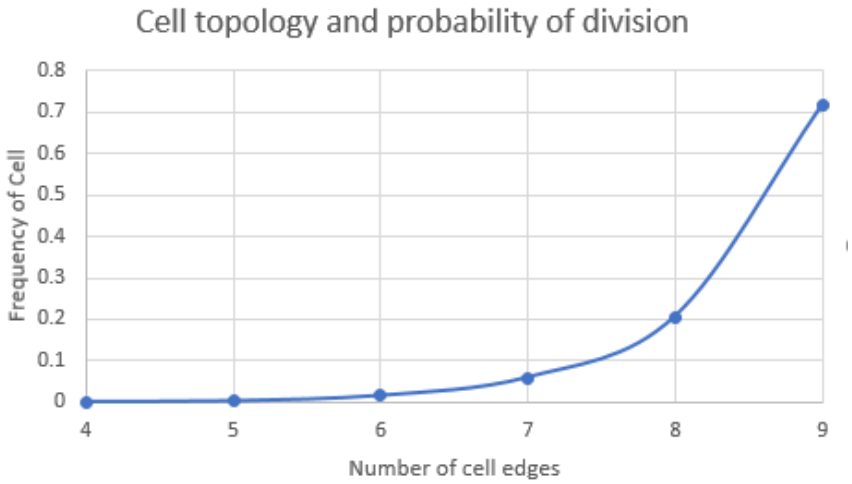



***Figure 3.4*** *Plot of data presented in final row of table 9, which has been inputted into Microsoft excel. The trend line illustrates the increased probability of mother cell division as the number of cell edges increases which seems to be exponential.*

In this model we will alter our original formulas for the three scenarios of cell division which were presented in the preliminary report. We initially assumed that $p_j$ is the probability of a mother cell being $i$ – sided before cell division occurring and our model assumed no spatial correlations between the sides of cells, and hence all cells in the population had an equal probability of dividing for each generation of proliferation (Gibson, 2009). Therefore, we assumed that $p_i^*(t) = p_i(t)$ where $p_i(t) = \frac{f_i{}^m(t)}{f_i(t)}$, which is the fraction of $i$ - sided mother cells in the population of proliferating tissue. It can be seen from figure 3.4 that the probability of a particular mother cell dividing increases at an exponential rate with an increase in the number of edges of cells. We will now assume that the probability of a dividing mother cell

being $i$ – sided is proportional to $\sigma_i \cdot \frac{{f_i}^m(t)}{f_i(t)}$ where the values of $\sigma_i$ were calculated using the data in Table 9. Figure 3.4 illustrates this exponential relationship which exists between the probability of mother cell division and the number of cell edges. This alteration leads to the following expression.

$$\bar{P}^* = \frac{\sigma_i p_i}{\sum_i \sigma_i p_i} \tag{3.3}$$

If we were to incorporate this property into our model involving matrices this will result in a system of non-linear differential master equations, which means that we cannot solve them analytically, but we will be able to solve them numerically. We have done this using a constructed MATLAB code which relies upon a defined set of initial conditions to generate solutions to the equations. We used a loop within MATLAB to solve the system of nonlinear equations.

At time t= 0, $\bar{P} = \bar{P}_0$ where the initial vector can be represented in the following form:

$$\bar{P}_0{}^* = \frac{\sigma_0 \bar{p}_0}{\sum \sigma \bar{p}}$$

If we consider the rate of change of the vector containing the evolving fractions of dividing mother cells, we can express this as the sum of two matrices A and B. These are the coefficients of the probabilities of a mother cell dividing into a pair of daughter cells with at least one of these being $i$-sided. However, the incorporation of exponential dependence on the number of dividing mother cell edges does not influence the probability of neighbouring cells dividing, and hence the expression becomes:

$$\frac{\bar{p}}{\Delta t} = \dot{\bar{P}} = \bar{A} \cdot \bar{P}^* + \bar{B} \cdot \bar{P} \tag{3.4}$$

After using concepts of linear algebra, we can amend and substitute expression (3.1) for 4 to 9 sided cells into the expression for the rate of change of evolving fractions of dividing mother cells (3.4) and we can incorporate the necessary matrices impacting the topology of the proliferating virtual tissue; dividing mother cells, generated daughter cells, neighbouring cells, and the identity matrix to give the following expression:

$$\begin{bmatrix}\dot{p}_4\\ \dot{p}_5\\ \dot{p}_6\\ \dot{p}_7\\ \dot{p}_8\\ \dot{p}_9\end{bmatrix} = (M+D)\begin{bmatrix}p^*{}_4\\ p^*{}_5\\ p^*{}_6\\ p^*{}_7\\ p^*{}_8\\ p^*{}_9\end{bmatrix} + (N-I)\cdot\begin{bmatrix}p_4\\ p_5\\ p_6\\ p_7\\ p_8\\ p_9\end{bmatrix} \tag{3.5}$$

We can substitute the amended results for the matrices M, N, and I into equation (3.5) from equation (3.2) and obtain a system of three nonlinear equations where for each scenario of division the only matrix that will change is the matrix for the resulting daughter cells, D.

*UNIFORMLY ORIENTATED*:
$$D = \begin{bmatrix} 2 & 1 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 1 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$\therefore \begin{bmatrix}\dot{p}_4\\ \dot{p}_5\\ \dot{p}_6\\ \dot{p}_7\\ \dot{p}_8\\ \dot{p}_9\end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & -\frac{1}{3} & \frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & \frac{2}{5} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & -\frac{3}{5} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & -\frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix}\sigma_4 p_4\\ \sigma_5 p_5\\ \sigma_6 p_6\\ \sigma_7 p_7\\ \sigma_8 p_8\\ \sigma_9 p_9\end{bmatrix} \cdot \frac{1}{\sum \sigma_i p_i} + \begin{bmatrix} -3 & 0 & 0 & 0 & 0 & 0 \\ 2 & -3 & 0 & 0 & 0 & 0 \\ 0 & 2 & -3 & 0 & 0 & 0 \\ 0 & 0 & 2 & -3 & 0 & 0 \\ 0 & 0 & 0 & 2 & -3 & 0 \\ 0 & 0 & 0 & 0 & 2 & -1 \end{bmatrix} \cdot \begin{bmatrix}p_4\\ p_5\\ p_6\\ p_7\\ p_8\\ p_9\end{bmatrix}$$

*BINOMIALLY ORIENTED*:
$$D = \begin{bmatrix} 2 & 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ 0 & 1 & 1 & \frac{3}{4} & \frac{1}{2} & \frac{5}{16} \\ 0 & 0 & \frac{1}{2} & \frac{3}{4} & \frac{3}{4} & \frac{5}{8} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{5}{8} \\ 0 & 0 & 0 & 0 & \frac{1}{8} & \frac{5}{16} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{16} \end{bmatrix}$$

$$\therefore \begin{bmatrix} \dot{p}_4 \\ \dot{p}_5 \\ \dot{p}_6 \\ \dot{p}_7 \\ \dot{p}_8 \\ \dot{p}_9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ 0 & 0 & 1 & \frac{3}{4} & \frac{1}{2} & \frac{5}{16} \\ 0 & 0 & -\frac{1}{2} & \frac{3}{4} & \frac{3}{4} & \frac{5}{8} \\ 0 & 0 & 0 & -\frac{3}{4} & \frac{1}{2} & \frac{5}{8} \\ 0 & 0 & 0 & 0 & -\frac{7}{8} & \frac{5}{16} \\ 0 & 0 & 0 & 0 & 0 & -\frac{15}{16} \end{bmatrix} \cdot \begin{bmatrix} \sigma_4 p_4 \\ \sigma_5 p_5 \\ \sigma_6 p_6 \\ \sigma_7 p_7 \\ \sigma_8 p_8 \\ \sigma_9 p_9 \end{bmatrix} \cdot \frac{1}{\sum \sigma_i p_i} + \begin{bmatrix} -3 & 0 & 0 & 0 & 0 & 0 \\ 2 & -3 & 0 & 0 & 0 & 0 \\ 0 & 2 & -3 & 0 & 0 & 0 \\ 0 & 0 & 2 & -3 & 0 & 0 \\ 0 & 0 & 0 & 2 & -3 & 0 \\ 0 & 0 & 0 & 0 & 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \end{bmatrix}$$

*EQUAL SPLIT*:
$$D = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\therefore \begin{bmatrix} \dot{p}_4 \\ \dot{p}_5 \\ \dot{p}_6 \\ \dot{p}_7 \\ \dot{p}_8 \\ \dot{p}_9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 2 & 1 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \sigma_4 p_4 \\ \sigma_5 p_5 \\ \sigma_6 p_6 \\ \sigma_7 p_7 \\ \sigma_8 p_8 \\ \sigma_9 p_9 \end{bmatrix} \cdot \frac{1}{\sum \sigma_i p_i} + \begin{bmatrix} -3 & 0 & 0 & 0 & 0 & 0 \\ 2 & -3 & 0 & 0 & 0 & 0 \\ 0 & 2 & -3 & 0 & 0 & 0 \\ 0 & 0 & 2 & -3 & 0 & 0 \\ 0 & 0 & 0 & 2 & -3 & 0 \\ 0 & 0 & 0 & 0 & 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \end{bmatrix}$$

The converged stationary solutions were found using the numerical simulation, normalized in a way that the sum of all its components equal one. The stabilised results were plotted to provide the following bar chart.

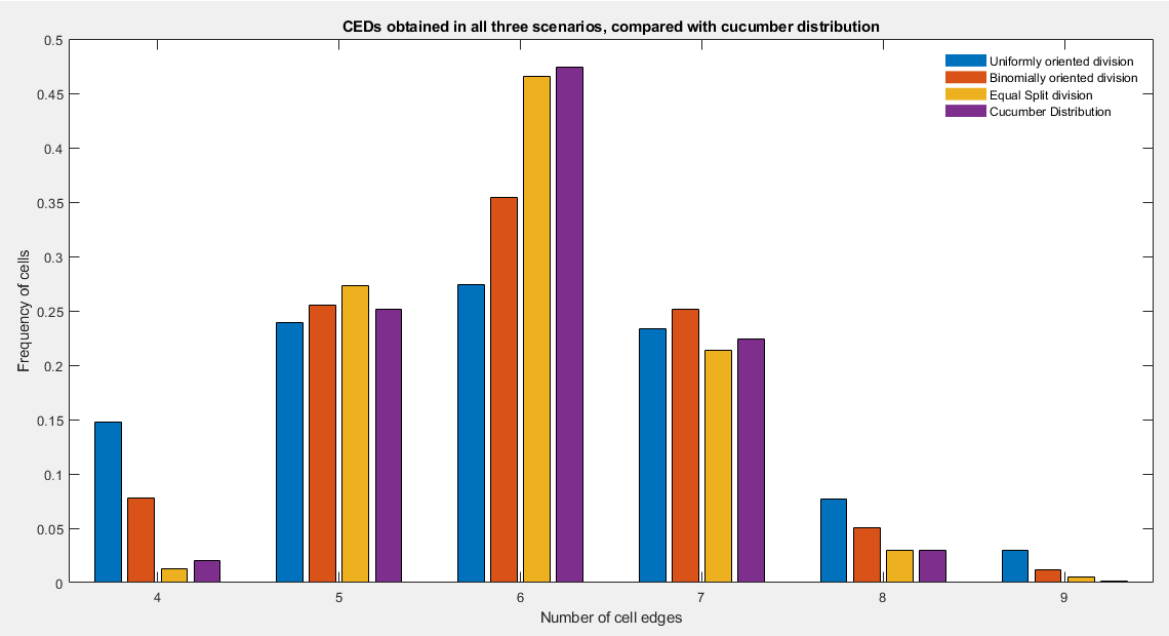



***Figure 3.3*** *Simulated CEDs for all three scenarios, for 4 to 9 sided cells along with experimental recordings (Lewis, 1928), including dependence of the frequency of cellular division on the number of edges of the dividing mother cell.*

From looking at the results displayed in figure 3.3 the stabilised CEDs generated from our non-linear models for each scenario of cell division reproduces the statistics of the experimental observations more accurately and are now more comparable, were less discrepancies between the experimental results and the modelled bar charts are observed. For all scenarios of cell division, the model predicts most polygonal cells in proliferating tissue are hexagonal, consistent with experimental observations. The equal split model gives a steady state converged solution which agrees with the experimental observations the most, indicating that this model performs the most effectively at predicting the stabilised topology. The uniformly orientated model generates a CED which is least like the universal shape, still overpredicting a population size of 4 sided cells and underpredicting a population size of 6 sided cells. However, this model predicts the population size of 5-sided and 7-sided cells remarkably well. The binomially orientated model generates a CED which is more comparable to experimental results.

To further our analysis, we will determine how similar the shapes of the stabilised CED produced from the equal split scenario of cell division and cucumber are. If these are statistically similar these can be treated as the same distributions which confirms our conclusion that the equal split scenario is the best model, where the necessary data has been presented in the following table:

| Number of Cell edges (i) | Equal split distribution $(F_E i)$ | Observed Cucumber distribution $(O_L i)$ | Expected Cucumber distribution $(E_L i)$ |
|---|---|---|---|
| 4 | 0.0128 | 20 | 12.8 |
| 5 | 0.2729 | 251 | 272.9 |
| 6 | 0.4659 | 474 | 465.9 |
| 7 | 0.2135 | 224 | 213.5 |
| 8 | 0.0298 | 30 | 29.8 |
| 9 | 0.0050 | 1 | 5 |
| Total Observations | | 1000 | 1000 |

***Table 10*** *Table displaying the observed and expected data for cucumber with the aim of concluding whether the equal split scenario of cell division (incorporating exponential relationship) and cucumber produce statistically similar CED shapes from non-linear model.*

- Null hypothesis $H_0$: Distributions for cucumber and equal split are statistically similar.
- Alternative hypothesis $H_1$: Distributions are not statistically similar.
- Significance level of $5\%$: $\alpha = 0.05$
- Significance level of $1\%$: $\alpha = 0.01$
- Degrees of freedom: $(6-1) = 5$
- Critical value under null hypothesis: ${X_5}^2 = 11.070$
- Critical value under null hypothesis: ${X_1}^2 = 15.086$

$x^2 = \frac{(20-12.8)^2}{12.8} + \frac{(251-272.9)^2}{272.9} + \frac{(474-465.9)^2}{465.9} + \frac{(224-213.5)^2}{213.5} + \frac{(30-29.8)^2}{29.8} + \frac{(1-5)^2}{5} = 9.6660\ (4.dp)$

Since the calculated test statistic is less than the critical value, we accept the null hypothesis at the 5% and 1% level of significance and reject the alternative hypothesis with 99% confidence. Hence the equal split distribution is statistically like the cucumber distribution and therefore can be treated as the same distribution.

To conclude this chapter, we introduced Matrices which form a system of linear ordinary differential equations describing the evolution of the polygonal cells in the tissue for each of the three scenarios of cell division. We show that to produce stabilised CEDs from using our continuous analytical model which reflect the topological dynamics of proliferating tissue, we need to incorporate the mathematical observation that the probability of cellular division increases exponentially with the number of edges, by making our model non-linear. We used a statistical test to show that mother cells should divide and produce two equally sized daughter cells.

# Chapter 4. Cellular Automata model

The understanding of geometric interactions that exist between cells and their neighbours due to various orientations of the division plane can give Mathematicians a better understanding of tissue patterns observed in natural epithelia (Kachalo, et al., 2015). It is the combination of mathematical modelling using differential equations and computer programming using Cellular automata (CA) which have an important role in the study of biological systems (Souza-e-Silva, et al., 2009). These Biological systems have been successfully modelled using CA computational methods. CA will be used to define a population of cells as a discrete dynamical system in space and time, where individual cells will behave according to a set of rules incorporated into the program (Alemani, et al., 2012). In this section we will describe a computational algorithm which replicates the events of a growing virtual tissue due to multiple events of proliferation where we incorporate the random orientation of the division plane. We aim to verify that our evolved CEDs generated from the continuous models for the various scenarios of cellular division, which is based on master equations, are accurate. To simulate a growing tissue due to proliferation, we are required to simulate cell division. The program is initiated with 10 polygonal cells and MATLAB runs the code until the population reaches a maximum of 1000 cells. The program randomly selects a set of numbers between 4 and 9. Each of the mother cells in the population is therefore associated with an integer between 4 and 9, corresponding to their respective number of edges before division occurs. The following steps were taken to initiate tissue growth at each time step corresponding to a single division event:

***Step 1*** – Randomly select a number from the existing set.

***Step 2*** - The randomly selected number from this existing set is replaced by two other numbers calculated according to the various scenarios of cell division. The original number corresponds to the number of edges of the mother cell and two new numbers corresponds to the number of edges of two daughter cells.

***Step 3*** – At each time step which division occurs, two other cells are chosen randomly from the population corresponding to the neighbours of the dividing cell for adding an extra edge to each one. For the equal split scenario if the mother cell has an even number of edges, then each daughter cell has $(i+4)/2$ edges. If the mother cell has an odd number of edges, then one daughter cell will have $(i+5)/2$ edges and the other daughter will have $(i+3)/2$ edges. At the end of each iteration, one mother cell disappears, two daughter cells appear, and two other cells receive an extra edge.

The process of tissue proliferation is repeated many times and eventually stabilises to an equilibrium CED, where figure 4.1 shows these for the independent models.

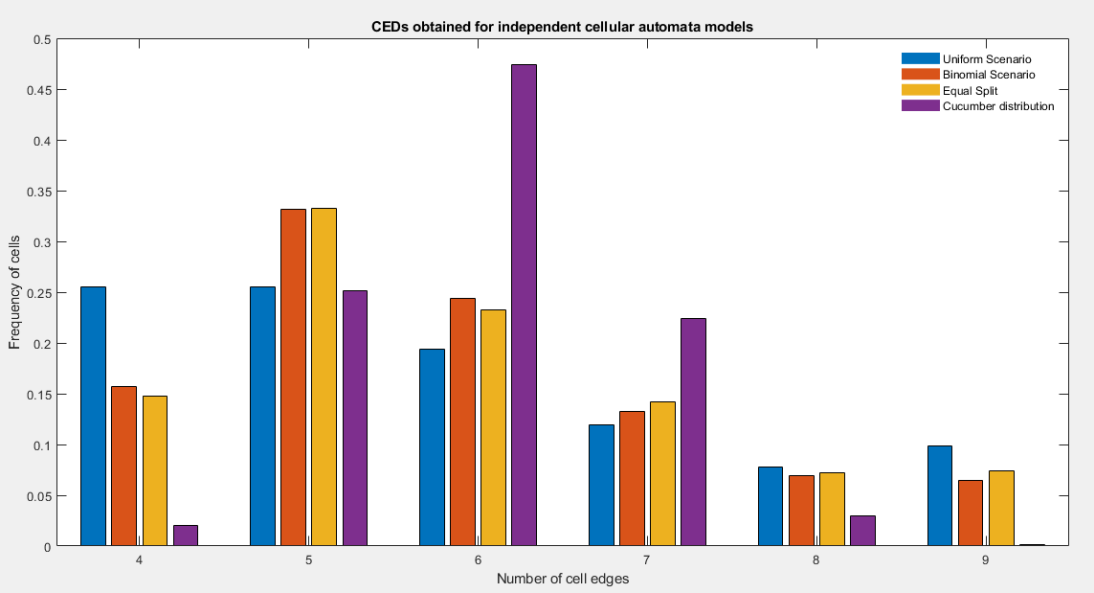



***Figure 4.1*** *Simulated stabilised CEDs for each respective scenario of cell division using the independent Cellular Automata code, in comparison with the experimental observations made by (Lewis, 1928).*

We can see that all the stabilised CEDs generated do not reproduce the shape of the cucumber CED. In the preliminary dissertation we concluded that the linear model represented by master equations assumes cells divide with a probability that does not depend on their edges. This is because the GPNP model which we relied upon for the construction of the linear model assumes that the probability of cell division is independent of the number of its edges. However, Lewis observed a natural law in proliferating cucumber epidermis; there is an exponential relationship between the probability of cells dividing and the cell shapes. Therefore, our bar charts which illustrates the evolution of CEDs does not reflect the experimental observations for the simple fact of the over simplicity in the CA assumptions. While we have successfully incorporated some biological assumptions of natural proliferating epithelia, this apparently is insufficient to successfully replicate the universal shape of the bar charts, presented in the Literature. Therefore, incorporating spatial correlations between cells may improve the accuracy of the bar charts obtained from the CA model. We will incorporate the exponential dependence of dividing mother cell edges by altering the independent cellular automata code slightly, including defining an array which contains the respective probabilities of each of the polygonal cells dividing. The stabilised CEDs for the dependent cellular automata model is displayed in figure 4.2.

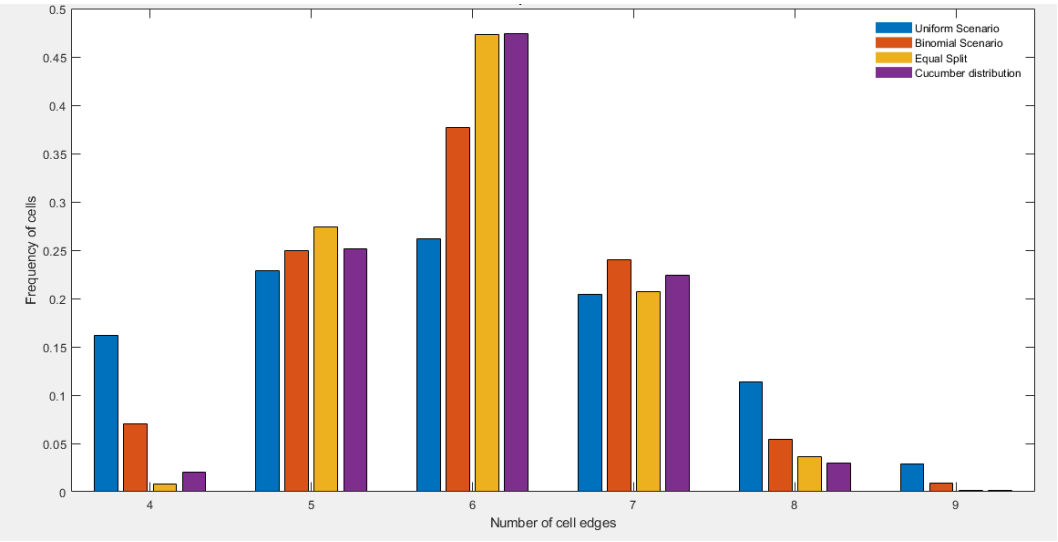



***Figure 4.2*** *Simulated stabilised CEDs for each respective scenario of cell division using the dependent Cellular Automata code, in comparison with the experimental observations made by (Lewis, 1928).*

From observation incorporating the exponential relationship into the cellular automata model generates stabilised CEDs that are representative of experimental observations, where we can now see that all models produce a hexagonal topology. Hence, we can be confident that it's important to consider all factors when it comes to tissue proliferation, where we have considered the equal split method of mother cell division and the exponential relationship between the topology of dividing mother cells and their respective probabilities of division occurring in a random event. This also confirms our results that we obtained using the model constructed upon continuous master equations.

In the same way as we analysed for the dependent model constructed upon master equations (see section 3.3), we will determine how similar the shapes of the stabilised CEDs produced from the equal split scenario of cell division and cucumber are. If these are statistically similar these can be treated as the same distributions which confirms our conclusion that the equal split scenario is the best model, where the stabilised CEDs from the equal split models is presented in figure 4.3.

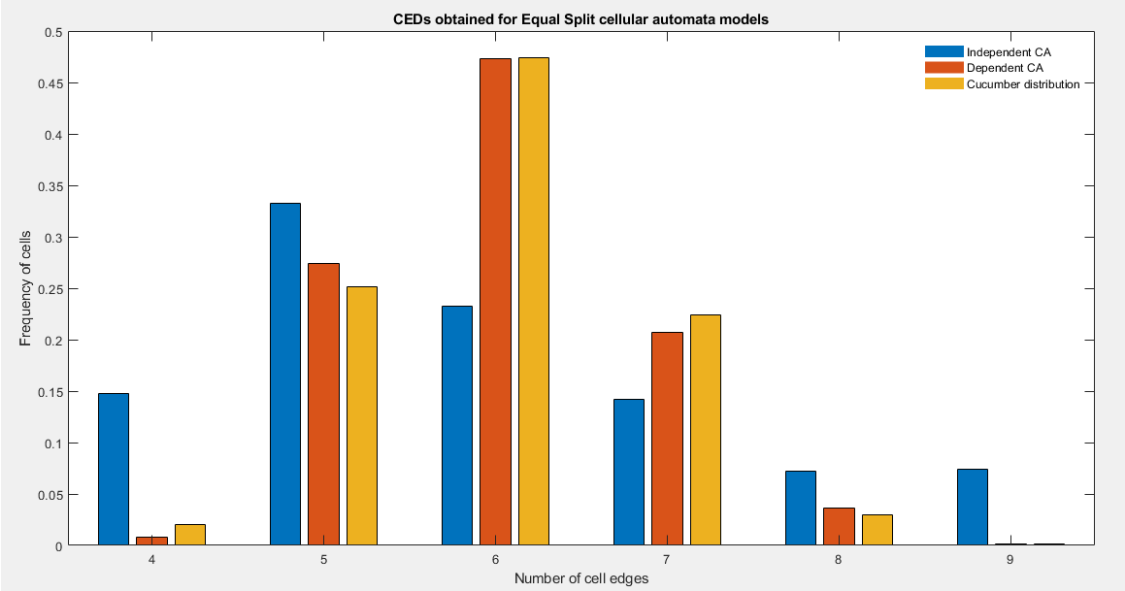



***Figure 4.3*** *Simulated stabilised CEDs for equal split cell division using both the dependent and independent Cellular Automata models, in comparison with the experimental observations made by (Lewis, 1928).*

The following table presents the raw data obtained from the produced stabilised distributions of cells by implementing the equal split scenario of cell division, required for the Xi-squared statistical test:

| **Number of Cell edges** (i) | **Equal split distribution** $(F_E i)$ | **Observed Cucumber distribution** $(O_L i)$ | **Expected Cucumber distribution** $(E_L i)$ |
|---|---|---|---|
| 4 | 0.008 | 20 | 8 |
| 5 | 0.274 | 251 | 274 |
| 6 | 0.473 | 474 | 473 |
| 7 | 0.207 | 224 | 207 |
| 8 | 0.036 | 30 | 36 |
| 9 | 0.001 | 1 | 1 |
| Total Observations | | 1000 | 1000 |

***Table 11*** *Table displaying the observed and expected data for cucumber with the aim of concluding whether the equal split scenario of cell division (incorporating exponential relationship) and cucumber produce statistically similar CED shapes using CA model.*

- Null hypothesis $H_0$: Distributions for cucumber and equal split are statistically similar.
- Alternative hypothesis $H_1$: Distributions are not statistically similar.
- Significance level of 5%: $\alpha = 0.05$
- Significance level of 1%: $\alpha = 0.01$
- Degrees of freedom: $(6 - 1) = 5$
- Critical value under null hypothesis: ${X_5}^2 = 11.070$
- Critical value under null hypothesis: ${X_1}^2 = 15.086$

$$x^2 = \frac{(20-8)^2}{8} + \frac{(251-274)^2}{274} + \frac{(474-473)^2}{473} + \frac{(224-207)^2}{207} + \frac{(30-36)^2}{36} = 22.3\ (1.dp)$$

Since the calculated test statistic is greater than the critical value at the 1%, we reject the null hypothesis at the 1% level of significance and accept the alternative hypothesis with 99% confidence. Hence the equal split distribution is not statistically like the cucumber distribution and therefore can cannot be treated as the same distribution. To conclude this chapter, it is obvious that incorporating the exponential relationship into our CA models does significantly improve the shape of all stabilised CEDs and it confirms that the mother cells do divide roughly equally.

# Chapter 5. Discussion

In this project we addressed how the dynamics of epithelial tissue topology could be described statistically and modified our analytical models' assumptions to determine if the generated stabilised CEDs reflected the universal shape better. We also produced a cellular automata code based upon a set of rules which incorporates the random orientation of the division plane during tissue growth. It was discovered that the shape of the universal CEDs observed in natural epithelia can be described by a log-normal fit. This becomes useful for mathematicians when constructing models and explaining the topology of proliferating epithelial tissue. If we extend the range of allowed shapes to exist in the population, were we assumed three sided cells could divide, it was shown that there is no difference between the shapes generated when mother cells divide according to the equal split scenario. If the cells were assumed to divide according to the binomial and uniform scenario, it was shown that the shapes of the distributions are still right skewed, and the cells are still mostly pentagonal which is inconsistent with experimental observations, where natural epithelial tissues display a hexagonal topology.

Lewis observed that there is an exponential relationship between the probability of cells dividing and the cell shapes in natural epithelia. We made a few modifications to our mathematical models. This involved altering our model assumptions by assuming cells divide with an exponential probability that depends on the number of dividing mother cell edges, making the model non-linear. The stabilised CEDs from the analytical nonlinear model were considerably more like the universal shape observed within the experiments of natural epithelia, displaying a hexagonal topology. We also discovered from the CA models that incorporating the exponential relationship between the sidedness of mother cells significantly improved the stabilised shapes of CEDs, again with mother cells appearing to divide according to the equal split scenario.

In conclusion, we have successfully improved our models where cells seem to divide approximately equally by incorporating the exponential relationship between sidedness of cells into our models. However, though there are still inconsistencies when cells are assumed to divide binomially and uniformly. There are strong spatial correlations between cells which must be incorporated into the model assumptions to implement the cellular mechanical events that occur, such as those events that occur at the junctional level at the apical side of cells. The vertex model by (Farhadifar, et al., 2007) considers the changes to the cell packing Geometry due to forces at the apical junctions. Our model does not consider rearrangements at the junctional level of the epithelium, where we have assumed only cell division influences the topological changes of a tissue over a discrete time. It would be interesting to investigate more mathematical models that consider the rearrangement of cells within the monolayer of epithelial tissues, allowing us to gain a better understanding of the dynamics of tissue development.

# References


Alberts, B., Johnson, A., Lewis, J., et al. (2002) 'Molecular Biology of the cell'. *Looking at the Structure of the cell*, 4th edn, New York: Garland Science.

Alemani, et al. (2012) 'Combining cellular automata and lattice Boltzmann method to model multiscale avascular tumor growth coupled with nutrient diffusion and immune competition'. *Journal of Immunological Methods*, 378 (1)(2), pp. 55-68. Available at https://doi.org/10.1016/j.jim.2011.11.009.

Besson, et al. (2015) 'Planar cell polarity breaks the symmetry of par protein distribution prior to mitosis in drosophila sensory organ precursor cells'. Curr Biol. 25 (8), pp.1104-1110.

Carlson, S.C. (2007) 'Topology'. *Encyclopaedia Britannica.*

Carter, R., Sánchez-Corrales, Y.E., Hartley, M., Grieneisen, V.A and Marée, A.F.M. (2017) 'Pavement cells and the topological puzzle'. *Development*., 144 (23), pp. 4386-4397. Available at https://dx.doi.org/10.1242%2Fdev.157073.

Cooper, G.M. (2000) 'Cell-Cell Interactions'. *The Cell: A Molecular Approach*, 2nd edn, Sunderland (MA): Sinauer Associates.

Delannay, R and Le Caër, G. (1994) 'Topological characteristics of 2D cellular structures generated by fragmentation'. *Phys. Rev. Lett.* 73 (11), pp. 1553-1556.

Farhadifar, R., Röper, J.C., Aigouy, B., Eaton, S and Jūlicher, F. (2007) 'The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing'. *Curr Biol*. (17), pp. 2095-2104.

Fehér, T., Papp, B., Pal, C and Pósfai, G. (2007) 'Systematic genome reductions: theoretical and experimental approaches'. *Chem. Rev.,* 107 (8), pp. 3498-3513.

Gibson, W.T and Gibson, M.C. (2006) 'Cell Topology, Geometry, and Morphogenesis in Proliferating Epithelia'. *Curr. Top. Dev. Biol.,* (89), pp. 87-114.

Gibson, M.C., Patel, A.B., Nagpal, R and Perrimon, N. (2006) 'The emergence of geometric order in proliferating metazoan epithelia'. *Nature*, 442 (7106), pp. 1038-1041.

Kachalo, et al. (2015) 'Mechanical model of geometric cell and topological algorithm for cell dynamics from single-cell to formation of monolayered tissues with pattern'. *PLoS one.* 10 (5): e0126484.
Available at: https://doi.org/10.1371/journal.pone.0126484.

Laskey, R.A., Alberts, B.M., Cooper, A., Lodish, F., Berntield, M.R., Stein, Wilfred, D., Cuffe, M, Chow, Christopher, Staehelin, L.A and Slack, J.M.W. (2021) 'Cell'. *Encyclopaedia Britannica.*

Le Bras, Stéphanie and Le Borgne, Roland. (2014) 'Epithelial cell division – multiplying without losing touch'. *Journal of Cell Science*. (127), pp. 5127-5137. Available at doi:10.1242/jcs.151472.

Lecuit, T and Lenne, P.F. (2007) 'Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis'. *Nat. Rev. Mol. Cell Biol.* (8), pp. 633-644. Available at http://dx.doi.org/10.1038/nrm2222.

Lewis, F.T. (1926) 'The effect of cell division on the shape and size of hexagonal cells'. *Anat. Recc,* 33 (5), pp. 331-355. Available at https://doi.org/10.1002/ar.1090330502.

Lewis, F.T. (1928) 'The correlation between cell division and the shapes and sizes od prismatic cells in the epidermis of cucumis'. *Anat. Recc*, 38 (3), pp. 341-376. Available at https://doi.org/10.1002/ar.1090380305.

Lynne, et al. (2011) 'Data Analysis Concepts and Observational Methods'. *Descriptive Physical Oceanography*, 6th edn, Amsterdam; Boston: Academic Press. pp. 147-186. Available at: https://doi.org/10.1016/B978-0-7506-4552-2.10006-X.

MATLAB (R2021a) *version 9.10.* The Mathworks Inc.

Molugaram, K and Rao, G.S. (2017) 'Curve Fitting'. *Statistical Techniques for Transportation Engineering*. pp. 281-292.
Available at: https://doi.org/10.1016/B978-0-12-811555-8.00005-2.

Patel, A.B., Gibson, W.T., Gibson, M.C and Nagpal, R. (2009) 'Modeling and Inferring Cleavage Patterns in Proliferating Epithelia'. *PLoS Computational Biology*. 5 (6): e1000412. Available at https://doi.org/10.1371/journal.pcbi.1000412.

Rodriguez-Boulan, Enrique and Macara, I.G. (2014) 'Organization and execution of the epithelial polarity programme'. *Nat Rev Mol Cell Biol*. 15 (4), pp. 225-242. Available at https://dx.doi.org/10.1038%2Fnrm3775.

Sahlin, P and Jönsson, H. (2010) 'A Modeling Study on How Cell Division Affects Properties of Epithelial Tissues Under Isotropic Growth.' *PLoS One*. 5 (7): e11750. Available at https://doi.org/10.1371/journal.pone.0011750.

Sandersius, S.A., Chuai, M., Weijer, C.J and Newman, T.J. (2011) 'Correlating cell behaviour with tissue topology in embryonic epithelia'. *PLoS ONE*. 6 (4): e18081. Available at: https://doi.org/10.1371/journal.pone.0018081.

Souza-e-Silva, et al. (2009) 'A Cellular Automata-Based Mathematical Model for Thymocyte Development'. *PLoS one.* 4 (12): e8233.
Available at https://doi.org/10.1371/journal.pone.0008233.

Ragkousi, K and Gibson, M.C. (2014) 'Cell division and the maintenance of epithelial order.' *J Cell Biol*. 207 (2), pp. 181-188. Available at https://dx.doi.org/10.1083%2Fjcb.201408044.

Thom, R. (1969) 'Topological models in Biology'. *Topology,* 8 (3), pp. 313-335.

Vrana, N.E., Lavalle, P., Dokmeci, M.R., Dehghani, F., Ghaemmaghami, A.M and Khademhosseini, A. (2013) 'Engineering functional epithelium for regenerative medicine and *in vitro* organ models: a review'. *Tissue Eng Part B Rev.* 19 (6), pp. 529-543.

Xin, Y and Karunarathna, M.C.M. (2018) 'Development of epithelial tissues: How are cleavage planes chosen?' *PLoS One.* 13 (11): e0205834. Available at https://doi.org/10.1371/journal.pone.0205834.

Yang, N, Ray, S.D and Krafts, K. (2014) 'Cell Proliferation'. *Encyclopedia of Toxicology*, 3 edt. pp. 761-765.

Yingzi, L., Naveed, H., Kachalo, S., Xu, L.X and Liang, J. (2012) 'Mechanisms of Regulating Cell Topology in Proliferating Epithelia: Impact of Division plane, Mechanical Forces, and Cell Memory'. *PLoS ONE.* 7 (8): e43108. Available at https://doi.org/10.1371/journal.pone.0043108.

Yokouchi, M., Atsugi, T., Logtestijn, M.V., Tanaka, R.J., Kajimura, M., Suematsu, M., Furuse, M., Amagai, M and Kubo, A. (2016) 'Epidermal cell turnover across tight junctions on Kelvin`s tetrakaidecahedron cell shape'. *eLife,* 5th edn. Available at https://dx.doi.org/10.7554%2FeLife19593.

# Appendices

All the graphs were generated from the computational software (Matlab, 2021).

## A.1 Appendix 1

The following code solves the system of equations corresponding to the three scenarios of cell division for the three to nine cell sided model.

```
%% 3 TO 9 SIDED CELL MODEL
%% Determine the eigensolutions for the matrices of the three division
scenarios.
clear all;
clc;
close all;
syms A B C V D W e
% Eigensolutions for uniform scenario.
% Define the Matrix A for the first scenario
A = [-3 2/3 1/2 2/5 1/3 2/7 0; 3 -10/3 1/2 2/5 1/3 2/7 1/3; 0 8/3 -7/2 2/5
1/3 2/7 1/3; 0 0 5/2 -18/5 1/3 2/7 1/3; 0 0 0 12/5 -11/3 2/7 1/3; 0 0 0 0
7/3 -26/7 1/3; 0 0 0 0 0 16/7 -5/3];

e = eig(A); % Column vector containing the eigenvalues of first matrix A.

[V,D] = eig(A);

% Eigensolutions for Binomial scenario.
% Define the Matrix B for the second scenario
B = [-3 1/2 1/4 1/8 1/16 1/32 0; 3 -3 3/4 1/2 5/16 3/16 1/16; 0 5/2 -13/4
3/4 5/8 15/32 5/16; 0 0 9/4 -7/2 5/8 5/8 5/8; 0 0 0 17/8 -59/16 15/32 5/8;
0 0 0 0 33/16 -61/16 5/16; 0 0 0 0 0 65/32 -31/16];

e = eig(B); % Column vector containing the eigenvalues of second matrix B.

[V_2,D_2] = eig(B);

% Eigensolutions for Equal split scenario.
% Define the Matrix C for the third scenario
C = [-3 0 0 0 0 0 0; 3 -2 1 0 0 0 0; 0 2 -3 2 1 0 0; 0 0 2 -4 1 2 1; 0 0 0
2 -4 0 1; 0 0 0 0 2 -4 0; 0 0 0 0 0 2 -2];

e = eig(C) % Column vector containing the eigenvalues of third matrix C.

[V_3,D_3] = eig(C)
```

## A.2 Appendix 2

The following code solves the system of equations corresponding to the three scenarios of cell division for the nonlinear analytical model.

```
function L=liverpool(S,C)
if S==1
   a4=1; a5=1; a6=1; a7=1; a8=1; a9=1; a10=1;
else
```

```
  a4=0.004233693;
a5=0.012209687;
a6=0.042432599;
a7=0.143233518;
a8=0.363681421;
a9=0.959153322;

end

P=[0 0 1 0 0 0]';
%   a4=a4/a9; a5=a5/a9; a6=a6/a9; a7=a7/a9; a8=a8/a9; a9=a9/a9;
for i=1:1:100000;
i;
  s8=1/(a4*P(1)+a5*P(2)+a6*P(3)+a7*P(4)+a8*P(5)+a9*P(6));
    s4=a4*s8;s5=a5*s8;s6=a6*s8;s7=a7*s8;s9=a9*s8;
%  s9=1/(a4*P(1)+a5*P(2)+a6*P(3)+a7*P(4)+a8*P(5)+a9*P(6));
%     s4=a4*s9;s5=a5*s9;s6=a6*s9;s7=a7*s9;s8=a8*s9;
 if(C==1)
%-------------------------------------------------------------------------
% display('Random Cleavage with the same probabilities')
%-------------------------------------------------------------------------
-

   A=[s4-3  s5   2/3*s6    2/4*s7    2/5*s8     2/6*s9;
       2    -3   2/3*s6    2/4*s7    2/5*s8     2/6*s9;
       0     2  -1/3*s6-3  2/4*s7    2/5*s8     2/6*s9;
       0     0   2        -2/4*s7-3  2/5*s8     2/6*s9;
       0     0   0           2      -3/5*s8-3   2/6*s9;
       0     0   0           0         2       -4/6*s9-1];

elseif(C==2)
%-------------------------------------------------------------------------
% display('Bionomial')
%-------------------------------------------------------------------------
   A=[s4-3  s5    2/4*s6    2/8*s7    2/16*s8     2/32*s9;
       2    -3    s6        6/8*s7    8/16*s8     10/32*s9;
       0     2   -1/2*s6-3  6/8*s7    12/16*s8    20/32*s9;
       0     0     2       -6/8*s7-3  8/16*s8    20/32*s9;
       0     0     0         2       -14/16*s8-3  10/32*s9;
       0     0     0         0         2          -30/32*s9-1];

elseif(C==3)
%-------------------------------------------------------------------------
% display('Equal Split(Middle)')
%-------------------------------------------------------------------------
A=[ s4-3    s5  0       0       0       0;
    2       -3  2*s6    s7      0       0;
    0       2   -s6-3   s7      2*s8    s9;
    0       0   2       -s7-3   0       s9;
    0       0   0       2       -s8-3   0;
    0       0   0       0       2       -s9-1];
else (C==4)
%-------------------------------------------------------------------------
% display('Random Cleavage with the Using the simulation')
%-------------------------------------------------------------------------
A=[s4-3    s5     0.248*s6     0.0313*s7    0.012*s8       0;
    2     -3       1.5*s6     0.9665*s7    0.314*s8    0.0621*s9;
    0      2      -0.748*s6-3 0.967*s7     1.346*s8    0.938*s9;
    0      0         2       -0.9687*s7-3  0.3144*s8   0.938*s9;
    0      0         0          2        -0.988*s8-3   0.0621*s9;
```

```
    0      0          0           0            2          -1*s9-1];
 end

B=0.01*A+eye(6);
P=B*P;
end
P;
tot=sum(P);
av_N=4*P(1)+5*P(2)+6*P(3)+7*P(4)+8*P(5)+9*P(6);
N=P/sum(P);
av_N=4*N(1)+5*N(2)+6*N(3)+7*N(4)+8*N(5)+9*N(6);
tot=sum(N);

T=table(s4,s5,s6,s7,s8,s9);
E=eig(A)
[V,D]=eig(A);
normalizeFromEIGEN=V(:,1)/sum(V(:,1))
N
T=table(s4,s5,s6,s7,s8,s9)
end
```

## A.3 Appendix 3

The following code is for the cellular automata models.

```
clear all;
MaxPopulation=5000;
MinimalCell=4;
MaximalCell=9;
population_size=10;
cells=randi([MinimalCell,MaximalCell],1,population_size); % 10 initial
cells
x = MinimalCell:1:MaximalCell;              % 6 categories - 4 to 9 number
of cell sides.
dd1=[];
 while (population_size<MaxPopulation)
mother=randi([1,population_size]);
mother_size=cells(mother);
% equal split
% if mod(mother_size,2)==0
%     daugthers=[mother_size/2+2,mother_size/2+2];
% else
%     daugthers=[(mother_size+1)/2+2,(mother_size-1)/2+2];
% end
%% Binomial Division
%Install Communications toolbox to make use of 'randsrc'
if mother_size==4 daugthers=[4,4]; end
if mother_size==5 daugthers=[4,5]; end
if mother_size==6
    dd=randsrc(1,1,[1,2,3;0.25,0.5,0.25]);
    if dd==1 daugthers=[4,6]; end %Probability of 1/4
    if dd==2 daugthers=[5,5]; end %Probability of 1/2
    if dd==3 daugthers=[6,4]; end %Probability of 1/4
%     dd1=[dd1,dd];
end
if mother_size==7
    dd=randsrc(1,1,[1,2,3,4;0.125,0.375,0.375,0.125]);
```

```
    if dd==1 daugthers=[4,7]; end %Probability of 1/8
    if dd==2 daugthers=[5,6]; end %Probability of 3/8
    if dd==3 daugthers=[6,5]; end %Probability of 3/8
    if dd==4 daugthers=[7,4]; end %Probability of 1/8
    %dd1=[dd1,dd];
end
if mother_size==8
    dd=randsrc(1,1,[1,2,3,4,5;0.0625,0.25,0.375,0.25,0.0625]);
    if dd==1 daugthers=[4,8]; end %Probability of 1/16
    if dd==2 daugthers=[5,7]; end %Probability of 1/4
    if dd==3 daugthers=[6,6]; end %Probability of 3/8
    if dd==4 daugthers=[7,5]; end %Probability of 1/4
    if dd==5 daugthers=[8,4]; end %Probability of 1/16
end
if mother_size==9

dd=randsrc(1,1,[1,2,3,4,5,6;0.03125,0.15625,0.3125,0.3125,0.15625,0.03125])
;
    if dd==1 daugthers=[4,9]; end %Probability of 1/32
    if dd==2 daugthers=[5,8]; end %Probability of 5/32
    if dd==3 daugthers=[6,7]; end %Probability of 5/16
    if dd==4 daugthers=[7,6]; end %Probability of 5/16
    if dd==5 daugthers=[8,5]; end %Probability of 5/32
    if dd==6 daugthers=[9,4]; end %Probability of 1/32
end
% random division
 dd=randi([1,mother_size-3]);
    if dd==1 daugthers=[4,mother_size]; end
    if dd==2 daugthers=[5,mother_size-1]; end
    if dd==3 daugthers=[6,mother_size-2]; end
    if dd==4 daugthers=[7,mother_size-3]; end
    if dd==5 daugthers=[8,mother_size-4]; end
    if dd==6 daugthers=[9,mother_size-5]; end
 cells(mother)=[];
neighbours=randi([1,population_size-1],1,2);
while (neighbours(2) == neighbours(1))
    neighbours(2)=randi([1,population_size]);
end
%cells
a1=neighbours(1); if cells(a1)<MaximalCell cells(a1)=cells(a1)+1; end
%c1=cells(a1)
a2=neighbours(2); if cells(a2)<MaximalCell cells(a2)=cells(a2)+1; end
%c2=cells(a2)
cells=[cells,daugthers];
population_size=length(cells);
% Drawing
if rem(population_size, 10) == 0
    freq=zeros(1,6);
    for i=1:1:population_size
        freq(cells(i)-3)=freq(cells(i)-3)+1;
    end
    y = freq./population_size;
    sum=0;
    for i=1:1:6
        sum=sum+y(i);
    end
%    bar(x,y)
%    drawnow
end
 end
```

```
%CA model including exponential dependence of sides
clear all;
MaxPopulation=5000;
MinimalCell=4;
MaximalCell=9;
population_size=10;
pd=[0.,0.0018,0.0149,0.0589,0.2062,0.7182]; % 10 initial cells
x = MinimalCell:1:MaximalCell;              % 6 categories - 4 to 9 number
of cell sides.
dd1=[];
 while (population_size<MaxPopulation)
mother=randi([1,population_size]);
mother_size=cells(mother);
if rand<pd(mother_size-3)
% equal split
if mod(mother_size,2)==0
    daugthers=[mother_size/2+2,mother_size/2+2];
else
    daugthers=[(mother_size+1)/2+2,(mother_size-1)/2+2];
end
%% Binomial Division
%Install Communications toolbox to make use of 'randsrc'
% if mother_size==4 daugthers=[4,4]; end
% if mother_size==5 daugthers=[4,5]; end
% if mother_size==6
%     dd=randsrc(1,1,[1,2,3;0.25,0.5,0.25]);
%     if dd==1 daugthers=[4,6]; end %Probability of 1/4
%     if dd==2 daugthers=[5,5]; end %Probability of 1/2
%     if dd==3 daugthers=[6,4]; end %Probability of 1/4
% %     dd1=[dd1,dd];
% end
% if mother_size==7
%     dd=randsrc(1,1,[1,2,3,4;0.125,0.375,0.375,0.125]);
%     if dd==1 daugthers=[4,7]; end %Probability of 1/8
%     if dd==2 daugthers=[5,6]; end %Probability of 3/8
%     if dd==3 daugthers=[6,5]; end %Probability of 3/8
%     if dd==4 daugthers=[7,4]; end %Probability of 1/8
%     %dd1=[dd1,dd];
% end
% if mother_size==8
%     dd=randsrc(1,1,[1,2,3,4,5;0.0625,0.25,0.375,0.25,0.0625]);
%     if dd==1 daugthers=[4,8]; end %Probability of 1/16
%     if dd==2 daugthers=[5,7]; end %Probability of 1/4
%     if dd==3 daugthers=[6,6]; end %Probability of 3/8
%     if dd==4 daugthers=[7,5]; end %Probability of 1/4
%     if dd==5 daugthers=[8,4]; end %Probability of 1/16
% end
% if mother_size==9
%
dd=randsrc(1,1,[1,2,3,4,5,6;0.03125,0.15625,0.3125,0.3125,0.15625,0.03125])
;
%     if dd==1 daugthers=[4,9]; end %Probability of 1/32
%     if dd==2 daugthers=[5,8]; end %Probability of 5/32
%     if dd==3 daugthers=[6,7]; end %Probability of 5/16
%     if dd==4 daugthers=[7,6]; end %Probability of 5/16
%     if dd==5 daugthers=[8,5]; end %Probability of 5/32
%     if dd==6 daugthers=[9,4]; end %Probability of 1/32
% end
% random division
% if mother_size==4 daugthers=[4,4]; end
% if mother_size==5 daugthers=[4,5]; end
```

```
% if mother_size==6
%     dd=randi([1,mother_size-3])
%     if dd==1 daugthers=[4,6]; end
%     if dd==2 daugthers=[5,5]; end
%     if dd==3 daugthers=[6,4]; end
% end
% if mother_size==7
%     dd=randi([1,mother_size-3])
%     if dd==1 daugthers=[4,7]; end
%     if dd==2 daugthers=[5,6]; end
%     if dd==3 daugthers=[6,5]; end
%     if dd==4 daugthers=[7,4]; end
% end
% if mother_size==8
%     dd=randi([1,mother_size-3])
%     if dd==1 daugthers=[4,8]; end
%     if dd==2 daugthers=[5,7]; end
%     if dd==3 daugthers=[6,6]; end
%     if dd==4 daugthers=[7,5]; end
%     if dd==5 daugthers=[8,4]; end
% end
% if mother_size==9
%     dd=randi([1,mother_size-3])
%     if dd==1 daugthers=[4,9]; end
%     if dd==2 daugthers=[5,8]; end
%     if dd==3 daugthers=[6,7]; end
%     if dd==4 daugthers=[7,6]; end
%     if dd==5 daugthers=[8,5]; end
%     if dd==6 daugthers=[9,4]; end
% end
cells(mother)=[];
neighbours=randi([1,population_size-1],1,2);
while (neighbours(2) == neighbours(1))
    neighbours(2)=randi([1,population_size]);
end
%cells
a1=neighbours(1); if cells(a1)<MaximalCell cells(a1)=cells(a1)+1; end
%c1=cells(a1)
a2=neighbours(2); if cells(a2)<MaximalCell cells(a2)=cells(a2)+1; end
%c2=cells(a2)
cells=[cells,daugthers];
population_size=length(cells);
end
 end
% Drawing
if rem(population_size, 10) == 0
    freq=zeros(1,6);
    for i=1:1:population_size
        freq(cells(i)-3)=freq(cells(i)-3)+1;
    end
    y = freq./population_size;
    sum=0;
    for i=1:1:6
        sum=sum+y(i);
    end
%    bar(x,y)
%    drawnow
end
%% Plot each CEDH for three scenarios
%% Plot CEDH for Uniform scenario
% bar(x,y)
```

```
% title('CED for Uniform scenario');
% xlabel('Number of cell sides');
% ylabel('Frequency of cells');

%% Plot CEDH for Binomial scenario
% bar(x,y)
% title('CEDH for Binomial scenario');
% xlabel('Number of cell sides');
% ylabel('Frequency of cells');

% Plot CEDH for Equal Split scenario
bar(x,y)
title('CED for Equal split scenario');
xlabel('Number of cell sides');
ylabel('Frequency of cells');
```