INAUGURAL – DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht-Karls Universität
Heidelberg

vorgelegt von
M.Sc. Diana-Patricia DANCIU
aus Craiova, Rumänien

Tag der mündlichen Prüfung: .....

Mathematical Modelling of Stem Cell Dynamics during Post-embryonic Organ Growth

Gutachter: Professor Dr. Anna Marciniak-Czochra

.....

Dedicated to my grandfather, S.

Abstract

The current work is devoted to the mathematical modelling of the development of fish respiratory organs, called gills or branchiae. The model organism chosen for the task is the Japanese rice fish (*Oryzias latipes*), more colloquially known as medaka. Their gills are analysed in the attempt to answer three main developmental questions *via* mathematical modelling, with possible applications beyond the scope of this thesis. Firstly, how many stem cells are needed to build the organ? What kind of heterogeneities exist among these stem cells? And, finally, what properties and relations with each-other do these stem cells have, that give the organ its shape?

Relying on experimental data from our collaborators in the group of Prof. Lazaro Centanin, Centre for Organismal Studies, Heidelberg University, we use a variety of methods to study the aforementioned aspects. These methods were selected, adapted and developed based on the goal of each project and on the available data. Thus, a combination of stochastic and deterministic techniques are employed throughout the thesis, including Gillespie-type simulations, Markov chains theory and compartmental models.

The study of stem cell numbers and heterogeneities is approached *via* stochastic simulations extended from the algorithm of Gillespie, and further improved by Markov chains methods. Results suggest that not only very few stem cells are sufficient to build and maintain the organ but, more importantly, these stem cells are heterogeneous in their division behaviour. In particular, they rely on alternating activation and quiescence phases, such that once a stem cell has divided, it becomes activated and divides multiple times before allowing another one to take the lead.

For the study of growth and shape of gills, multiple deterministic models based on different assumptions and investigating various hypotheses have been developed. All these models have a compartmental structure, with increasing number of compartments governed by indicator functions which, in turn, depend on explicit or implicit algebraic equations. For each model, the existence, uniqueness and non-negativity of solutions are proved, the analytical solutions are found and their regularity is discussed. The models are compared based on their ability to reproduce part of the data, and the best one is selected. The chosen model is then applied to further data and speculations on hypotheses supporting the model are made. Results suggest that the main stem cell types, responsible for growing the organ, slow down their proliferation in time, either due to ageing or to the lack of sufficient nutrients.

The main results and strengths of this thesis consist of the high variety of models developed and methods employed, their capability to answer important biological questions and, even more, to uncover new insights on mechanisms previously unknown.

Heidelberg, May 2019 *Diana-Patricia Danciu*

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der mathematischen Modellierung der Entwicklung von Atmungsorganen der Fische, den sogenannten Kiemen. Der für die Aufgabe gewählte Modellorganismus ist der japanische Reisfisch (*Oryzias latipes*), umgangssprachlich Medaka genannt. Mit dem Versuch, drei Hauptentwicklungsfragen mittels mathematischer Modellierung zu beantworten, werden ihre Kiemen analysiert, wobei mögliche Anwendungen über den Rahmen dieser Arbeit hinausgehen. Erstens, wie viele Stammzellen werden zum Aufbau des Organs benötigt? Welche Art von Heterogenitäten gibt es unter diesen? Und schlielich, welche Eigenschaften und Beziehungen zueinander haben diese Stammzellen, die dem Organ seine Form geben?

Basierend auf experimentellen Daten unserer Mitarbeiter in der Gruppe von Prof. Lazaro Centanin verwenden wir eine Vielzahl von Methoden, um die oben genannten Aspekte zu studieren, Methoden auszuwählen, anzupassen und basierend auf dem Ziel jedes Projekts und den verfügbaren Daten zu entwickeln. Daher wird in der gesamten Arbeit eine Kombination aus stochastischen und deterministischen Techniken eingesetzt, zu denen Gillespie-Simulationen, die Markov-Ketten-Theorie und Kompartimentmodelle gehören.

Die Untersuchung von Stammzellzahlen und -heterogenitäten wird durch stochastische Simulationen erreicht, die auf dem Algorithmus von Gillespie beruhen und durch Markov-Kettenmethoden weiter verbessert werden. Die Ergebnisse deuten darauf hin, dass nicht nur sehr wenige Stammzellen ausreichen, um das Organ aufzubauen und zu erhalten. Wichtiger jedoch ist, dass diese Stammzellen in ihrem Teilungsverhalten heterogen sind. Sie sind insbesondere auf abwechselnde Aktivierungs- und Ruhephasen

angewiesen, so dass eine Stammzelle, sobald sie sich geteilt hat, aktiviert wird und sich mehrmals teilt, bevor eine andere die Führung übernehmen kann.

Für die Untersuchung des Wachstums und der Form von Kiemen wurden mehrere deterministische Modelle entwickelt, die auf unterschiedlichen Annahmen basieren und verschiedene Hypothesen untersuchen. Alle diese Modelle haben eine Kompartimentstruktur mit einer zunehmenden Anzahl von Kompartimenten, die durch Indikatorfunktionen gesteuert werden, die wiederum von expliziten oder impliziten algebraischen Gleichungen abhängen. Für jedes Modell werden Existenz, Eindeutigkeit und Nicht-Negativität von Lösungen nachgewiesen, die analytischen Lösungen werden gefunden und ihre Regularität wird diskutiert. Die Modelle werden basierend auf ihrer Fähigkeit einen Teil der Daten zu reproduzieren verglichen, und das beste wird entsprechend ausgewählt. Das gewählte Modell wird auf weitere Daten angewendet und es werden Spekulationen zu Hypothesen aufgestellt, die das Modell unterstützen. Die Ergebnisse legen nahe, dass beide Hauptstammzelltypen, die für das Wachstum des Organs verantwortlich sind, ihre Proliferation mit der Zeit verlangsamen, entweder aufgrund von Alterung oder Nahrungsmangel.

Die wichtigsten Ergebnisse und Stärken dieser Arbeit bestehen in der großen Vielfalt der entwickelten Modelle und Methoden, ihrer Fähigkeit, wichtige biologische Fragen zu beantworten und noch mehr neue Erkenntnisse über bisher unbekannte Mechanismen zu gewinnen.

Heidelberg, Mai 2019 *Diana-Patricia Danciu*

Rezumat

Prezenta lucrare este dedicată modelării matematice a dezvoltării organelor respiratorii la peşti, numite branhii. Organismul ales ca model pentru aceast studiu este peştele de orez japonez (*Oryzias latipes*), cunoscut mai simplu sub numele de medaka. Acest studiu efectuat pe branhiile peştelui medaka îşi propune obţinerea prin intermediul modelării matematice a unor răspunsuri pentru trei ntrebări de bazua din biologie, cu posibile aplicaţii dincolo de scopul acestei teze. O primă întrebare se referă la numărul celulelor stem necesare pentru construirea organului. Apoi, ne interesează un răspuns privind tipul eterogenităţilor existente între aceste celule stem. Şi, în cele din urmă, este important să aflăm ce proprietăţi şi relaţii dintre aceste celule stem conduc la forma specială a acestui organ, branhia.

Bazându-ne pe datele experimentale ale colaboratorilor noştri din cadrul grupului Prof. Lazaro Centanin, folosim o serie de metode pentru a studia aspectele menţionate mai sus, metode selectate, adaptate şi dezvoltate în funcţie de scopul fiecărui proiect şi de datele disponibile. Astfel, pe parcursul studiului, şi al prezentei teze de doctorat, s-au combinat o serie de tehnici stocastice şi deterministe, tehnici care includ simulări de tip Gillespie, teoria lanţurilor Markov şi modele compartimentale.

Studiul numărului de celule stem şi al eterogenităţilor lor este abordat prin simulări stocastice, extinse de la algoritmul lui Gillespie, şi îmbunătăţite ulterior prin metodele lanţurilor Markov. Rezultatele obţinute sugerează nu numai că, pentru construirea organului este suficient un număr mic de celule stem, ci şi mai important, că aceste celule stem sunt eterogene în ceea ce priveşte comportamentul lor de divizare. Mai exact, s-a observat că divizarea celulelor stem se bazează pe faze alternante de activare şi repaus, astfel încât,

odată ce o celulă s-a divizat, aceasta devine activ şi se divide de mai multe ori, înainte de a permite altei celule să preia conducerea.

Pentru studiul dezvoltării şi formei branhiilor, au fost create mai multe modele deterministe bazate pe reguli diferite şi investigând diverse ipoteze. Toate aceste modele au o structură compartimentată, cu un număr de compartimente care creşte în timp, creştere guvernată de funcţii indicator care, la rândul lor, depind de ecuaţii algebrice explicite sau implicite. Pentru fiecare model, se demonstrează existenţa, unicitatea şi non-negativitatea soluţiilor, se calculează soluţiile analitice şi se discută regularitatea lor. Modelele sunt comparate pe baza capacităţii lor de a reproduce o parte din date, cel mai bun dintre aceste modele fiind în final cel selectat. Modelul ales este evaluat pe date suplimentare şi se fac apoi speculaţii privind ipotezele care susţin modelul. Rezultatele obţinute în studiu sugerează că ambele tipuri principale de celule stem, responsabile de creşterea organului, îşi încetinesc proliferarea în timp, fie din cauza îmbătrânirii, fie a lipsei de factori nutritivi suficienţi.

Principalele rezultate şi puncte forte ale acestei teze constau în diversitatea de modele create şi de metode folosite, capacitatea lor de a răspunde la întrebări cheie din biologie şi, mai mult, de a descoperi noi mecanisme, necunoscute anterior.

Heidelberg, Mai 2019 *Diana-Patricia Danciu*

Contents

Chapter 1
Introduction

"The imagination of nature is far, far greater than the imagination of man."

- Richard Feynman

Mathematical modelling has been playing an increasingly important role in the understanding of processes from many research fields, from natural sciences to engineering and to social sciences. In particular, in biological and medical sciences, many mathematical models have been developed to study development [1, 2] and disease [3]. Not only does mathematical modelling help in uncovering new insights on the biological systems under study, but biology drives the development of new mathematical techniques needed to understand such systems [4, 5, 6, 7, 8].

This work applies mathematical methods in the field of developmental biology, more specifically in stem cell research, and studies the stem cell dynamics during post-embryonic organ growth in fish. Developmental Biology belongs to the fundamental research category, dealing with topics such as organogenesis or stem cell differentiation and studies in this field make the basis for further applied investigations such as cancer, ageing or other diseases. Mathematical models to study various aspects in developmental biology are constructed, by applying mathematical methods to data from the respiratory organ of fish, coming from experiments on the model organism Medaka (*Oryzias latipes*) [9]. Aspects such as the properties of and relations between various stem cells are investigated, and determining the numbers and functional heterogeneities among these stem cells is aimed for. To our knowledge, there have been no previous mathematical models for the study of gill development.

The thesis employs both stochastic and deterministic methods adapted to the specific investigations of two related projects and to the data available. For the first project, a stochastic algorithm is created and implemented to simulate the biological system, and

subsequently Markov chains theory is used to improve its efficiency. The second project develops various compartmental models with growing domain, all consisting of a system of ordinary differential equations coupled to indicator functions governing the addition of new compartments, which depend on inputs from two algebraic equations. These models are analysed and compared *via* model selection methods. Finally, conclusions are drawn based on insights gained from the modelling, and suggestions regarding the plausibility of different biological hypotheses are made.

1.1 Overview of methods and results

Medaka, also known as the Japanese rice fish, recommends itself as a suitable model organism to study aspects in developmental biology, due to its rapid development, body transparency during juvenile stages, and isogenic property. The organ placed under the microscope is the branchia (also known as gill) which, because of its modular structure, is perfectly suited for our studies. As can be seen in Fig. 1.1, the gills have a stereotypic spatio-temporal organisation and are composed of four pairs of double-sided branchial arches, which in turn consist of a sequence of filaments [10, 11, 12]. Branchial arches grow along two orthogonal axes, longitudinally by the elongation of filaments, and transversally by addition of new ones at each of the two extremes.

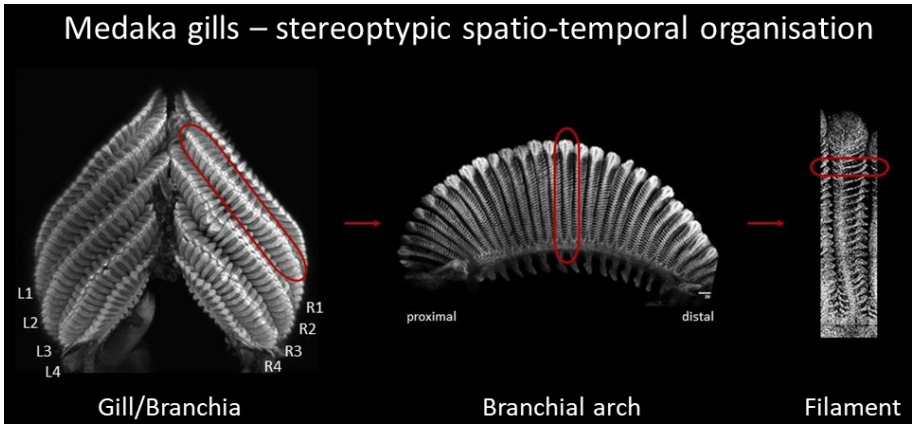


Fig. 1.1: Structure of a gill: one branchia (left) is composed of eight branchial arches (middle), each of which is built from a sequence of filaments (right).

1.1.1 Stem cell numbers and functional heterogeneities

By taking advantage of this organised growth "protocol", stochastic modelling can shed light on the numbers of stem cells needed for building these organs, as well as on the heterogeneities existing among them. Stochastic modelling, and in particular methods such as Markov chains and Gillespie-type algorithms, have been often employed in mathematical biology [8, 13, 14, 15, 16, 17, 18, 19] and prove highly useful in our studies, as well. The clonal data used in this (first) project record the labelling status (labelled or unlabelled) of filaments from multiple branchial arches from various fish [20, 21].

Hence, by developing an adapted Gillespie-type algorithm one can simulate the growth process and produce an *in silico* data set similar to the experimental one. These two can then be compared *via* an objective function describing the "distance" between data and simulation results and the desired parameters can be estimated by minimising this function. The computational method presented is approximate, very time-consuming and demands a high computational effort, but is quite easy to understand.

A further improvement of the previous modelling approach uses Markov chains theory. A two-state Markov chain is defined for describing transitions between a labelled and an unlabelled filament, transitions representing successions in the sequence of such filaments within a branchial arch. By computing the entries of the specific transition probability matrix, one can obtain exact probabilities of observing events of filament distributions similar to those existing in the experimental data. Hence, this method is exact and very efficient, but is quite abstract and thus more difficult to grasp and develop without certain comprehension gained from the previous approach.

The main insight resulted from the mathematical modelling of the system is related to the functional heterogeneities among the stem cells driving the growth of branchial arches, fact which was previously unknown. The results of this project show that the division process is not entirely random: we find that once a stem cell has divided to generate a filament, it has a much higher chance to divide again in the detriment of another niche resident. This behaviour corresponds to a scenario of activation and quiescence phases: when a stem cell divides, it becomes active and subsequently divides multiple times before another one takes over the task of filament generation. The so-called *probability of division* is found to have a surprisingly high value, of approximately *0.9*. In addition, sim-

ulation results suggest that only a small number of functionally heterogeneous active stem cells is, in fact, sufficient for building and maintaining the organ: *2-3* stem cells per niche.

A more detailed description of the results related to the stem cell numbers and their functional heterogeneities can be found in Chapter 7, which concludes the first part of the thesis.

1.1.2 Modelling of growth and shape

Continuing the study of post-embryonic gill growth, a second project uses deterministic modelling to describe the organ development, and looks at the shapes of branchial arches in order to understand the behaviour of the stem cells driving their growth. Deterministic modelling, in the form of ordinary and partial differential equations, have had a rich history within the field of mathematical biology, in the study of development [1, 22, 2, 23] and disease [3, 24, 25, 26, 27]. For this project, data recording lengths of filaments have been gathered from multiple branchial arches of various fish and ordinary differential equations models have been employed.

Five different mathematical models have been developed, all having similar structures but being based on different assumptions. The models are inspired by previously developed compartmental models [28, 29] and consist of a system of ordinary differential equations, each describing the behaviour within each compartment representing one filament. Considering that not only do filaments elongate, but also new ones are added, a method to increase the discrete number of compartments was needed. The solution came in the form of indicator functions, their "choice" being governed by two algebraic equations describing the proliferation of branchial arch stem cells. The relative novelty of these models comes from the method of approaching the discrete domain growth. Usually, the modelling of systems with growing domain is approached *via* partial differential equations [4, 30, 31] or individual-based models [8, 32, 33, 34], but the specificity of our biological system is given by the interplay between continuous and discrete mechanisms. Few instances of similar methods can be found [22, 35].

Each model is analysed mathematically, proving the existence, uniqueness and non-negativity of the solutions, and their analytical solution is found. Subsequently, the models

are compared by fitting to data from one representative branchial arch and the best one is selected according to model selection criteria. The selected model is then applied to data from various other branchial arches and its capability to reproduce the shape of each such arch is established.

Based on the insights gained from the results of the model, we can state that filament stem cells reduce their proliferation with their length, for example due to the increasing time needed for sufficient nutrients to travel to their location. In addition, branchial arch stem cells also increase their cell cycle duration but, in contrast, we can speculate that this does not happen due to nutrients transport but, for example, due to ageing. The models thus prove that the branchial arches *slow down* their growth along *both* orthogonal axes.

The detailed overview of the results related to the branchia growth and shape can be found in Chapter 12, which concludes the second part of the thesis.

1.2 Outline of the Thesis

The thesis starts with an overview of the mathematical methods used for the investigations performed (**Chapter 2**), namely short descriptions of stochastic algorithms, Markov chains theory and differential equations properties. In addition, parameter estimation techniques and model comparison methods are discussed.

Chapter 3 follows with an introduction into the biological background used in this work, in particular properties of fish model organisms and their post-embryonic growth, as well as an in detail description of the organ under study - the branchia. Subsequently, the thesis is divided into two parts, describing two separate but related projects, which study various aspects of post-embryonic growth of medaka gills.

The **first part** (**Chapters 4-7**) presents the study of the number of stem cells needed to build the respiratory organ of fish and their heterogeneities. The aim of this project is to understand how many stem cells are necessary to grow the fish gill and to maintain it throughout the life of the fish, and what kind of properties these stem cells possess.

In particular, **Chapter 4** describes the clonal data used in this thesis, the experimental procedures and the transformations performed on these data, before the mathematical modelling. Following lineage tracing experiments, recorded is the labelling status (la-

belled or unlabelled) of certain filaments within the organ for four different fate-restricted stem cell types.

Chapter 5 starts the modelling within this part of the thesis by considering two simplified homogeneous models for two extreme scenarios. This analysis is performed in order to study whether the specific cells under study are indeed stem cells and not progenitor cells, instead, and the mathematical insights are able to support our hypothesis. In addition, these investigations suggest that functional heterogeneities resulting from their division behaviour exist among the stem cells. These properties come from the dynamics of stem cells, which have activation and quiescence phases thus leading to a setting in which cells take turns in being activated and continue dividing until another sibling becomes active and takes the lead in this task. The results of this chapter, together with a thorough description of the biological procedures have been published [21].

Chapter 6 develops the model by incorporating the newly gain insights, through the introduction of a new heterogeneity parameter in the form of a probability describing the chances of a stem cell to divide again after it has just divided. Here two different approaches are employed: the Stochastic Simulations approach (**Section 6.2**) and the Markov Chains approach (**Section 6.3**). In the first case, an algorithm for stochastically simulating large data sets of *in silico* branchial arches is developed and these are compared to the experimental data in order to estimate the desired parameters: the number of stem cells and their probability of division. The second method studies the same aspects in a more abstract way, by making use of Markov chains theory to circumvent the need for producing such big simulated data sets. This method is not only more efficient in terms of computational time and effort, but is also exact. The two approaches provide agreeing results, as desired, and these are presented together with the parameter estimation procedure in **Section 6.4**.

Finally, **Chapter 7** summarises and concludes the **first** part, while also providing an outlook on possible applications.

The **second part** (**Chapters 8-12**) of the thesis presents the study of organ growth and, in particular, of properties and relations between different stem cell types, by analysing the shapes of branchial arches. These gill structures grow on two orthogonal axes (longitudinal - driven by filament stem cells, and transversal - driven by branchial arch stem

cells) and have a curved, asymmetric profile, the study of which can shed light onto the dynamics of stem cells driving growth.

Chapter 8 begins the part with a description of the experimental procedure and the data collected. These record lengths of filaments of various branchial arches from multiple fish.

Starting in **Chapter 9**, five models developed for our investigations are presented. These models consist of a system of ordinary differential equations coupled to algebraic equations *via* specific indicator functions. They are all based on the same structure, which can thus be written in generalised form, but are outlined in the order of growing complexity and their assumptions are summarized:

- ◇ The *Linear Growth Model* assumes constant proliferation of both stem cell types driving growth of branchial arches along the two orthogonal axes.
- ◇ The *Elongation-Decay Model* considers a gradual decrease in the proliferation of filament stem cells, from here on referred to as a decay in growth or proliferation. This decay is governed by an exponential function with negative exponent, dependent on the length of filaments. The branchial arch stem cells are still assumed to divide at constant rates.
- ◇ The *Elongation-Generation-Decay Model* goes on step further and introduces a decay in the proliferation of arch stem cells, as well, in addition to the previously considered one. This second decay is time-dependent and is modelled *via* a negative feedback function. Thus the growth decreases along both axes.
- ◇ The *Alternative-Elongation-Generation-Decay Model* represents a variant of the previous model, in which the exponential function describing the filament growth decay is replaced by a negative feedback, similar to that for the arch stem cells, which is still dependent on filament length. The arch stem cells behave as in the previous model.
- ◇ Finally, the *Space-dependent (Alternative-) Elongation-Generation-Decay Model* keeps the assumptions on filament stem cells from the previous model, but switches from a time-dependent to a space-dependent decay in the proliferation of arch stem cells. Namely, this decay is also modelled by a negative feedback but it depends on the total number of filaments (equivalently, on the "distance" between the two niches), instead of time.

In **Chapter 10**, the two most general models are analysed mathematically and proofs for the existence, uniqueness and non-negativity of solutions are presented. In addition, the analytical solution is determined for each model and its regularity is discussed.

Chapter 11 applies the models developed to the experimental data and compares them *via* model selection methods to determine the best out of them. First, the transformations made to the data and the numerics behind these are explained. The five models are then compared by applying each to data from one representative branchial arch and by computing their Akaike Information Criterion for model selection. The results of each model are displayed and the *Alternative-Elongation-Generation-Decay Model* is selected as the best one out of the five. To end the chapter, the selected model is further applied to data from other branchial arches and its capabilities to reproduce any experimental shape are shown. Here it is speculated that stem cells reduce their division due to age, and against the nutrient consumption hypothesis.

Finally, **Chapter 12** summarizes the results of the **second** part and presents possible applications that go beyond the biological model.

To conclude, **Chapter 13** presents a summary of the thesis, reviews the main findings of the work and ends with an outline of further research envisioned for the near-future.

Chapter 2
Mathematical Background

"All models are wrong, but some are useful."
- George E. P. Box

This chapter provides an overview of various mathematical methods needed for the projects presented in the thesis.

Mathematical modelling is a powerful tool for describing biological systems, and once the model is tested and able to reproduce the data at hand, after an iterative cycle of refinement, it can be used for making predictions and drawing conclusions related to the biological mechanisms under study [4, 32]. There exist many methods of describing a system but it is important to take into account a series of aspects such as system dimension and variability.

The two main complementary approaches to a biological system are stochastic and deterministic techniques [36]. One of the methods used in the case of a system with a small population size, for example few cells interacting or few molecules reacting, is the Gillespie Algorithm. The algorithm relies on stochastic simulations which are able to take into account the variability within the behaviour of its components. On the other hand, deterministic modelling *via* systems of ordinary or partial differential equations, possibly coupled with algebraic equations, is more appropriate when the population size in the biological system is large, as it is the case, for example, of the hematopoetic system ($N = 10^8$ [37]), because everything is modelled as a bulk and each cell or interaction is not described individually. In this case, numerical simulations of the model are faster and every time the result is the same for a certain set of parameters and initial conditions. In general, if the same biological system can be described by both stochastic and deterministic models, the average of multiple stochastic simulations converges to the deterministic solution but this is not always the case. A scenario in which the deterministic solution

does not provide the average of multiple stochastic simulations, is when the system has multiple stable equilibrium states, and the stochastic fluctuations can drive the solution from one stable state to another.

2.1 Stochastic Simulations

Stochastic simulations record the evolution of variables which change randomly with certain probabilities. Initially, such stochastic simulations using the so-called Gillespie Algorithm were first introduced by Daniel Gillespie for simulating chemical or biochemical reaction systems [13, 38]. The general steps of such an algorithm for chemical reactions may be outlined as follows:

1. **Initialize the system:** Number of molecules, reaction constants and random number generator.
2. **Select reaction:** This is a Monte-Carlo step - based on randomly generated numbers one can choose which reaction will occur next and the time interval before this happens.
3. **Execute reaction and update data:** Update the molecule count based on the reaction that occurred and increase the time span by the time interval from point 2.
4. **Iterate:** Terminate the algorithm or repeat steps 2-3.

Over the years, several adaptations of the Gillespie Algorithm have been developed, mainly in order to improve efficiency for large molecule numbers, or suitable for different systems. A few of these methods are shortly described below.

- ◇ The **Direct Method** is the one originally introduced by Gillespie in 1976 [38]. This method computes a reaction probability by multiplying the rate constant of each reaction with the concentration of its substrates.Then two random numbers are generated: one for choosing the reaction to occur based on the relative probabilities of all reactions, and one for the execution time of the chosen reaction. Since the method uses two random numbers the complexity is linear in time and storage with the number of reaction types, hence making it computationally expensive for large systems.

- ◇ The **First Reaction Method** was also introduced by Gillespie [38] and is a more intuitive way of implementing step 2 by computing a "tentative time" τ_i for each reaction to occur and then selecting the reaction with the smallest τ_i. However this requires one random variable generated per possible reaction thus making it very inefficient.

Other similar methods exist, some examples being the **Next Reaction Method** [39] or the **Random Substrate Selection** [40]; they are, however, not discussed in this thesis.

2.2 Markov Chains

The following overview on Markov Chains theory [41, 42] is selected and adapted from the lecture notes for the Markov Chains second-year course (2011), by Prof. Richard Webber, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge.

Let I be a countable set, $I = \{i, j, k, ...\}$, where each $i \in I$ is a state and the set I is called state-space. In addition, a probability space $(\Omega, \mathscr{F}, P)$ is defined, where Ω is a set of outcomes, $\mathscr{F}$ is a set of subsets of Ω and for $A \in \mathscr{F}$, $P(A)$ is the probability of A. Further, a row vector $\lambda = (\lambda_i, i \in I)$ is called a measure if $\lambda_i \geq 0$, for all $i \in I$. When $\Sigma_{i \in I} \lambda_i = 1$, it is called a distribution or a probability measure.

A transition probability matrix (or transition matrix) is defined by $P = (p_{ij} : i, j \in I)$ with $p_{ij} \geq 0$. This is a stochastic matrix having the property that $\Sigma_{j \in I} p_{ij} = 1$, i.e. each row of P is a distribution over I.

Definition 2.2.1 (Markov Chain) *We call a sequence of random variables $(X_n)_{n \geq 0}$ a Markov chain with initial distribution λ and transition matrix P, Markov(λ, P), if for all $n \geq 0$ and $i_0, \ldots, i_{n+1} \in I$ the following hold:*

- ◇ $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$
- ◇ $\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}$

This means that the next event only depends on the current one, and not on the entire history of events up to this point, i.e. the chain is memoryless. Another equivalent relation states that

$$\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdot \ldots \cdot p_{i_{n-1} i_n} \tag{2.1}$$

Within each Markov Chain, a class is a set of states within which we can always return to the state from which we start. So if we are in state i at some point, and transitions $i \leftrightarrow j$ and $j \leftrightarrow k$ can happen, it also means that $i \leftrightarrow i$, and thus i, j, k are part of the same class. We call a closed class a set of states $C \subset I$ such that

$$i \in C,\ i \to j \Rightarrow j \in C,$$

i.e. a class from which there is no escape. Further, a state i is an absorbing state if $\{i\}$ is a closed class.

Definition 2.2.2 (Irreducibility) *A Markov Chain in which I is a single class is called irreducible. In other words, a Markov Chain is irreducible if it is possible to get to any state from any state, i.e. $i \to j$ for every $i, j \in I$.*

Definition 2.2.3 (Hitting probabilities and mean hitting times) *Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. We define by hitting time of state i (or subset $A \subset I$) the first instance when we move to i (or A). This is a random variable $H^A : \Omega \to \mathbb{N}$ given by*

$$H^A(\omega) = \inf\{n \geq 0 :\ X_n(\omega) \in A\}, \quad \omega \in \Omega. \tag{2.2}$$

The probability that starting from i we hit A at some point, called hitting probability, is

$$h_i^A = \mathbb{P}_i(H^A < \infty). \tag{2.3}$$

In addition, the mean hitting time for $(X_n)_{n\geq 0}$ to reach A is

$$k_i^A = \mathbb{E}(H^A) = \Sigma_{n<\infty} n\mathbb{P}_i(H^A = n) + \infty\mathbb{P}(H^A = \infty). \tag{2.4}$$

Informally, we write

$$h_i^A = \mathbb{P}_i(\textit{hit } A \textit{ at some point}), \quad k_i^A = \mathbb{E}_i(\textit{time to hit } A \textit{ the first time}),$$

which can be easily computed.

Markov Chains are a powerful tool to study systems in which an event only depends on the previous one, with applications in Physics, Chemistry, Biology, Speech recognition

or Social sciences. In biology, such examples include birth-death processes, substrate-enzyme kinetics or changes in gene frequencies [43].

2.3 Deterministic Compartmental Modelling

In the case of deterministic modelling, one main method of describing a biological system is by classifying the "populations" under study based on a finite number of properties, and thus splitting them into different compartments. Examples of populations are different species, different types of cells or molecules. Such a compartmental model is described by a system of differential equations, for example ordinary differential equations (ODEs), each of which explains the dynamics of one particular population from the biological system (2.5). Possible actions within the system are interactions between compartments such as transitions from one compartment to another, or inflow (e.g. birth or production) and outflow (e.g. death or degradation) to and from the system. Some traditional examples of compartmental models are predator-prey (e.g. Lotka-Volterra [44, 45]) and epidemiology (e.g. SIR [46]) models.

$$\frac{du(t)}{dt} = g(t,u(t);p), \quad u : [0,\infty) \to \mathbb{R}^n,\ g : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n,\ p \in \mathbb{R}^m \tag{2.5}$$

In the system (2.5), we have n compartments u_i, each described by a function g_i depending on time t, the compartments u_i (or subset, thereof) and the m-dimensional parameter vector p. Such a model is well-posed if the following are true:

- ◇ Global existence: there exists a solution $u(t)$ for $t \in [0,\infty)$
- ◇ Uniqueness: the solution of the problem is unique
- ◇ Non-negativity: the solution is non-negative for non-negative initial data, i.e $[0,\infty)$ is an invariant set (or a subset thereof)

The first two conditions can be proven by applying the theorems of Peano and Picard-Lindelöf [47], respectively. In addition, the local stability of the (hyperbolic) equilibrium points (if any) can be investigated by the Hartmann-Grobman Theorem [47].

2.4 Parameter Estimation

Parameter estimation deals with determining the parameters of a system based on data, by minimising a so-called objective function, which describes the "distance" between the model and the data [48, 49, 50, 51].

As an example we will present a short overview of parameter estimation approaches in the case of a dynamical system with parameters p given by:

$$\begin{gathered}\frac{dx}{dt}(t) = f(t,x(t),p) \quad x(t_0) = x_0 \\ \text{where } x : [t_0,t_f] \to \mathbb{R}^d,\ p \in \mathbb{R}^{n_p},\ f \in \mathscr{C}^1([t_0,t_f] \times \mathbb{R}^d \times \mathbb{R}^{n_p})\end{gathered} \tag{2.6}$$

We denote by Y_{ij} the data of measurement $i = 1,\dots,n$ and observable $j = 1,\dots,m$, where n is the total amount of data and m the number of observables. The data Y_{ij} satisfies

$$Y_{ij} = g_j(t_i,x(t_i),p) + \varepsilon_{ij}, \tag{2.7}$$

for some continuously differentiable observation function g, normally distributed measurement errors $\varepsilon_{ij} \sim \mathscr{N}(0,\sigma_{ij}^2)$, and the sample points t_i which are ordered such that $t_0 \leq t_1 < \cdots < t_n < t_f$.

The aim is to estimate the parameters p based on the measurements Y_{ij}, by minimising an appropriate objective function given by the principle of maximum-likelihood [52]:

$$\min_{x,p}\ h(p) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{(Y_{ij} - g_j(t_i,x(t_i),p))^2}{\sigma_{ij}^2} \tag{2.8}$$

Various methods for minimising such a function exist. In this thesis, the minimisation was employed *via* specific routines implemented in the programming languages *Mathematica*®, version 10.2 (`NMinimize`) and *MATLAB*®, release R2018b (`lsqnonlin`). The methods used by each of these routines are described below.

2.4.1 The Nelder-Mead Method

The algorithm of Nelder and Mead [53] for function minimisation is the most common direct search method and is shortly described bellow. It is one of the methods available in the `NMinimize` routine for numerical minimisation, of the *Mathematica*® programming language [54].

Let us first consider the minimization of a function $h(p_1,\ldots,p_n)$, $n \in \mathbb{N}$, with p the parameter vector. Further, at each iteration, let the points $x_0, x_1, \ldots, x_m$ define the corners of a simplex. Each point x_i is given by a sepcific set of parameters p. A simplex is defined as a polytope with $m+1$ vertices in m dimensions and in the general case it does not have to be regular. As an example, for $m = 2$ the regular simplex is a triangle, while for $m = 3$ it is a tetrahedron. For simplicity, let us order the points such that $h(x_0) \leq h(x_1) \leq \ldots \leq h(x_m)$. At each iteration, the worst point (according to the value of h evaluated at that point) is replaced by a newly generated "better" point, so the first point to be replaced is x_m.

Let c be the centroid of the remaining points. In the case of regular polytopes, the centroid has coordinates obtained by computing the averages of vertices coordinates. For non-regular or even non-convex polytopes the formulae get more complicated. A new trial point x^* is generated by reflecting the worst point x_m through the centroid:

$$x^* = (1+\alpha)c - \alpha x_m,$$

where $\alpha > 0$ is a so-called reflection coefficient. Thus the trial point x^* lies on the line joining x_m and c with a proportionality relation between the lengths of segments $[x^*c]$ and $[x_mc]$: $[x^*c] = \alpha[x_mc]$. The choice of whether to keep this trial point depends on its relation with the remaining points as follows.

- ◇ If $h(x_0) \leq h(x^*) \leq h(x_{m-1})$, then replace x_m by x^*.
- ◇ If $h(x^*) \leq h(x_0)$, it means that the search direction looks promising and a new trial point is generated by expansion, i.e. by reflecting x^* through the centroid:

$$x^+ = \beta x^* + (1-\beta)c,$$

with $\beta > 1$ the expansion coefficient. This time β sets a ratio between segments $[x^+c]$ and $[x^*c]$: $[x^+c] = \beta[x^*c]$. If $h(x^+) \leq h(x_0)$ we replace x_m by x^+ and restart the process. However, if $h(x^+) > h(x_0)$ then the expansion has failed and we use x^* for replacing x_m before restarting the algorithm.

◇ If upon the first reflection, $h(x^*) > h(x_m)$ and thus x^* is worse, then we need to employ a contraction instead. The new trial point x^- upon contraction is generated by:

$$x^- = \begin{cases} \gamma x_m + (1-\gamma)c, & \text{if } h(x^*) \geq h(x_m) \\ \gamma x^* + (1-\gamma)c, & \text{if } h(x^*) < h(x_m) \end{cases}$$

The contraction coefficient $\gamma \in (0,1)$ is the ratio between segments $[x^-c]$ and $[x_mc]$. Next, if $h(x^-) \leq \min\{h(x_m), h(x^*)\}$ we accept the point to replace x_m and restart the process. Otherwise, the contracted point is worse than the better of the points x_m and x^*. In the case of such a failed contraction, all points x_i get replaced by $\frac{x_i + x_0}{2}$ and the process is restarted.

A failed expansion corresponds to advancing into a local minimum valley. A failed contraction, which occurs less often, can be encountered when the direction is promising for some of the simplex vertices but is not favourable for others. In this case a further contraction would move these points away from the valley, contrary to what we need. The proposed solution in the case of a failed contraction, contracts the simplex towards the "best" vertex and thus gradually brings the points toward the valley bottom.

Possible stopping criteria implemented in *Mathematica*® are based on maximum number of iterations or on comparison of a prescribed tolerance with either the distance between best function values in the new and old polytope or with distances between the new best point and the old best point.

2.4.2 The Trust-Region Method for Nonlinear Minimization

The Trust-Region Method [55] is used by the *MATLAB*® programming language routine `lsqnonlin` for least-squares minimization and is described below [56].

For simplicity, let us consider the case of an unconstrained minimization problem, where a function $h(x)$ with $h : \mathbb{R}^n \to \mathbb{R}$ should be minimized. Start at a point $x \in \mathbb{R}^n$, given by a specific set of parameters p, and try to improve $h(x)$ by moving to a new point $y \in \mathbb{R}^n$ with $h(y) < h(x)$. The main idea of the method is to use a simpler function q for approximating h, such that the behaviour of function h in a neighbourhood $\mathscr{N}(x)$ is reasonably reflected by q. This neighbourhood is called the trust region and gives the algorithm its name. A trial step s is computed by minimising (or approximately minimising) over $\mathscr{N}(x)$, which provides us with the trust-region subproblem (2.9).

$$\min_s \{q(s),\ s \in \mathscr{N}(x)\} \tag{2.9}$$

If $h(x+s) < h(x)$, then the current point x is updated to $x+s$; otherwise, the current point is kept, the trust region is shrunk and the search step is repeated. The following tasks of the method are to decide how to choose the function q for approximating h, how to choose and adapt the trust region $\mathscr{N}(x)$, and how accurately to solve the subproblem (2.9).

The choice of the quadratic approximation function q is done by selecting the first two terms from the Taylor expansion of h around the point x, and the neighbourhood $\mathscr{N}(x)$ is usually chosen to be spherical or ellipsoidal. The subproblem is rewritten as (2.10)

$$\min \left\{ \frac{1}{2} s^T H s + s^T k, s.t. \|Ds\| < \Delta \right\} \tag{2.10}$$

where k is the gradient of h at the current point x, H is the Hessian matrix, D is a diagonal scaling matrix, Δ is a positive scalar, and $\|.\|$ is the l_2-norm.

Various algorithms for solving (2.10) exist [57], which rely on computing a full eigensystem and a Newton process applied to the equation (2.11).

$$\frac{1}{\Delta} - \frac{1}{\|s\|} = 0. \tag{2.11}$$

Even though this type of algorithms are able to find an accurate solution to (2.11), they require much computational time, proportional to several factorisations of H. Therefore, in order to avoid this issue, the Optimization Toolbox solvers follow an approximation approach relying on restricting the trust-region subproblem to a two-dimensional subspace S [58, 59], which then needs to be determined. For this purpose, a preconditioned con-

jugate gradient process described below is used. The subspace S is defined as the linear space spanned by s_1 and s_2, with s_1 being in the direction of the gradient k and s_2 either an approximate Newton direction given by the solution of (2.12)

$$Hs_2 = -k \tag{2.12}$$

or a direction of negative curvature given by (2.13)

$$s_2^T H s_2 < 0. \tag{2.13}$$

The main idea behind determining the subspace S is to force global convergence, *via* the steepest descent or the negative curvature direction, and to achieve fast local convergence *via* a Newton step. Once the subspace S has been chosen, equation (2.10) can easily be solved. As a consequence, the unconstrained minimisation using the Trust-Region Method can be summarised as follows:

1. Formulate the two-dimensional trust-region subproblem.
2. Solve (2.10) to determine the trial step s.
3. If $h(x) > h(x+s)$ replace x by $x+s$.
4. Adjust Δ.

These steps are repeated until convergence. The adjustment in Δ is done according to standard rules, and in particular, Δ is decreased if $h(x) \leq h(x+s)$ [60, 61].

2.5 Model Comparison

In the process of deriving a model, one is always faced with the task of introducing as much complexity as necessary for capturing most of the essential properties of the system under study (biological in this case), but at the same time one needs to be cautious not to include too many parameters, as this could lead to overfitting. In such cases, when multiple models are at hand, a model comparison needs to be performed for selecting the "best" model. The best model is not the true model, in the sense that no model will ever explain the system in its entirety, but is the most accurate one out of the available ones.

Probably the most mainstream method used for choosing the best model is that of information criteria. Such criteria consist of a function which takes as arguments the parameter vector θ, the data x and the model g_i with $i \in \mathscr{M}$, the set of models, and return a value used for comparison. Examples include Akaike's (AIC), Bayes' (BIC) and Takeuchi's (TIC) Information Criteria. In the following, the theory of the Akaike Information Criterion and its variant, the Corrected Akaike Criterion (AICc) [62, 63] will be described.

Akaike's seminal paper [64] proposed the use of the Kullback-Leiber (KL) information, which is a measure of the distance between the model and the reality f, based on the parameters θ. However, the KL cannot be computed, as the true f is unknown. The main theory was developed for statistical models and everything relies on the idea that there exists a unique value for θ which minimises the KL distance $I(f,g)$ (2.14), *via* Maximum Likelihood (ML) estimation.

$$I(f,g) = \begin{cases} \displaystyle\int f(x) \log\left(\frac{f(x)}{g(x|\theta)}\right) dx \geq 0, & \text{for continuous distributions} \\ \displaystyle\sum_{i=1}^{k} p_i \log\left(\frac{p_i}{\pi_i}\right) \geq 0, & \text{for discrete distributions} \end{cases} \tag{2.14}$$

where f and p_i represent the "true" distribution, while g and π_i the approximate one of the model. This result can be rewritten as (2.15).

$$I(f,g) = C - \mathbb{E}_f[\log(g(x|\theta))], \quad \text{where } C \text{ is a constant} \tag{2.15}$$

The unknown value of the parameter θ which minimises $I(f,g)$ depends on the true f, the model g, the parameter space and the nature and structure of the data x. Let us denote by θ_0 the "true" minimising value of the parameter. If one knew that g is the best model approximating f, then the Maximum Likelihood Estimator (MLE) $\hat{\theta}$ would approximate θ_0. Since one cannot know the actual value of θ_0 but does know the value of the MLE $\hat{\theta}$, the focus of the model selection switches to minimising the expected KL distance instead of minimising the known KL distance, over the set $\mathscr{M}$ of models considered. Akaike has shown [64] that the issue of a KL criterion is that of minimising

$$\mathbb{E}_y \mathbb{E}_x[\log(g(x|\hat{\theta}(y)))]$$

where x and y are random samples from the same distribution and both expectations are taken with respect to the truth f. Even though estimating (2.5) by the maximum log-likelihood $\log(\mathscr{L}(\hat{\theta}|x))$ of each model g_i seems appropriate, there is a bias that appears, bias which is approximately equal to the number K_i of estimable parameters in the model g_i. Then Akaike's Information Criterion [65, 66] reads

$$AIC = -2\log(\mathscr{L}(\hat{\theta}|y)) + 2K \tag{2.16}$$

Thus, *via* the *AIC* (2.16), one has an estimate of the expected, relative distance between the fitted model and the unknown reality which generated the data. The *AIC* value includes indeed the trade-off between having a good fit (*via* the log-likelihood) and avoiding over-fitting (*via* the term K). Thus, the best model is selected by choosing that with the lowest *AIC* from the considered ones. In this sense, the actual value of the *AIC* does not matter, but instead the differences $\Delta_i = AIC_i - AIC_{min}$.

The general approach of likelihood theory which stays at the base of *AIC* computation, needs to be adapted for the case of Least-Squares (LS) estimation, with normally distributed errors. In this case, if all models assume normally distributed errors with constant variance, the *AIC* for LS derived from LS regression statistics [67] is

$$AIC = n\log(\hat{\sigma}^2) + 2K \tag{2.17}$$

where

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^{\,2}}{n}, \quad \text{with } \hat{\varepsilon}_i \text{ the estimated residuals, and } n \text{ the sample size.}$$

Thus the *AIC*, as well as other information criteria not described here, are in agreement with the principle of parsimony saying that a model with "... the smallest possible number of parameters for adequate representation of the data" should be preferred. [68]

The corrected AIC (2.18), introduced by Sugiura [69], is a second-order variant of the AIC that aims to correct the bias which appears if the sample size is small.

$$AICc = AIC + \frac{2K(K+1)}{n-K-1} \tag{2.18}$$

The use of *AICc* (2.18) is recommended unless the sample size is large with respect to the number of estimated parameters (say when $n/K < 40$).The AICc will be used instead of AIC throughout this thesis, for selecting the most appropriate model.

Chapter 3
Biological Background

"Look deep into nature, and then you will understand everything better."
- Albert Einstein

This chapter presents a short overview of the biology needed for understanding and easily navigating the rest of this thesis.

3.1 Model organism

In recent years, the *Oryzias latipes*, also known as the Japanese rice fish or, in short, medaka, has emerged as a model organism for studying organogenesis as well as other species-specific features such as sex-determination or adaptation to different water salinity [9, 70]. Medaka is a complementary model organism to other mainstream ones, such as zebrafish.

Due to its rapid development and its transparent body during juvenile stages (Fig. 3.1) [71], medaka is a convenient organism for investigating organogenesis and development. Organs that have been studied so far include and are not limited to the retina [72, 73], the brain [74] and the liver [75, 76].

In this thesis the post-embryonic growth of the respiratory organ of medaka fish is placed under the microscope (Fig. 3.1c). Features such as the relation between growth and homeostatic stem cells, or between different types of stem cells as well as their co-ordination are investigated. These studies can help in understanding how organs adapt to the permanent changing organismal size (Fig. 3.1a) [77], and could be extrapolated to mammalian organisms and humans, either during early development or in the case of diseases.

(a) Organismal growth in medaka

(b) Transparent body

Fig. 3.1: (a) Size difference between a medaka hatchling (left) at 10 days post-fertilization (dpf) and an adult (right) at 6 months post-fertilization (mpf) - image by Prof. Dr. Làzaro Centanin, Centre for Organismal Studies, Heidelberg University. (b) Two medaka showing the transparency of their body. One can easily see the spine and skeleton, the stomach (dark colour) and the respiratory organs (shown by arrows) - image adapted from the Birney group, EMBL-EBI, Hinxton.

3.2 Structure of the organ

The branchia (also known as gill) is the respiratory organ of the fish and is the biological system under study throughout this thesis. Each branchia has a stereotypic modular organisation, being composed of four pairs of branchial arches, each containing a sequence of filaments which, in turn, are composed of multiple stacked lamellae (Fig. 3.2). Filaments grow in size (number of lamellae), as well as in numbers throughout the fish life, thus driving the organ growth along two orthogonal axes (transversal and longitudinal). Each

branchial arch is, in fact, composed from two rows of intercalated filaments, denoted sides *A* and *B*, thus obtaining a total of 16 arches per gill.

Lineage tracing consists in labelling a cell in such a way that the label is passed on to the progeny, thus resulting in a conglomerate of labelled cells that share an origin - a clone. Various labelling methods exist, but in this work a genetic Cre-Lox strategy is employed (further described in Chapter 4, Fig. 4.1). These experiments are therefore used for identifying the progeny of a single cell, providing information about their number, location and differentiation status. By employing such lineage tracing experiments, stem cell niches were discovered at the periphery of each branchial arch, responsible for filament generation, and also at the tip of each filament important for their elongation [20, 21]. In the following we will refer to these as branchial arch or peripheral niches (containing arch stem cells) and filament or tip niches (hosting filament stem cells), respectively.

Another method of labelling cells is through IdU (iodo-deoxyuridine) experiments. IdU is incorporated in proliferating cells, those in the S-phase of the cell cycle (when DNA duplication happens), the label is passed to the progeny and then diluted in time with every division event. When incubating medaka with IdU, the gill filaments incorporate

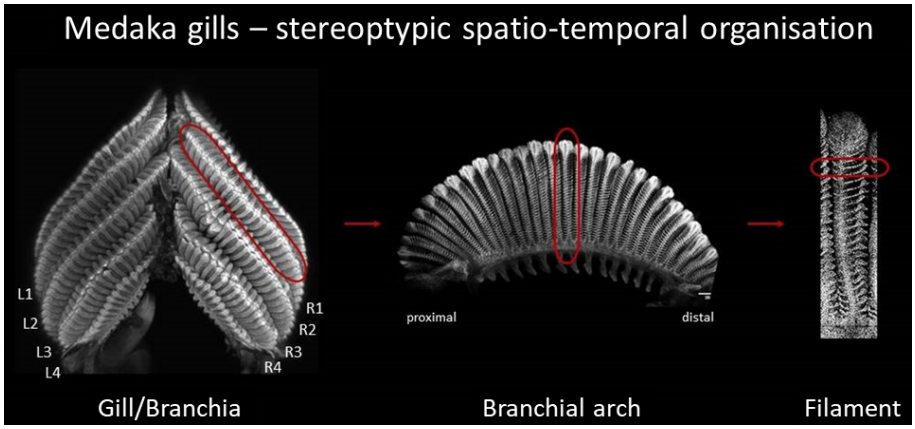


Fig. 3.2: Structure of a gill: one branchia (left) is composed of eight branchial arches (middle), each of which is built from a sequence of filaments (right). The branchial arches on the left are annotated as L1-L4 from the longest to the smallest, and those on the right by R1-R4, analogously. The A side of the branchial arch (middle) is shown, while the B side cannot be seen in this figure, as it is found behind. Image by Julian Stolper, Centanin group, Heidelberg University [20].

the label along the entire filament. What is expected when the fish is then allowed to grow after the incorporation, if stem cells are situated at the tip of the filament, is that the tip and the newly produced lamellae would be IdU-negative, since the label would be lost due to division events (Fig. 3.3). These results reinforce the previous assertions about the existing stem cell niches [21].

3.3 Organ growth

The stem cell niches at the periphery of branchial arches are responsible for filament generation. Hence, following divisions of stem cells in this niche new filaments are inserted in the arch. Focusing on one of the peripheries and keeping in mind that the other side develops similarly, the simplest case to understand and to explain graphically is when a new filament is produced by the division of one stem cell (Fig. 3.4): starting with a few embryonic filaments, when one branchial arch stem cell A divides, its progeny A' introduces a new filament in the arch. In the following time step, two actions can take place: the stem cell A that remained in the niche divides again and its daughter-cell A'' generates a new filament, or the filament stem cell A' in the newest filament divides to elongate it. This process continues throughout the life of the fish. To be noted is the fact that filaments are believed to be independent of each other, and their growth only depends on the division of stem cells in the niches. Furthermore, each of the two halves of the arch grows by the same mechanism but independently of each other, being driven by the properties of the two peripheral niches.

3.4 Filamental patterns

Each filament is composed of multiple cell types having various functions. Such examples include pillar cells responsible for controlling the blood flow through lamellae and reducing the pressure drop during the flow of blood across the gill; pavement cells playing an important role in gas transport and secretion of protons [78]; neuroepithelial cells important as oxygen sensors and regulators of blood flow [20] - to name a few. Listing all gill cells and their functions is not within the scope of this thesis, but important to remember

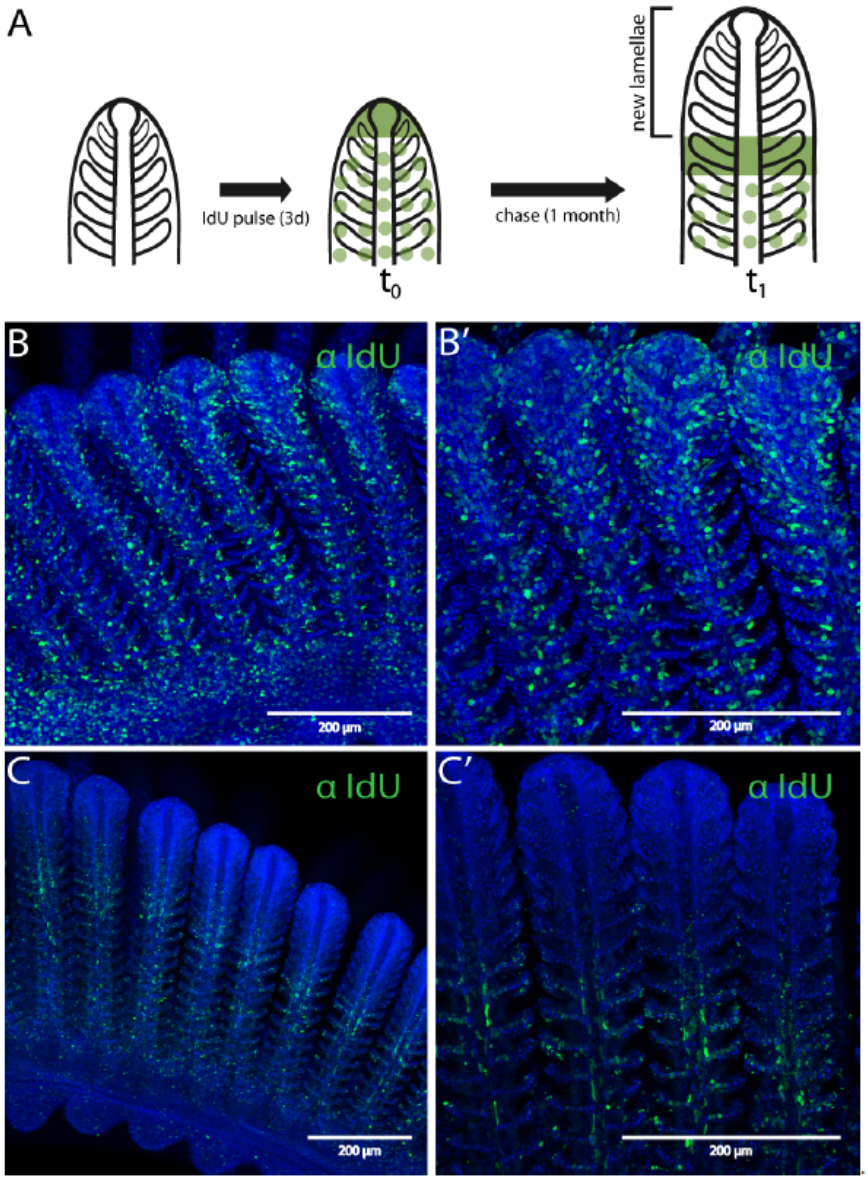


Fig. 3.3: (A) Scheme of the experimental design and expected site of lamellae generation. Putative highly proliferative tissue is labelled in dark green. Adult medaka treated with 33 mg/L IdU for 3 days (t0) were grown for 1 month (t1). (B) Proliferative cells in the filament tissue at t0. (B') Higher magnification of proliferative cells in the filament tissue at t0. Nuclei were labelled with DAPI (blue). (C) Filaments of adult fish with unlabelled tips and IdU positive cells in a transition zone at t1. (C') Higher magnification of filament tips and transition zone containing IdU positive cells. Nuclei were labelled with DAPI (blue). Image by Julian Stolper, Centanin group, Heidelberg University [20].

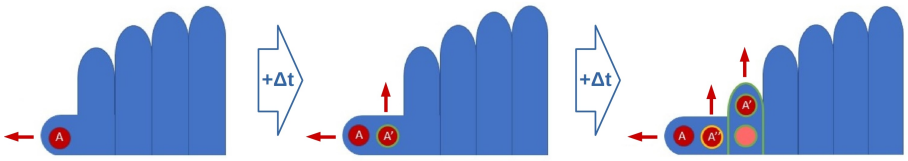


Fig. 3.4: Growth of a branchial arch

is that the generation of each filament requires the coordination of multiple fate-restricted stem cells giving rise to the different specialized tissue types. In each peripheral niche there exist at least four types of fate restricted stem cells [21]. Through lineage tracing experiments four different filamental patterns produced by the four labelled stem cell types have been observed. Since the aim of this study concerns determining the number of stem cells of each type, we will not focus on the function of each such filamental tissue pattern and we will simply refer to them as patterns 1-4 corresponding to cell types 1-4 (Fig. 3.5).

In addition to "clean" patterns in which the labelling corresponds to only one of the fate-restricted stem cells in the branchial arch niche being labelled, mixed or partial patterns were also observed. Mixed patterns appear when two or more labelled fate-restricted cells are selected together for generating a new filament. Partial patterns appear if at some point, one of the fate restricted stem cells is lost and is replaced by another one, which has a different label than the lost one. Another case is when half of the filament (longitudinally) presents one pattern and the other half another one. One could hypothesise that in fact, each tip niche hosts two stem cells of the same fate, each giving rise to one half of the filament (Fig. 3.6).

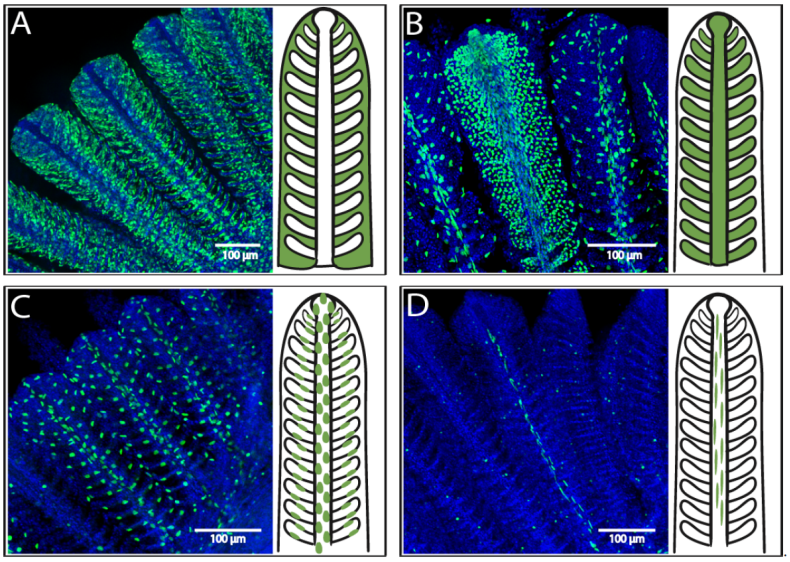


Fig. 3.5: (A) Inter lamellar pattern (Pattern 1). The majority of cells are labelled on the epithelium and between lamellae. (B) Intra lamellar pattern (Pattern 2). Clonal filaments contain labelled nuclei inside of each lamellae and in the middle of the filament. (C) Epithelial pattern (Pattern 3). Single nuclei are labelled in the filament epithelium and on the lamellae epithelium. (D) Intra filamental pattern (Pattern 4). Labelled elongated cells, located in the middle of the filament. Image by Julian Stolper [20].

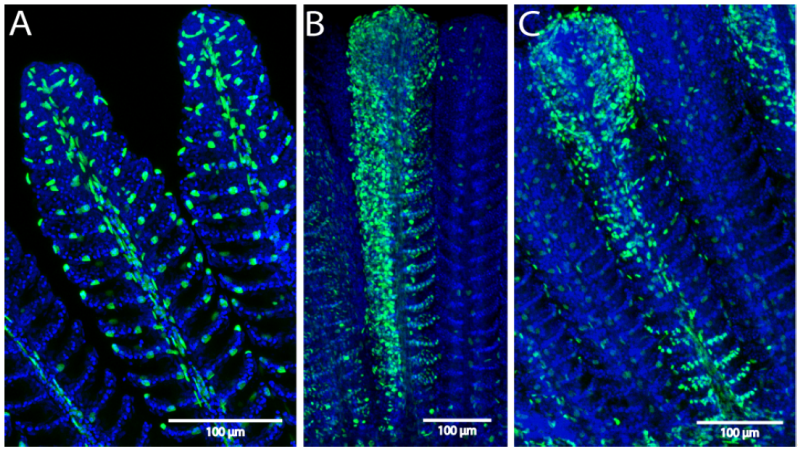


Fig. 3.6: (A) Example of clonal filaments consisting of different patterns. Displayed filament contains the intra filamental and epithelial pattern. (B) Clonal filament labelled with intra lamellar cells but one half is labelled with inter lamellar cells. (C) Clonal filament with the apical part labelled with another pattern than the basal part of the filament. Image by Julian Stolper [20].

Part I

Mathematical Modelling of Stem Cell Numbers and Functional Heterogeneities

One interesting and important question in developmental biology concerns the relation between growth and homeostatic stem cells. Are the same stem cells able to perform both tasks, or are there two separate stem cells populations, each responsible for one of the tasks? In particular, one approach for coming closer to understanding such an aspect is to determine the number of stem cells necessary for building and maintaining an organ. Such a study could further our understanding of cancer stem cell behaviour in higher vertebrates and could have applications in wound healing and organ regeneration. By using the medaka fish model organism we aim to determine the number of stem cells needed to form the respiratory organ, which is of high importance in order to understand the coordination of fate-restricted stem cells. The fish gill is a suitable system for qualitatively investigating a stem cell niche hosting multiple coordinated fate-restricted stem cells, as is described in the following. This question, although seemingly basic, is in fact very difficult to approach in many biological systems. The difficulty stems from the lack of specific stem cell markers or from the lack of structure in the organ under study. By using methods such as stochastic simulations and Markov Chains theory, together with numerics and parameter estimation, in this part we aim to answer various questions related to stem cell behaviour during post-embryonic growth of the organ.

Chapter 4
Experimental Procedure and Data Acquisition

This chapter describes the experimental procedures used in acquiring the data and the type of data adjustment performed for the further use in mathematical modelling. It is important to understand these aspects before going further into the mathematical approach.

In the study of branchiae in general, and in particular those of medaka, one can take advantage of the organ spatio-temporal organization [21], which has been described in Section 3.2 (Fig. 3.2). As explained in Section 3.3, filaments are generated by divisions of stem cells in the niches at the basal extremities of the branchial arches (Fig. 3.4). The lineage tracing experimental procedure labels few cells at juvenile stage and since the label is passed on to the progeny, all filaments generated by a labelled stem cell will be labelled, and correspondingly, the filaments produced by an unlabelled stem cell will be unlabelled. The labelling is done by using the Gaudi^{RSG} (Red-Switch-Green) fish line presented in [72]. This ubiquitously expresses a red fluorescent protein (RFP), which prevents the expression of a nuclear green fluorescent protein (GFP). Upon heat-shock or induction with Tamoxifen, the RFP is cut out allowing for the GFP to be expressed in the nuclei, and also to be inherited by the progeny (Fig. 4.1). As this is an ubiquitous expression, the labelling of the epithelium on top of branchial cells hinders live imaging procedures. Therefore, the method chosen to study the recombined medaka gills is to have the fish sacrificed, dissected and their gills extracted and split into the respective branchial arches (R1-R4 and L1-L4).

Coming back to the study of post-embryonic growth of medaka gills, let us now focus our attention on one stem cell niche and thus on one half-arch. A “half-arch” is defined

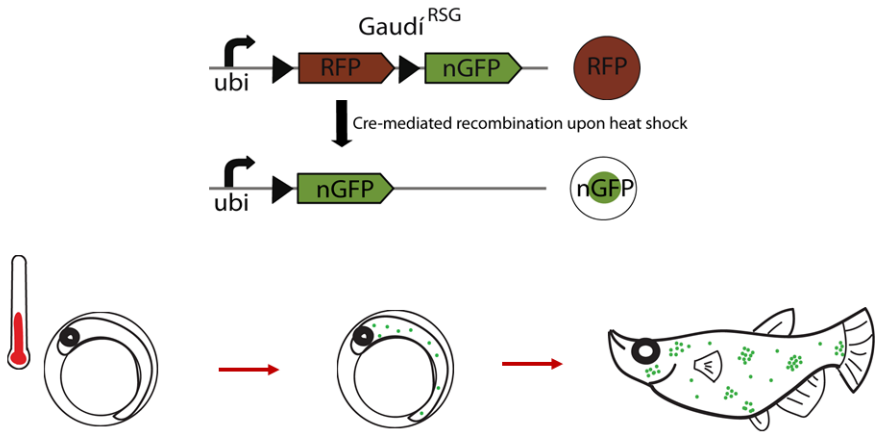


Fig. 4.1: Experimental procedure: Gaudi^{RSG} fish line in which, upon heat shock, the Red-Fluorescent-Protein (RFP) is cut allowing for the Green-Fluorescent-Protein (GFP) to be expressed in the nuclei. As a consequence cells are labelled with GFP, which is then transmitted to their progeny as the fish grows. Image by Julian Stolper, Centanin group, Heidelberg University.

as the sequence of filaments next to one niche, which are generated by the stem cells in that niche. Therefore, each branchial arch is composed of two half-arches, but they must not necessarily be equal in the number of filaments, i.e. “half” here does not mean 50% of the filaments.

The experimental images show branchial arches containing labelled (1) and unlabelled (0) filaments (Fig. 4.2) and thus the data record the filamental label for each branchial arch, in arrays of 1s and 0s. Thus, one can say that the data records the history of divisions of the labelled and unlabelled stem cells from the niches at the extremes, not being able however to distinguish between cells carrying the same label.

Considering the way in which branchial arches grow, there are certain insights that can be drawn. For example, if the stem cell niche hosts only one stem cell, then the entire half-arch will have the label of that particular stem cell. On the other hand, if more than one stem cell reside in the niche, and they can be either labelled or unlabelled, one expects to observe a mixture of labelled and unlabelled filaments in the half-arch. What is in fact most often observed in the experimental images is an intermediate scenario displaying long stretches of consecutive filaments having the same label close to the niche [21].

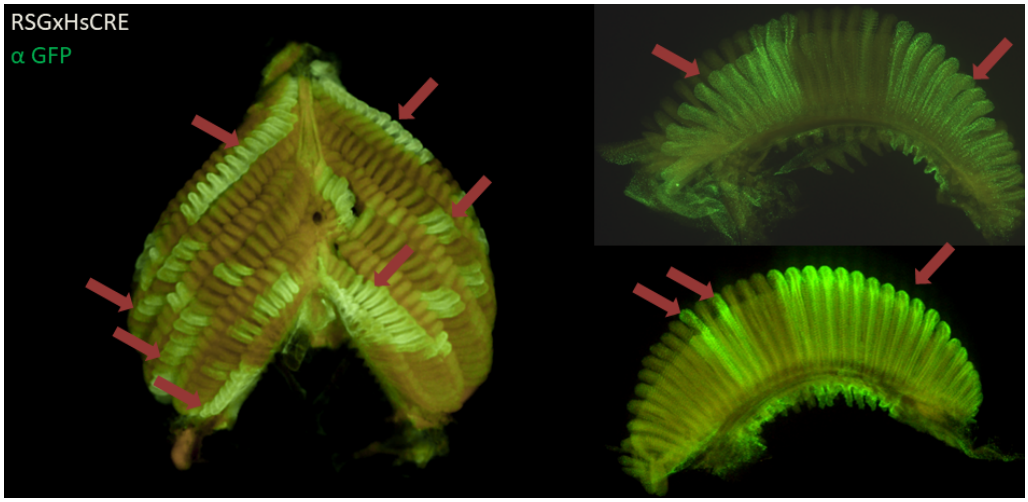


Fig. 4.2: Experimental images. Left: Ventral view of a recombined adult medaka gill. One can see the labelled filaments within branchial arches. Right: Two branchial arches which display clonal filament stretches (either labelled or unlabelled). Red arrows point to the GFP-labelled filaments or streaks. Image by Julian Stolper.

Since the goal of the project is to estimate the number of stem cells in each niche, we focused on stretches of 8 filaments from each extreme, in order to exclude the embryonic filaments from the study, as they were formed before the experimental procedure. In the following we shall refer to these 8-filament long stretches as “mini-arches”. The choice of number 8 was made after excluding branchial arches with less than 25 filaments in total, and by considering the fact that in general there exist approximately 5-9 embryonic filaments per branchial arch.

The data record the distribution of labelled and unlabelled filaments, the labelled case having four different options corresponding to the four filamental patterns observed. Recall that one can occasionally observe a mixture of patterns within the same branchial arch or even within the same filament, resulted from the coordination of multiple labelled fate-restricted stem cells from the niche. What is important to keep in mind is that these patterns are produced by different fate-restricted stem cells and by analysing each cell type independently one avoids the issue of patterns mixing (Fig. 4.3).

Thus for each particular pattern, we obtain a data set containing as many rows as mini-arches, each of them consisting of binary entries corresponding to labelled or unlabelled

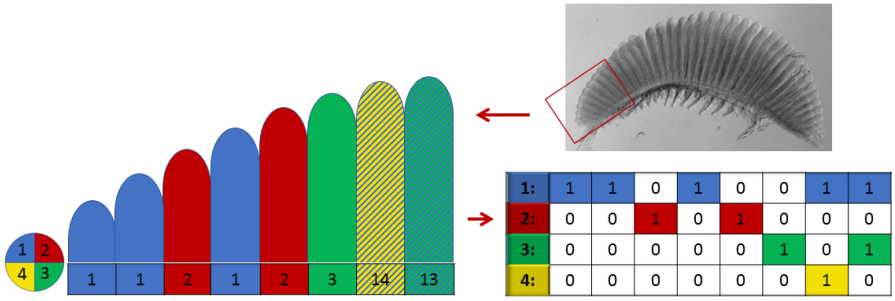


Fig. 4.3: Example data from toy mini-arch with mixtures of patterns within the arch and within certain filaments (e.g. 13 - mixture of patterns 1 and 3 in the last filament on the right). For analysing each pattern, shown by a different color in the scheme, one has to edit the data such that it consists of 1 and 0 for whether the filament is labelled with that particular pattern or not.

filaments. Those with mixed patterns ab with $a,b \in \{1,\ldots,4\}$ are considered as labelled both in the Pattern a and in the Pattern b data sets. In each row, columns are organised such that the row starts with the value corresponding to the oldest filament and continues towards the youngest one. In other words, the left mini-arches are reversed, while the right ones remain in the initial order.

Chapter 5
Homogeneous Model for Stem Cell Numbers

This chapter presents the first simple model developed, which aids in the better understanding of the biological system and in the further development of hypotheses and assertions.

In order to analyse the data, the notion of switches is introduced, allowing the study of stem cell divisions history.

Definition 5.0.1 *A switch is defined as the transition from a labelled to an unlabelled filament or viceversa, within an arch.*

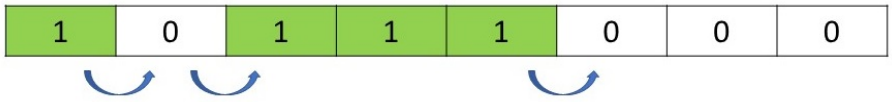


Fig. 5.1: Toy mini-arch with three switches and four labelled filaments

Therefore switches give us an idea of the alternation of labels (i.e. 1 or 0) in a mini-arch.

However, just counting the number of switches is not sufficient for the purpose of obtaining an overall information on stem cell divisions: for example, an entirely labelled branchial arch has no switches, but the same is true for an entirely unlabelled branchial arch. Therefore, the number of labelled filaments is counted, in addition. This number of labelled filaments in each arch depends in turn on the labelling efficiency of the ex-

periment. If the labelling efficiency was 100% then all stem cells in the niche would be labelled, therefore producing a completely labelled branchial arch.

5.1 Labelling efficiency

In what concerns the goal of counting the number of stem cells in the peripheral niche, the labelling efficiency (from hereon denoted *probLab*) reflects the ratio between the labelled stem cells and the total number of stem cells in that niche. This depends on the experimental procedure and the environmental effects and is unknown. Nevertheless it can be approximated from the data as follows.

For each branchial arch, the first filament in the data, i.e. the oldest post-embryonic filament considered, indicates whether the *first* stem cell which divided post-embryonicaly was labelled (and had a pattern 1-4) or unlabelled. In longer branchial arches this might not be in fact the first stem cell which divided post-embryonicaly but the estimation is sufficiently good for our purpose. Therefore, by counting the number of labelled first filaments (for each pattern in particular) and dividing by the total number of arches (5.1), the labelling efficiency for each pattern is estimated and reads as seen in Table 5.1.

$$probLab = \frac{\text{Number of labelled first filaments}}{\text{Total number of branchial arches}} \tag{5.1}$$

Pattern	Pattern 1	Pattern 2	Pattern 3	Pattern 4
probLab	0.13	0.06	0.04	0.01

Table 5.1: Labelling efficiency for each pattern.

The values of the labelling efficiency differ from one pattern to another and at first sight this seems counter-intuitive. However, considering that these patterns are produced by stem cells which are fate restricted to specific lineages, this is not surprising since different cell types can assimilate tamoxifen in different quantities or can react differently to the heat shock, tamoxifen and heat shock being the two methods used for inducing the label. What is still relatively surprising is the big difference between these values: for example between Pattern 1 and Pattern 4 there is a ten-fold difference in the labelling efficiency. This aspect will be revisited later on, in Chapter 6.

5.2 Derivation and application of the homogeneous model

To begin with, one question that arose was whether the generation of filaments is performed by progenitor cells or by a self-renewing stem cell. Here, by progenitors, we mean daughter-cells of one or more stem cells, which have been produced before the labelling procedure. Then, upon induction, some of those progenitors will be labelled. If filaments are produced by these progenitors, with each filament being generated by one such cell, then depending on the labelling in the niche, one expects a random mixture of labelled and unlabelled filaments arising in the mini-arch. On the other hand, if all filaments are produced by one *bona fide* self-renewing stem cell then the entire mini-arch would possess the label of that stem cell.

Accordingly, two models describing the two extreme scenarios were considered:

- ◇ **M1**: the stem cell model - all filaments are produced by the same stem cell
- ◇ **M2**: the progenitor model - each filament is generated by a different progenitor cell

For this particular investigation, the notion of patterns was ignored, and we focused primarily on whether a filament possesses any label or not. We used mini-arches of 6 filaments, instead of 8 as described before, and considered branchial arches with more than 20 filaments in total, instead of 25. We chose this approach because we wanted to analyse each gill separately in order to see what influence the labelling efficiency has on the results. By excluding only arches with less than 20 filaments, a bigger dataset could be used, with more arches per gill. Hence the labelling efficiency was computed separately for each gill (Table 5.2) as previously described in Section 5.1, and 22 gills were analysed. These 22 branchiae were selected out of a total of 36, and they represented the cases which had at least 8 branchial arches with more than 20 filaments each.

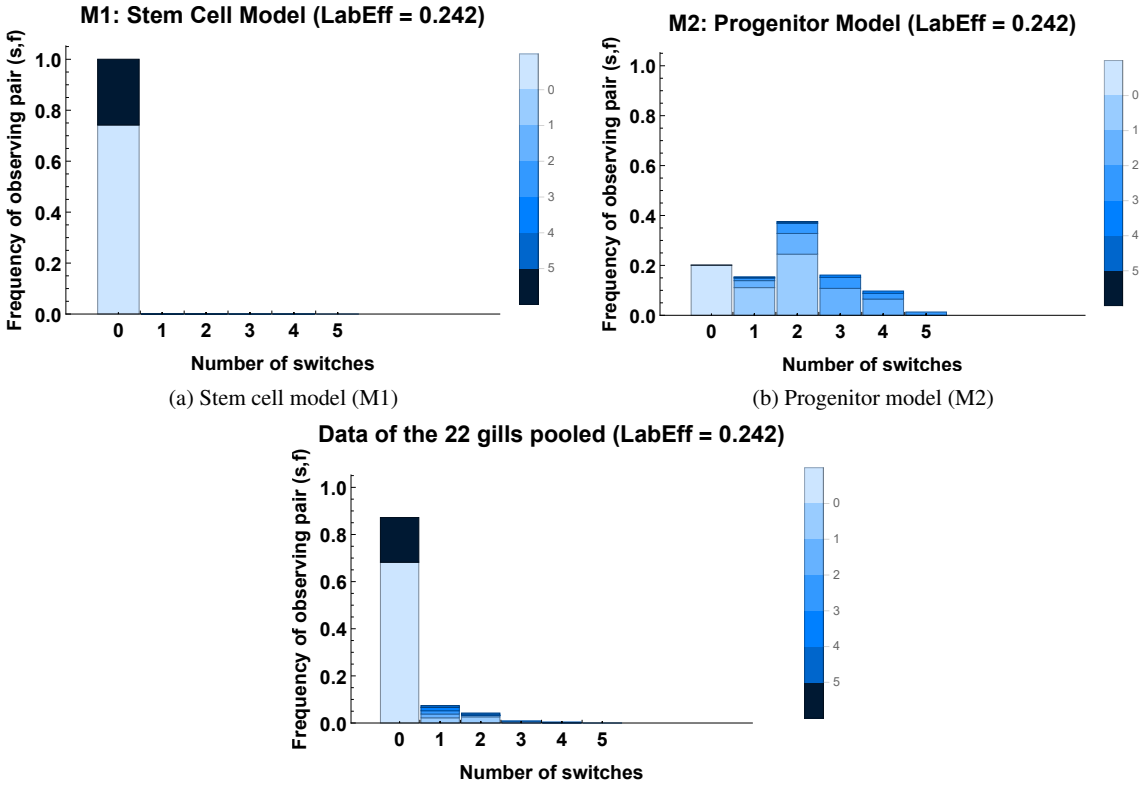
Gill	1	2	3	4	5	6	7	8	9	10	11
probLab	0.077	0.088	0.09	0.194	0.2	0.209	0.211	0.241	0.243	0.287	0.289
Gill	12	13	14	15	16	17	18	19	20	21	22
probLab	0.298	0.305	0.32	0.33	0.332	0.357	0.364	0.39	0.419	0.472	0.488

Table 5.2: Labelling efficiency for separate gills. Here we ignore the patterns and we treat as labelled filaments (1) any filament which possesses any label/pattern (1-4).

We performed stochastic simulations for the above two models, starting with a set of filament generating cells and empty arrays representing the developing mini-arches. Such stochastic simulations simply "chose the cell to divide" thus filling a 6-entry-long array corresponding to the mini-arch with binary values.

- ◇ **M1**: This case was straightforward as only one cell was considered which was labelled with a probability *probLab* equal to the labelling efficiency for each particular gill. Next, since this stem cell produces the entire mini-arch, all six filaments will possess the label of that cell, therefore obtaining at the end of multiple trials a dataset for each gill containing arrays of six 1s or six 0s with the ratio between those representing the labelling efficiency *probLab* for that particular gill. For modelling this scenario, we drew random numbers from a Bernoulli distribution where the probability parameter equals the experimental labelling efficiency of our dataset: Bernoulli$(probLab)$. Each draw represents one arch, with a success corresponding to having the stem cell (and the entire arch) labelled (1), and a failure to it being unlabelled (0). We performed 1000 trials.
- ◇ **M2**: In a similar manner we considered the case of having six progenitor cells in the niche, each producing one filament. This time, we need six realisations of a Bernoulli$(probLab)$ process, for each branchial arch, to simulate how many progenitor cells are labelled and in which order they divide. Again, 1000 trials of the six-outcome Bernoulli process were performed.

Through plots as in Fig. 5.2 the results from each model were initially compared to the data coming from pooling all the above 22 gills together. In the case of **M1** only zero-switches can be seen, while in **M2** more variation is observed. To compare each model to the experimental data for independent gills, we computed an objective function in the form of a sum of square differences for each gill and each model. The smaller this objective function is, the better the fit between experimental data and simulations. Both the number of switches and of labelled filaments in each branchial arch were recorded. There exist 19 possible pairs (s, f) with s switches and f labelled filaments, ranging from (0,0), (0,6) up to (5,3). We calculated for each pair i, of the form (s, f), the frequency of observing it in the data from each gill j, $fD_i^{(j)}$, and in simulations of 5000 filament stretches per gill j, $fS_i^{(j)}$. The objective function $f^{(j)}$ was computed for each gill as an



(a) Stem cell model (M1)

(b) Progenitor model (M2)

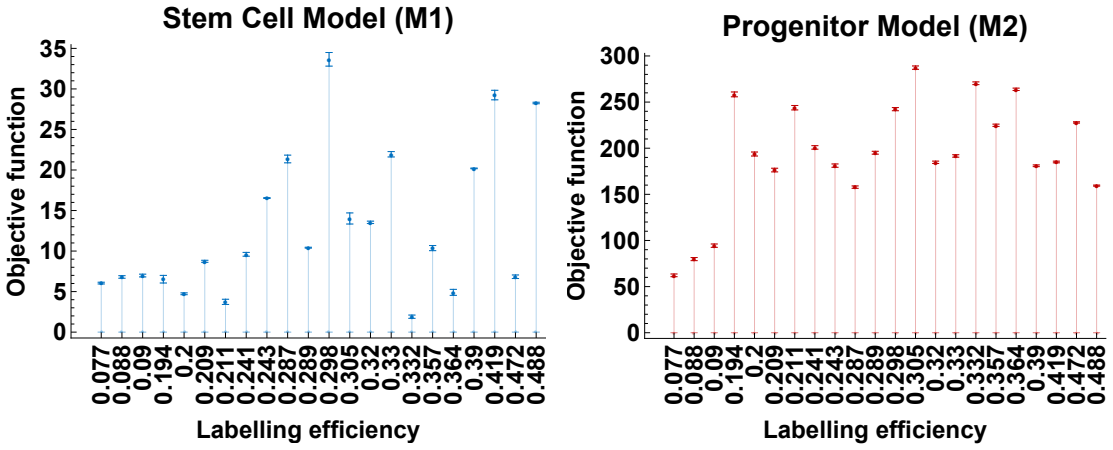
(c) Data

Fig. 5.2: Extreme scenarios: in each plot the x-axis shows the number of switches, while the y-axis records the frequency of observing a certain number of switches. The colour code represents the number of labelled filaments from light to dark blue corresponding to 0-6 labelled filaments. (a) Stem cell model M1. (b) Progenitor model M2. (c) Data. Simulations were run with an overall labelling efficiency of 0.242, as estimated from the data. In these plots, all 22 branchiae were pooled together to first get an overview of the comparison between each model and the dataset.

adjusted sum of square differences:

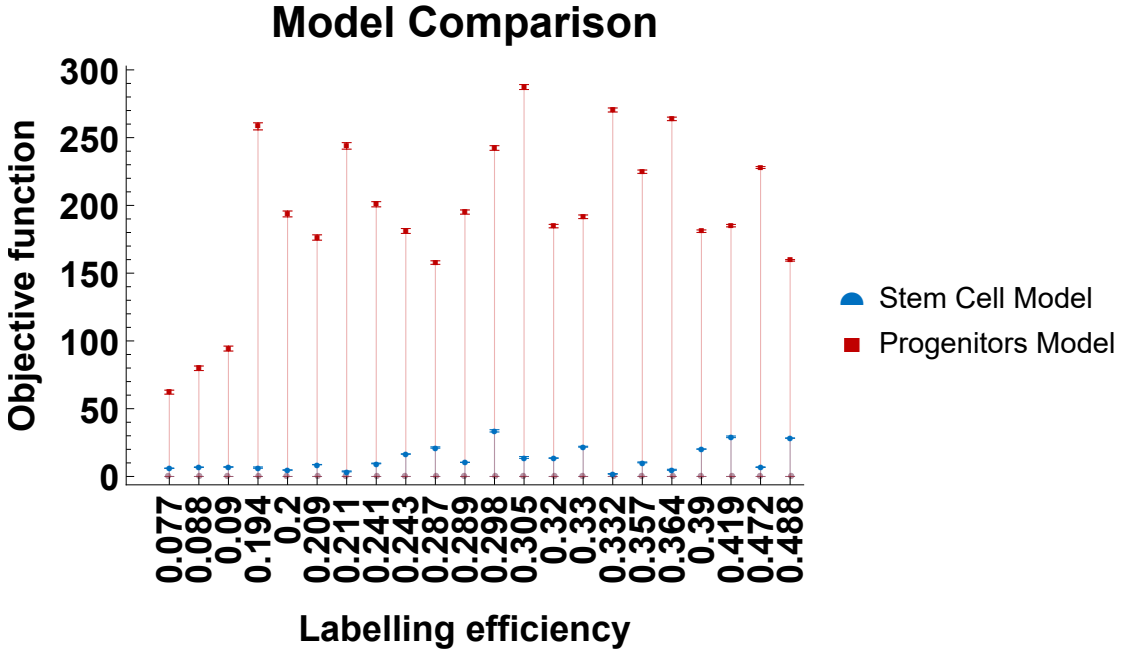
$$f^{(j)} = \frac{\Sigma_{i=1}^{19}\left(fD_i^{(j)} - fS_i^{(j)}\right)^2}{19} \cdot 10^4 \tag{5.2}$$

This was done for both the stem cell and the progenitors model. The factor 10^4 was introduced for avoiding small numbers thus facilitating the comparison between results. The procedure was repeated 1000 times, producing 1000 objective functions per gill and



(a) Stem Cell Model

(b) Progenitor Model



(c) Model Comparison

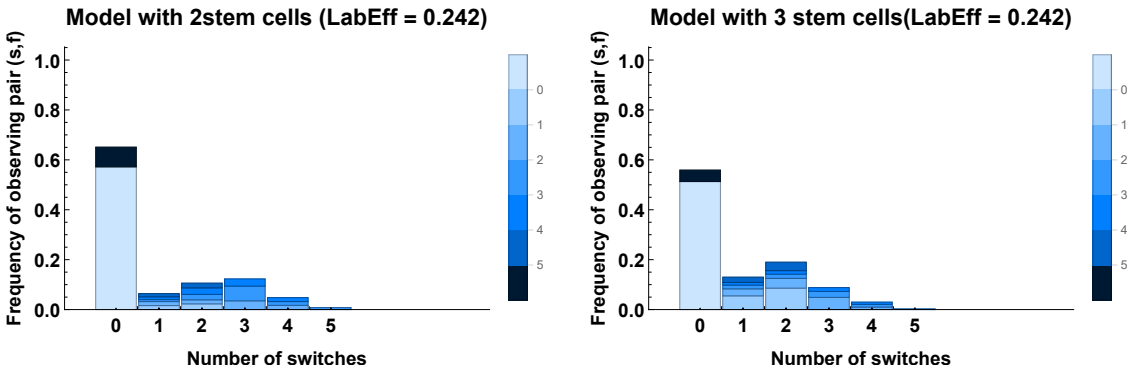
Fig. 5.3: The plots depict the objective function values (y-axis) with respective error bars for each individual gill ordered by their labelling efficiency (x-axis). (a) Stem cell model M1. (b) Progenitor model M2. (c) Plots (a) and (b) combined in one graph for comparison. One can easily notice the better fit model M1 has to the data, given by the smaller values of the objective functions for each of the 22 gills considered.

per model, and therefore obtaining an average value and a standard deviation for each gill for each model. These were plotted in Fig. 5.3, plot which suggests that **M1** fits the data much better than **M2** for each gill. These methods together with further insights from various biological experiments were described in [21].

We can thus conclude that filaments are generated by stem cells, but also that more than one stem cell should reside in the niche since the data contain cases of branchial arches with at least one switch. Accordingly, similar stochastic simulations for the cases with two and three homogeneous stem cells in the niche were performed using the following steps:

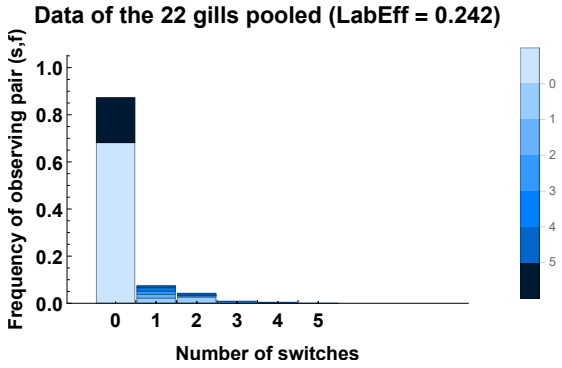
1. Initialize number of stem cells in the niche, $n \in \mathbb{N}$, labelling efficiency, $probLab \in (0,1)$, and empty arch (array).
2. Decide how many stem cells are labelled ($L \in \mathbb{N}$ cells) and unlabelled ($U \in \mathbb{N}$ cells):
$$L \sim \text{Binomial}(n, probLab), \quad U = n - L.$$
3. Randomly select whether a labelled or unlabelled cell divides, to add a 1 or a 0 to the arch. The random selection is weighted by the values L and U, as all stem cells have equal chance of being selected but the label of the next filament is, of course, influenced by the relation between L and U.
4. Repeat step 3. until the arch has six filaments, i.e. array has six entries.

After performing 1000 trials, the results were compared to the data from the 22 pooled gills (Fig. 5.4). As can be seen from the plots, the more stem cells included in the model the worse the fit gets. This suggests that the niche hosts few active stem cells but something is still missing from our understanding of how stem cells divide to produce filaments post-embryonically.



(a) Model with 2 stem cells

(b) Model with 3 stem cells



(c) Data

Fig. 5.4: Plots as in Fig. 5.2. (a) Simulation of the case with two stem cells in the niche. (b) Simulation of situation with three stem cells. (c) Data from the 22 gills pooled. A labelling efficiency of 0.242 was used, as before.

Chapter 6
Heterogeneous Model for Stem Cell Numbers Analysis

"Truth is much too complicated to allow anything but approximations."
- John von Neumann

This chapter extends the methods previously used to include heterogeneity among the stem cells in the niche. Two different techniques for modelling and analysing the biological system are presented, leading to agreeing insights and results.

After comparing the previous extreme scenarios and homogeneous models to the data, the hypothesis that few *heterogeneous* stem cells should reside in the niche was reached. Considering that filaments are independent, the heterogeneity is the missing aspect mentioned in the previous chapter and comes from the division behaviour of the niche stem cells.

The heterogeneity parameter introduced in the model was *the probability that the stem cell which has just divided will divide again in the next time step*, further called *probability of division* and denoted by *probDiv*. This idea can be explained by phases of activation and quiescence of the stem cells, i.e. when a stem cell of a certain fate divides it becomes active, increasing its probability of division, and therefore dividing multiple times before becoming quiescent and allowing another stem cell of the same fate to take over the task of filament generation through coordination with the cells of other fates. For this model we included the analysis of the four different patterns and considered 8-filaments long mini-arches from branchial arches with at least 25 filaments in total. Branchial arches from all gills were pooled together and only one average labelling efficiency was computed per cell type (and hence per pattern). **Assumptions 6.0.1** describe the insights gathered so far before beginning the modelling task.

Assumptions 6.0.1

- ◇ *A filament is generated by the divisions of (at least) four fate-restricted stem cells residing in the peripheral niche, thus when analysing each pattern separately we can talk about one stem cell of each type per filament generated.*
- ◇ *One labelled stem cell (with respective pattern 1-4) in the peripheral niche produces a labelled filament.*
- ◇ *One unlabelled stem cell (0) in the niche gives rise to an unlabelled filament.*
- ◇ *The number of labelled stem cells in the niche, $L \in \mathbb{N}$, depends on the labelling efficiency probLab $\in (0,1)$ and on the total number of stem cells in the niche $n \in \mathbb{N}$:*

$$L \sim Binomial(n, probLab)$$

- ◇ *The number of unlabelled stem cells, $U = n - L$, $U \in \mathbb{N}$.*
- ◇ *The cell which has just divided will divide again in the next time step with probability probDiv $\in [0.5,1)$ - see Section 6.1. Here time step refers to a division event and not to a time point reached after a specific time interval. Since branchial arches with their labelled filaments present to us a "history" of stem cell divisions we do not need actual information on time points, but just on division events, which for simplicity will be hereon referred to as time steps.*
- ◇ *Parameters n and probDiv will be further determined* via *parameter estimation methods, by making use of the value for probLab and formulae for L and U.*

Subsequently, as we aim to compute the frequency of observing each pair (s, f) of switches and labelled filaments, we need to compute conditional probabilities of a filament of label a to be followed by a filament of label b. Here $a, b \in \{0,1\}$, as we analyse patterns separately. Table 6.1 summarises the notations used throughout this chapter.

Parameter	Domain	Description
n	$\mathbb{N}$	Number of stem cells in the niche
L	$\mathbb{N}$	Number of *labelled* stem cells in the niche
U	$\mathbb{N}$	Number of *unlabelled* stem cells in the niche
$probLab$	$(0,1)$	Labelling efficiency - we choose open interval, as some cells are labelled, but not all
$p = probDiv$	$[0.5,1)$	Probability of division - probability that the stem cell which has just divided will divide again in the next time step
s	$\{0,1,\ldots,7\}$	Number of switches within a 8-filament long mini-arch
f	$\{1,2,\ldots,8\}$	Number of labelled filaments within a mini-arch

Table 6.1: Table summarising the meaning of each parameter used.

6.1 Probabilities computation

For simplicity, we rename $p = probDiv \in [0.5,1)$ and we choose $p \geq 0.5$ based on our previous hypotheses. A probability of division of $p = 0.5$ corresponds to an entirely functionally homogeneous system, in which the probability of the previously diving cell to divide again is equal to that of another random cell of the same type to take over this task. A probability of division $p > 0.5$ corresponds to our hypothesis of a functionally heterogeneous system with activation and quiescence phases. In addition, $p < 1$, since $p = 1$ would be equivalent to a case when the entire mini-arch is created by one stem cell, case which we proved infeasible - see Fig. 5.2 in Chapter 5.

Recall the formula

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \tag{6.1}$$

for conditional probabilities and suppose i is the cell which has just divided and j the cell about to divide. Further, denote by cL the event of choosing a labelled cell, and by cU the event of choosing an unlabelled cell. Then one can compute the probability that j is labelled or unlabelled, conditional on whether i was labelled or unlabelled, as in (6.2).

Conditional probabilities of cell labellings:

For $n \geq 2$:

$$pLL := \mathbb{P}(j \text{ labelled} | i \text{ labelled}) = \mathbb{P}(j = i) + \mathbb{P}(j \neq i)\mathbb{P}(cL) = p + (1-p)\frac{L-1}{n-1}$$
$$pUU := \mathbb{P}(j \text{ unlabelled} | i \text{ unlabelled}) = \mathbb{P}(j = i) + \mathbb{P}(j \neq i)\mathbb{P}(cU) = p + (1-p)\frac{U-1}{n-1}$$
$$pLU := \mathbb{P}(j \text{ labelled} | i \text{ unlabelled}) = \mathbb{P}(j \neq i)\mathbb{P}(cL) = (1-p)\frac{L}{n-1}$$
$$pUL := \mathbb{P}(j \text{ unlabelled} | i \text{ labelled}) = \mathbb{P}(j \neq i)\mathbb{P}(cU) = (1-p)\frac{U}{n-1},$$

For $n = 1$:

$$pLL := \mathbb{P}(j \text{ labelled} | i \text{ labelled}) = 1$$
$$pUU := \mathbb{P}(j \text{ unlabelled} | i \text{ unlabelled}) = 1$$
$$pLU := \mathbb{P}(j \text{ labelled} | i \text{ unlabelled}) = 0$$
$$pUL := \mathbb{P}(j \text{ unlabelled} | i \text{ labelled}) = 0 \tag{6.2}$$

Of course,

$$\mathbb{P}(j \text{ labelled}) = pLL\frac{L}{n} + pLU\frac{U}{n} = \left(p + (1-p)\frac{L-1}{n-1}\right)\frac{L}{n} + (1-p)\frac{L}{n-1}\frac{U}{n} = \frac{L}{n},$$
$$\mathbb{P}(j \text{ unlabelled}) = pUU\frac{U}{n} + pUL\frac{L}{n} = \left(p + (1-p)\frac{U-1}{n-1}\right)\frac{U}{n} + (1-p)\frac{U}{n-1}\frac{L}{n} = \frac{U}{n},$$
$$\mathbb{P}(j \text{ labelled} \cup j \text{ unlabelled}) = \frac{L}{n} + \frac{U}{n} = 1$$

as expected.

6.2 Stochastic simulations approach

One of the methods used for analysing the biological system was by performing stochastic simulations that produce a large set of *in silico* mini-arches as in the data set, for comparison.

In accordance with **Assumptions 6.0.1** and based on the conditional probabilities computed in (6.2), the algorithm shown in the flowchart from Fig. 6.1 was implemented. By using this algorithm we produce simulated *in silico* data set to compare to the available biological data, with the final goal of estimating the number of active stem cells needed for organ growth and maintenance, n, and the probability that the fate-restricted stem cell which has just divided will be the next one to divide again, p. The algorithm is outlined below.

1. The algorithm starts by computing the labelling efficiency for each cell type (i.e. stem cells of fates giving rise to each patter 1-4) according to the formula (5.1).
2. With these values at hand, the main part consisting of the stochastic simulations begins, which is visualized through the big green rectangle in the flowchart.
    a. For each mini-arch to be simulated, the program chooses random parameters n and p based on which the mini-arch will be filled with filaments, i.e. the 8-cell-long array will be filled with binary values.
    b. Out of the n stem cells of a particular fate, the number of labelled ones, L, is computed according to the Binomial distribution with parameters n and $probLab$, with a success representing the event of having a cell labelled.
    c. The remaining cells, U, are thus unlabelled. Having all the necessary information about the contents of our virtual niche, the mini-arch can be simulated as follows.
        i. One starts with an empty array representing the mini-arch before any filament has been generated.
        ii. The first stem cell to divide and generate a filament to be added to the mini-arch is chosen according to the amount of labelled and unlabelled stem cells in the niche. This means making a weighted random choice of whether to add a 1 or a 0 to the empty array, with weights given by probabilities of choosing a labelled cell $\frac{L}{n}$, or an unlabelled cell $\frac{U}{n}$, respectively.

iii. After the first entry in the array has been added, the following ones depend on the previously inserted value.

iv. Thus, if the previously added filament was labelled, the new one will also be labelled with a probability pLL, computed in (6.2). This happens either if the same (labelled) stem cell divides again, or if another labelled stem cells does.

v. The procedure is similar for all other cases corresponding to conditional probabilities pUL, pLU, pUU, and continues until the array is filled with 8 entries, thus generating one virtual mini-arch.

d. This part of the algorithm (i-v) is repeated 10^4 times to generate a large table of simulated mini-arches to compare with the available biological data set.

3. Once the simulated data set is obtained, the number of switches and the number of labelled filaments is determined for each arch, as it was previously done for the biological data.
4. Subsequently, the frequency of observing a certain pair (s, f) of switches and labelled filaments, out of the total 33 possibilities in the case of 8-filament long mini-arches, is calculated for both the simulation results and the biological data.
5. These frequencies are then used to construct an objective function similar to the one in (5.2), but this time for 8-filament long mini-arches instead of 6, for comparison with the frequencies obtained from the biological dataset (6.3).

$$f = \frac{\sum_{i=1}^{33} (fD_i - fS_i)^2}{33} \tag{6.3}$$

Here fD_i and fS_i are the frequencies in the data and the simulation results of observing pair $i \in \Pi$, with $\Pi = \{(0,0),(0,8),\ldots,(7,4)\}$ the set of the 33 possible pairs (s, f).

6. The algorithm concludes with the parameter estimation procedure for obtaining the best parameters n and p that provide the most accurate fit between the simulation results and the data, performed by minimizing the objective function. For this purpose the *Mathematica*® routine `NMinimize` was used with the Method "NelderMead", outlined in Section 2.4.

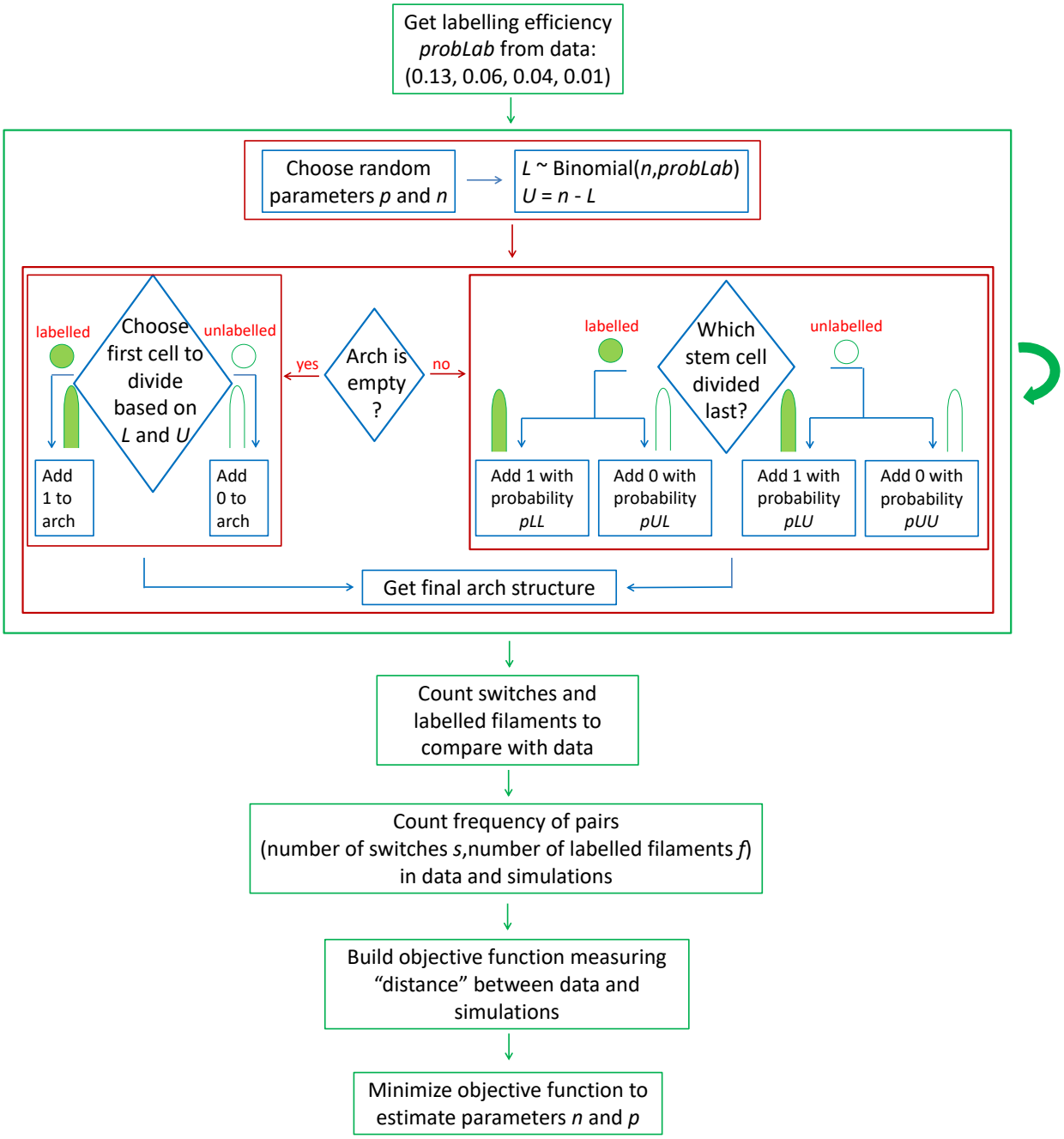


Fig. 6.1: Flowchart of the stochastic algorithm.

6.3 Markov Chains approach

The second approach employed Markov Chains theory to compute the exact probabilities of observing a certain pair (s, f), to be later used for comparison to the frequencies obtained from the biological data, and is described below.

On a closer look, one notices that the transitions from one filament to the next, i.e. the division of a cell based on the previous one, and depending on the respective label, can in fact be modelled by a two-step Markov Chain as seen in Fig. 6.2. This is true since, according to **Assumptions 6.0.1**, the outcome of the next cell division only depends on the current cell division, in agreement with the Markov Chain property. A broader description of Markov Chains theory has been outlined in Section 2.2.

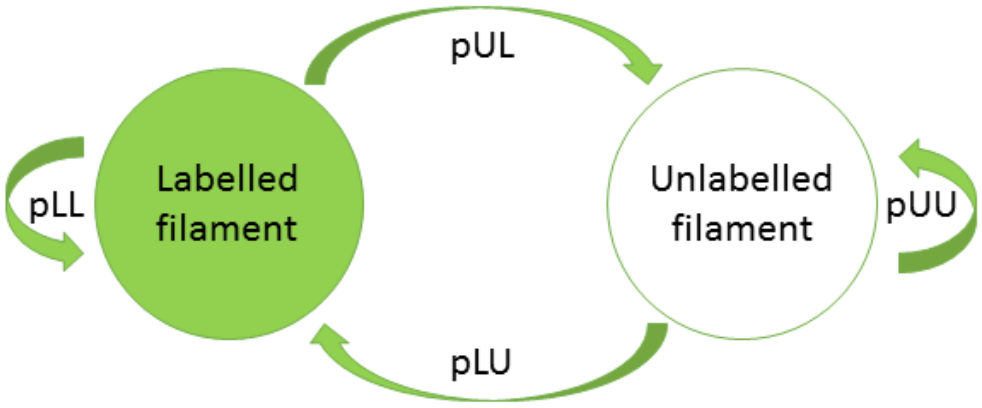


Fig. 6.2: Markov Chain transitions diagram: transitioning from a labelled to an unlabelled filament (with probability pUL) or viceversa (pLU) gives rise to a switch.

The Markov Chains approach can be used instead of the stochastic simulations approach for the purpose of determining the number of active stem cells of each fate residing in the niche. This time we compute the exact probabilities of observing a pair (s, f) of switches and labelled filaments rather than approximate frequencies, which is more reliable and much less computationally expensive, as it does not require long simulations of multiple trials of large data sets of *in silico* branchial arches.

Thus, the generation of filaments can be defined as a homogeneous stochastic Markov process $\{X_t\}$, with $t \in \mathscr{T} \subset \mathbb{N}$ the division step, $\mathscr{N} = \{0,1\}$ the state space corresponding to the labelling status of the filament and $P = (p_{ij})$ the probability transition matrix. Note that the division step is equivalent to the filament index, where indexing ranges from the oldest to the youngest filament.

Accordingly, the k^{th} step transition matrix $P^{(k)}$ with entries as in (6.4) can be computed, where the initial distribution is $\lambda = \left(\frac{L}{n}, \frac{U}{n}\right)$. The previous probabilities $pLL, \ldots, pUU$ will be renamed to $p_{11}, \ldots, p_{22}$ as entries of the probability transition matrix.

Probability transition matrix $P^{(k)}$ entries:

$$\begin{cases} p_{11}^{(k)} := pLL^{(k)} = \dfrac{L}{n} + \left(p - p\dfrac{L-1}{n-1} - \dfrac{U}{n(n-1)}\right)\left(p - \dfrac{1-p}{n-1}\right)^{k-1} \\ p_{22}^{(k)} := pUU^{(k)} = \dfrac{U}{n} + \left(p - p\dfrac{U-1}{n-1} - \dfrac{L}{n(n-1)}\right)\left(p - \dfrac{1-p}{n-1}\right)^{k-1} \\ p_{12}^{(k)} := pUL^{(k)} = \dfrac{U}{n} - \dfrac{U}{n}\left(p - \dfrac{1-p}{n-1}\right)^{k} \\ p_{21}^{(k)} := pLU^{(k)} = \dfrac{L}{n} - \dfrac{L}{n}\left(p - \dfrac{1-p}{n-1}\right)^{k}, \quad \text{for } n \geq 2 \end{cases} \tag{6.4}$$

Proposition 6.3.1 (Irreducibility)
Recall that $p \in [0.5, 1)$. The Markov Chain described by the transition matrix at the k-th step

$$P^{(k)} = \begin{pmatrix} p_{11}^{(k)} & p_{12}^{(k)} \\ p_{21}^{(k)} & p_{22}^{(k)} \end{pmatrix}$$

is irreducible for $L \geq 1$ and $U \geq 1$ and has a closed class otherwise.

Proof. First of all, if $n = 1$ (i.e. only one cell) then either $L = 1$ and $U = 0$ or viceversa. This means that our transition matrix $P = P^{(k)} = I_2$ the identity matrix (6.2)*, so one never jumps out of the starting state, thus producing a stretch of filaments of one label.*

Now let $n \geq 2$. Since a closed class implies no transitions out of the particular class, it suffices to inspect whether any of the elements $p_{12}^{(k)}$ and $p_{21}^{(k)}$ is equal to 0*.*

◇ *If* $L = 0$ *and* $U \geq 2$*,*

$$p_{12}^{(k)} = 1 - \left(p - \frac{1-p}{n-1}\right)^k, \qquad p_{21}^{(k)} = 0$$

but $\lambda = (0,1)$ *so we start in the unlabelled state and stay there.*

◇ *If* $U = 0$ *and* $L \geq 2$*,*

$$p_{12}^{(k)} = 0, \qquad p_{21}^{(k)} = 1 - \left(p - \frac{1-p}{n-1}\right)^k$$

but $\lambda = (1,0)$ *so we start in the labelled state and stay there.*

Next, set $L \geq 1, U \geq 1$*. Once again, one needs to investigate whether any of* $p_{12}^{(k)}$ *and* $p_{21}^{(k)}$ *is equal to* 0*. This is equivalent to*

$$1 - \left(p - \frac{1-p}{n-1}\right)^k = 0, \quad k \geq 1 \quad \Leftrightarrow \quad pn - 1 = n - 1 \quad \Leftrightarrow \quad p = 1$$

which is a contradiction with our problem setting. □

As previously stated, the main idea of the Markov Chains approach is to calculate the analytical probabilities of observing a certain pair $(s, f) \in \Pi$ of s switches and f labelled filaments in the 8-filament long mini-arch. Each such probability is computed as a sum of probabilities of possible Markov Chain trajectories producing the required number of switches and labelled filaments. For example the probabilities of producing an entirely unlabelled mini-arch (with 0 switches and 0 labelled filaments) and an entirely alternating mini-arch (with 7 switches and 4 labelled filaments) are computed below:

◇ $\mathbb{P}(s = 0, f = 0) = \mathbb{P}(00000000) = \mathbb{P}(X_1 = 0, X_2 = 0, \ldots, X_8 = 0) = \lambda_2 \cdot p_{22}^7$ or

◇ $\mathbb{P}(s = 7, f = 4) = \mathbb{P}(10101010) + \mathbb{P}(01010101) = \lambda_1 \cdot p_{12}^4 \cdot p_{21}^3 + \lambda_2 \cdot p_{12}^3 \cdot p_{21}^4$.

Similarly, this can be done for all the 33 possible pairs (s, f) to obtain a list of probabilities with entries corresponding to each such pair, depending on the number of stem cells in the niche n, the probability of division p and the labelling efficiency *probLab*. These analytical probabilities $\mathbb{P}(s, f)$ replace the frequencies fS_i of observing pair $i \in \Pi$ in the simulation results from the Stochastic simulations approach (Section 6.2). Therefore, not only are these probabilities exact and computationally cheaper to obtain, but they can be

used as in (6.3) to obtain an objective function that needs to be minimized for estimating the parameters n and p.

The two methods (Stochastic simulations and Markov chains approaches) produce the same results - see Section 6.4. Additionally, by using the Markov chains approach one can compute the expected hitting times. In other words, if one starts in state $i \in \{1,2\}$ (where 1 represents the labelled filaments state and 2 the unlabelled one), what is the expected number of steps after which the other state is reached? By using the methods described in Section 2.2 the following result was obtained.

Proposition 6.3.2 (Mean hitting times)
The expected time needed to hit the other state than the one in which the chain starts (i.e. $1 \to 2$ or $2 \to 1$) can be written as a function of the transition from the initial state to itself (i.e. p_{11} or p_{22}, respectively).

Proof. Let $k_i^j = \mathbb{E}$(time to hit j when starting in i). Then

$$\begin{cases} k_1^1 = 0 \\ k_2^1 = 1 + p_{22}k_2^1 \end{cases} \quad \textit{and} \quad \begin{cases} k_2^2 = 0 \\ k_1^2 = 1 + p_{11}k_1^2 \end{cases} \quad \Rightarrow \quad \begin{cases} k_1^2 = \dfrac{1}{1-p_{11}} = \dfrac{n-1}{(1-p)U} \\ k_2^1 = \dfrac{1}{1-p_{22}} = \dfrac{n-1}{(1-p)L} \end{cases} \tag{6.5}$$

□

Let us consider some examples:

- ◇ For $n = 2, L = 1, U = 1, p = 0.9$: $k_1^2 = k_2^1 = 10$ steps. Keeping in mind that we consider 8-filament long mini-arches, this result suggests the emergence of completely labelled or completely unlabelled mini-arches depending on the initial starting state.
- ◇ For $n = 2, L = 2, U = 0, p = 0.9$: $k_1^2 = \infty$ as wanted, since when only labelled stem cells reside in the niche the chain never transitions to an unlabelled state. On the other hand, $k_2^1 = 5$, but in fact one never starts in the unlabelled state so one can discard this case.
- ◇ Similarly, for $n = 2, L = 0, U = 2, p = 0.9$: $k_2^1 = \infty$ and the same reasoning as in the previous point applies.

In conclusion one can easily compute the mean hitting times by inspecting the transition matrix P or, equivalently, the values for parameters n, p and L.

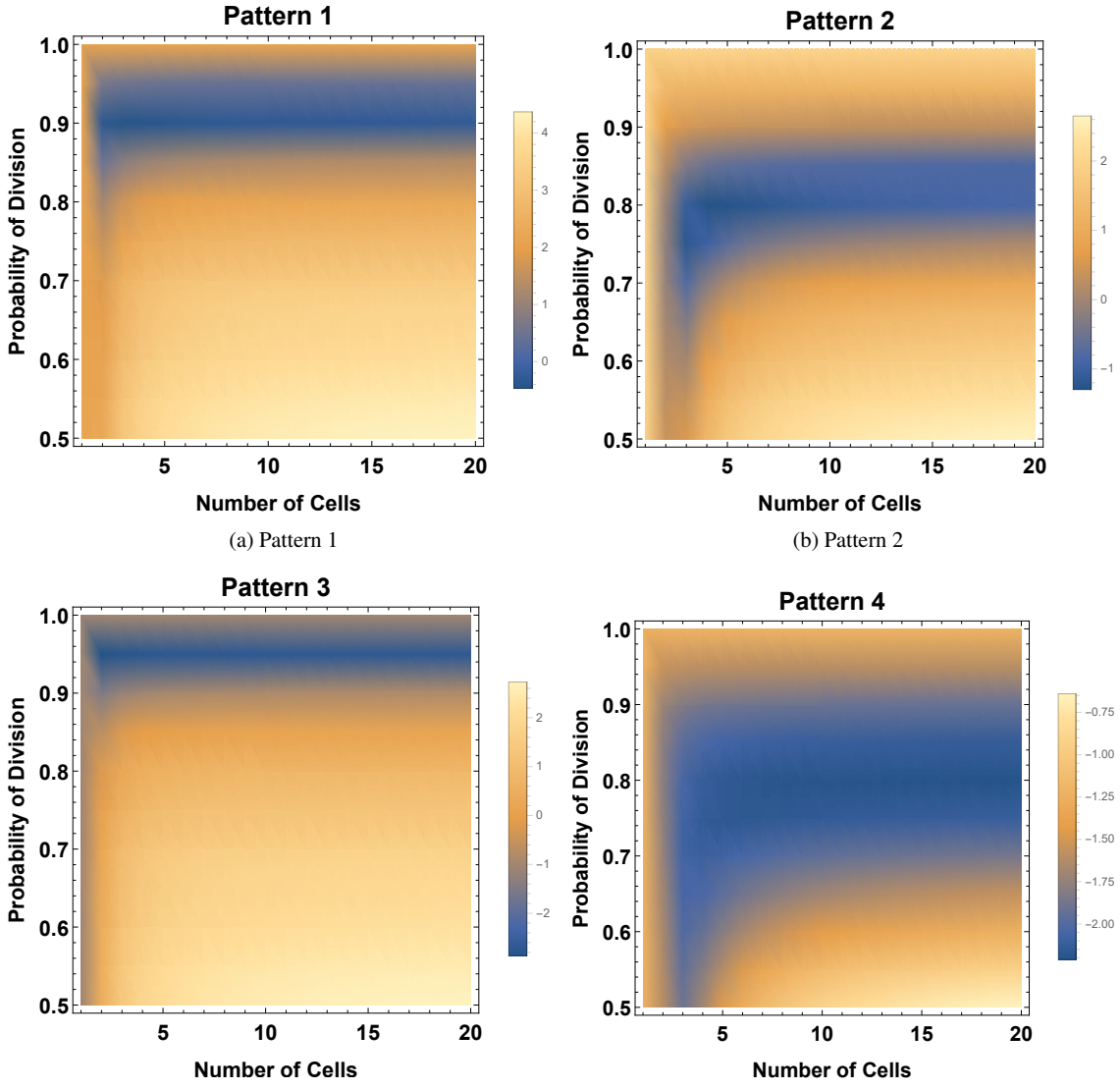
6.4 Parameter estimation and results of the heterogeneous model

For both the Stochastic Simulations (Section 6.2) and the Markov Chains (Section 6.3) approaches, one wishes to find the parameters n - number of stem cells in the niche and p-probability of division for which the model(s) provide the best fit to the data. In order to do so, a formal parameter estimation method was employed, namely an l_2-type objective function was built

$$f = \frac{\sum_{i=1}^{33} (fD_i - fS_i)^2}{33} \tag{6.6}$$

where fD_i are the frequencies in the data and fS_i the probabilities/frequencies resulted from the Markov Chains approach/Stochastic Simulations approach of observing pair $i \in \Pi$, with $\Pi = \{(0,0),(0,8),(1,1),\ldots,(7,4)\}$ the set of the 33 possible pairs (s,f).

To first get an overview of the best parameter sets for each pattern, a parametric sweep as seen in Fig. 6.3 was performed. Here the logarithms of the objective functions are plotted against the number of stem cells and the probability of divisions, with dark blue corresponding to low values while bright orange to high ones. In these plots one can easily notice the high importance of having a big probability of division in order to produce a good fit to the experimental data, especially in the case of the first three patterns. Pattern 4 however shows a greater variation in the "best" value for the probability of division suggesting parameter non-identifiability. To mention here and take into account is the fact that this aspect is not necessarily the outcome of our approaches but instead, in part, as a result of the data - practical non-identifiability. As previously mentioned, the labelling efficiency in the case of Pattern 4 is extremely low ($probLab = 0.01$), which means that irrespective of how many cells of fate 4 reside in the niche, the probability of having some labelled one(s) is very low. And of course, when all the cells are unlabelled we can only observe cases with no switches regardless of whether the probability of division is high or not. As a further comment on the low labelling efficiency of Pattern 4, we need to point out that in the cases of mixed overlapping patterns within the same filament, pattern 4 is difficult to observe in combination with any of the other patterns (since all cells carry a GFP label), so it can easily and unknowingly be overlooked, thus resulting in few cases with a labelled Pattern 4.



(a) Pattern 1

(b) Pattern 2

(c) Pattern 3

(d) Pattern 4

Fig. 6.3: Parametric sweep plots representing the logarithms of the objective functions with respect to the number of cells n and probability of division p. The smaller the objective function, the better fit to the data. Hence, the coordinates in the (n, p) plane of the darkest point represent the best parameter values. Subfigures (a)-(d) correspond to patterns 1-4.

Once the probabilities of division have been set to a big value (e.g. approximately 0.9 for Pattern 1), the number of stem cells can vary. Nevertheless, we are interested in *the number of stem cells needed for building the organ*, which means we are looking for the minimum number of stem cells n for each pattern. As seen in the parametric sweep plots, these minima correspond to surprisingly low values: 2-3 cells depending on the cell-type (1-4).

Now, in order to determine the best parameter sets for each pattern, for both approaches the objective functions were minimized in *Mathematica*® using the `NMinimize` routine, by employing the "Nelder-Mead" method, previously described in Section 2.4. As one would expect after having studied the sweep plots, the parameter sets contain low values for the numbers of cells n and high values for the probabilities of division p, for each pattern. This can be seen in Fig. 6.4 which depicts the fit for the estimated parameters, for each pattern. The plots use the same parameter values for both the Markov Chains approach, which is based on exact probabilities of observing a pair (s, f), and the Stochastic Simulations approach which uses approximate frequencies of finding such a pair. Even if the Stochastic Simulations approach is approximate, by producing a large number of trials, the frequencies get very close to the exact probabilities (Table 6.2) and thus the fits from subfigures 6.4(a) and (b) are almost indistinguishable.

In summary, the minimization results suggest that only *few active stem cells* which have *large probabilities of division* are needed for building the respiratory organ of fish. This enforces our hypothesis of functional heterogeneities among the stem cells. A step by step summary of how the biological system behaves according to our hypothesis is outlined below.

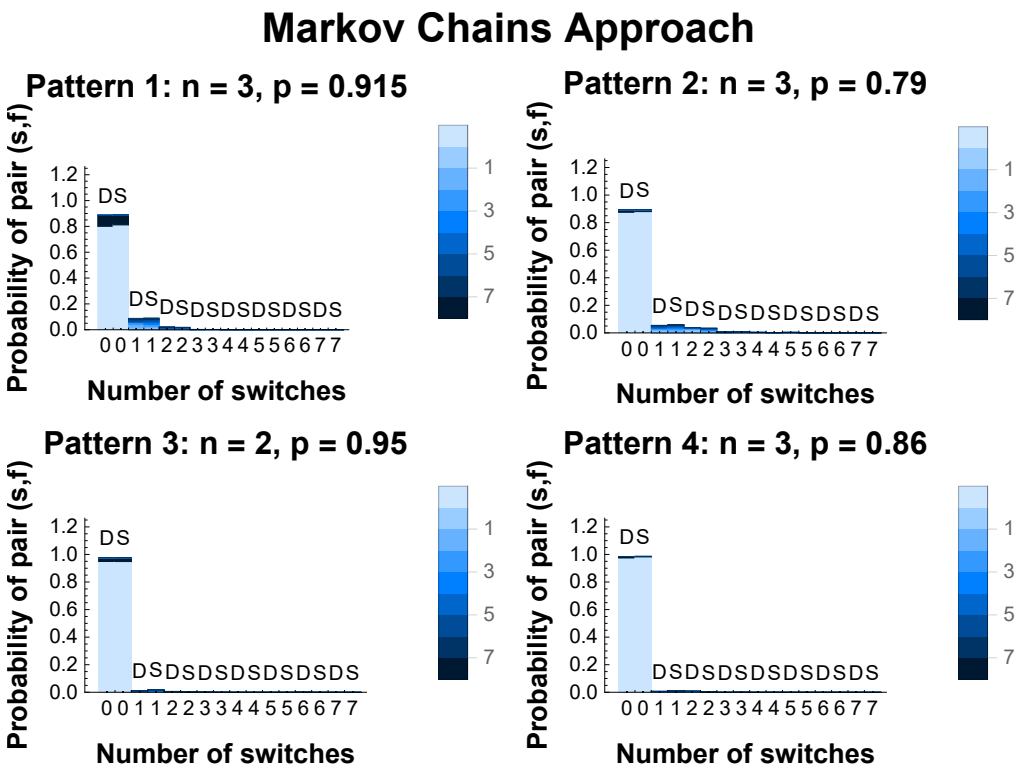
- ◇ When the post-embryonic development begins, and after (some of) the cells in the niches have been labelled, a first post-embryonic filament needs to be generated and hence a fate-restricted cell should be chosen for division from among its neighbours of the same fate. The initial choice is random.
- ◇ Once a cell A has been chosen and has generated the first filament, it becomes active and thus its probability to divide again next highly increases. It continues to produce filaments as long as it is active, thanks to its high probability of division.

◇ At some point, it may happen that a new cell B of that same type is selected for division. Naturally, it gets activated and with its new high probability of division it generates filaments until it gets "overthrown" from its duty, thus becoming quiescent. If B possesses the same label as A, it simply continues the streak of filaments with that label until it turns quiescent, otherwise it produces a switch.

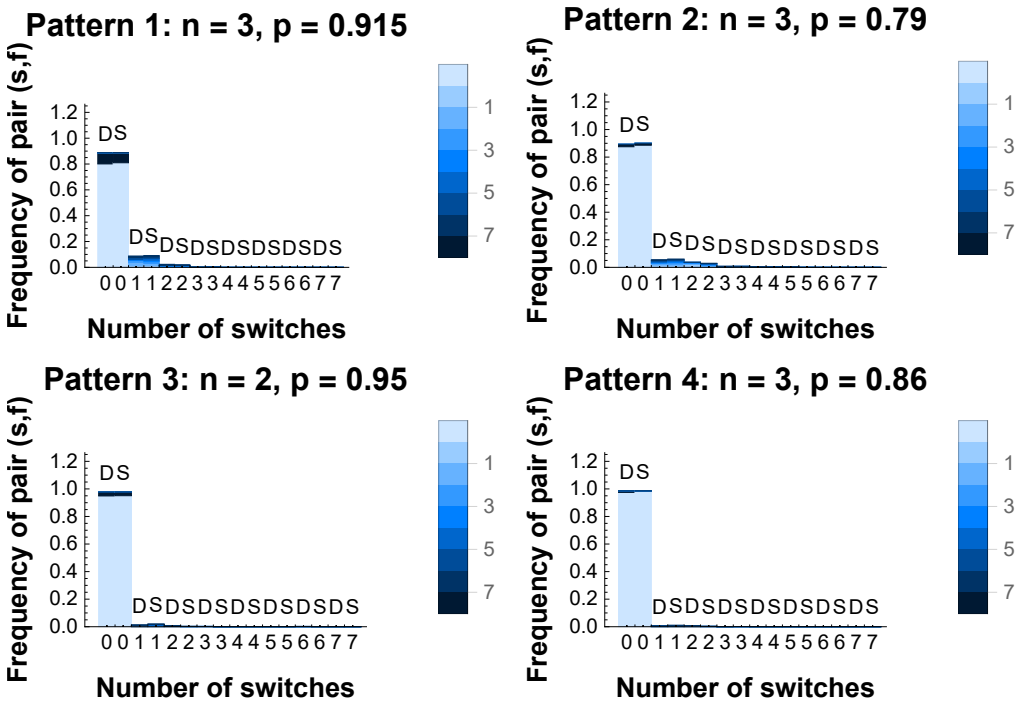
◇ This process of subsequently activating cells continues throughout the fish life, keeping in mind their permanent growth.

Markov Chains Probabilities - Pattern 1									
s / f	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	0.8	-	-	-	-	-	-	-	0.085
1	-	0.013	0.025	0.015	0.003	0.004	0.014	0.011	-
2	-	0.006	0.002	0.002	0.004	0.002	0.	0.008	-
3	-	-	0.	0.	0.	0.002	0.	-	-
4	-	-	0.	0.	0.	0.	0.	-	-
5	-	-	-	0.	0.	0.	-	-	-
6	-	-	-	0.	0.	0.	-	-	-
7	-	-	-	-	0.	-	-	-	-
Stochastic Simulations Frequencies - Pattern 1									
s / f	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	0.811	-	-	-	-	-	-	-	0.076
1	-	0.014	0.024	0.013	0.013	0.012	0.012	0.012	-
2	-	0.003	0.003	0.003	0.002	0.002	0.	0.002	-
3	-	-	0.	0.	0.	0.002	0.	-	-
4	-	-	0.	0.	0.	0.	0.	-	-
5	-	-	-	0.	0.	0.	-	-	-
6	-	-	-	0.	0.	0.	-	-	-
7	-	-	-	-	0.	-	-	-	-

Table 6.2: Probabilities of observing pair (s, f) in the case of Pattern 1, computed with the Markov Chains approach and Stochastic Simulations approach for the estimated parameters $n = 3$, $p = 0.915$, for comparison. Note that for the later, from the pair $(s, f) = (3, 2)$ onwards the 0. values are rounded from values multiplied with 10^{-m}, $m \geq 4$, going all the way to $m = 10$ for pair $(s, f) = (7, 4)$.



(a) Markov Chains approach



(b) Stochastic Simulations approach

Fig. 6.4: (Caption next page)

Fig. 6.4: (Figure previous page) Comparisons between data (D) and simulations (S) for the Markov Chains Approach (a) and the Stochastic Simulations Approach (b) for each of the four patterns. x-axis represents the number of switches and the colour code corresponds to the number of labelled filaments (light blue - zero, dark blue - eight).

Returning to the notion of expected hitting times, it is interesting to observe what values these produce, now that we have our optimized parameters. Recall the previously determined formulae:

$$\begin{cases} k_1^1 = 0 \\ k_2^1 = 1 + p_{22} k_2^1 \end{cases} \quad \text{and} \quad \begin{cases} k_2^2 = 0 \\ k_1^2 = 1 + p_{11} k_1^2 \end{cases} \quad \Rightarrow \quad \begin{cases} k_1^2 = \dfrac{1}{1-p_{11}} = \dfrac{n-1}{(1-p)U} \\ k_2^1 = \dfrac{1}{1-p_{22}} = \dfrac{n-1}{(1-p)L} \end{cases}$$

where k_i^j is the mean hitting time of state j if we started in state i, with 1 corresponding to a labelled and 2 to an unlabelled state. Recall that in our case, time is discrete and corresponds to division events. However, we must accept non-integer values for the mean hitting times because, as the name suggests, they are averages of expected division steps. Based on these equations and using the optimized parameter values we get the following hitting times for each pattern.

Mean Hitting Times for each Pattern:

Pattern 1	L=1	L=2
$1 \to 2$	11.76	23.53
$2 \to 1$	23.53	11.76

Pattern 2	L=1	L=2
$1 \to 2$	4.76	9.52
$2 \to 1$	9.52	4.76

Pattern 3	L=1
$1 \to 2$	20
$2 \to 1$	20

Pattern 4	L=1	L=2
$1 \to 2$	7.14	14.28
$2 \to 1$	14.28	7.14

In inspecting these values, one should also keep in mind the values for the labelling efficiencies for each pattern. For example, by forgetting about these one would at first

be surprised by the fact that in case of Pattern 4, if we have one unlabelled cell out of the total of three stem cells, the expected number of divisions between an unlabelled and a labelled cell is 7 since our data barely has any case with switches. But if we account for the labelling efficiency of 0.01, we immediately realise that the cases of having some labelled cells in the niche are scarce. Thus, if then all cells are unlabelled we obviously have an infinite hitting time for the transition $2 \to 1$. Note that these tables do not show the cases when all cells in the niche are of the same label because a transition could never happen. Further the transitions of type $i \to i$, $i \in \{1,2\}$ are not presented because their mean hitting times are always zero. These results further our understanding of why such few cases with switches are observed.

Chapter 7
Conclusion and Discussion for the Stem Cell Numbers Modelling

In this part we have been investigating the fascinating question of how many active stem cells are needed to produce and maintain an organ. We have chosen to approach this question in a fish model organism, namely the *Oryzias latipes* Japanese rice fish also known as medaka. Fish, due to their permanent growth throughout life, possess adult stem cells which not only maintain homeostasis but are also responsible for post-embryonic growth, making them totally different from adult stem cells in mammals and higher vertebrates and in fact more similar to embryonic stem cells in these organisms. Therefore, these fish adult stem cells divide rapidly and often in order to keep up and scale the organ in accord with the organism growth, thus being perfectly suited for studying post-embryonic organ development.

We have chosen to study these aspects in the respiratory organ of the fish, known as branchia or gill. This choice was made because of the modular spatio-temporal organisation of the organ which recommends it as nicely suited for such an investigation. Gills are built from four pairs of double-sided branchial arches, which in turn consist of a sequence of filaments. The data recorded the distribution of green fluorescent protein labelling in filaments within a branchial arch, in 36 viable gills. This was recorded for each pattern created by labelling of one of the four fate-restricted stem cell types from the niche.

The study has been approached *via* two stochastic methods which reassuringly produced agreeing results. First, the experimental data have been calibrated into a form fitted for comparison with the models. In particular, the arches were split into 8-filament long mini-arches, for each of which the frequencies of observing certain numbers of switches

and labelled filaments were computed - one frequency per pair (s, f) of s switches and f labelled filaments, per pattern.

For the modelling part, on one hand, a stochastic algorithm adapted from that of D. Gillespie has been developed in order to produce a large set of *in silico* virtual mini-arches for which the respective analogue frequencies were computed. This algorithm has been rerun with various values for the two parameters which had to be determined and, by employing parameter estimation methods with the help of *Mathematica*® routines, best fit values for the number n of active stem cells in the niche and for the probability p of a cell to divide again have been found.

On the other hand, Markov Chains theory has been used to compute the probabilities of observing a certain pair (s, f) for specific parameter values n and p. As opposed to the stochastic simulations, the Markov Chains approach compute exact probabilities instead of approximate frequencies. In addition, due to actual formulae for the said probabilities, these computations are immediate in contrast to the previous case where much more time is needed for determining the frequencies. As before, the probabilities had to be computed for various values of parameters, within the parameter estimation algorithm, and finally the tuple (n, p) that best fits the model to the data had been found for each pattern.

Even though the Markov Chains approach is much more advantageous in terms of computational time and accuracy, a relative advantage of the Stochastic Simulations approach is that, by actually simulating *in silico* mini-arches as in the dataset, it is more intuitive to model and work with. It is a good starting point in planning how to approach the mathematical problem for understanding the biological system, before switching to a more accurate method such as Markov Chains theory.

Indeed, the results produced by the two approaches lead to the same conclusion: for the post-embryonic growth of a gill, only few active stem cells (of each fate) are needed in the niche and, more importantly, these fate-restricted stem cells are functionally heterogeneous, in the sense that they take turns in getting activated. Once a stem cell divides it becomes active and generates multiple filaments before it goes back to quiescence and allows another cell of the same type to become active and take over its task.

For a stronger validation of the models, more data would be of highly use. In particular, in the case of Pattern 4, we observe practical non-identifiability for the probability of division. This is a result of the low labelling efficiency obtained for this pattern. The

low labelling efficiency can also be a result of the difficulty of observing this pattern in combination with other ones (i.e. mixed patterns), in which Pattern 4 can be hidden beneath other more striking patterns such as 1 and 2. These issues could be avoided if, for example, cell-type specific markers existed for each of the four stem cell types, or at least for the fourth type. If this were the case, different colours could be used for different patterns and the problems of not observing Pattern 4 would vanish.

The investigation on the number of active fate-restricted stem cells in the niche is the first step in understanding the coordination of the four lineages during filament generation. How do these four different cell types get recruited together to a newly forming filament, since this process should be quite robust considering that it takes place hundreds of times during the life of a medaka fish? Further, how is their activity coordinated in order to maintain the ratio of cell types in the filaments? Investigating such avenues could extend our understanding of the activity of mammalian cancer stem cells which possess similar properties to adult stem cells in fish, as they have the ability to both drive growth and maintain homeostasis. Such an overall understanding of fish stem cells could also have possible applications in wound healing and organ regeneration in higher vertebrates, processes which are driven by stem cells directed back into active and rapid proliferation due to injury and stress conditions.

Part II

Mathematical Modelling of Branchia Growth and Shape

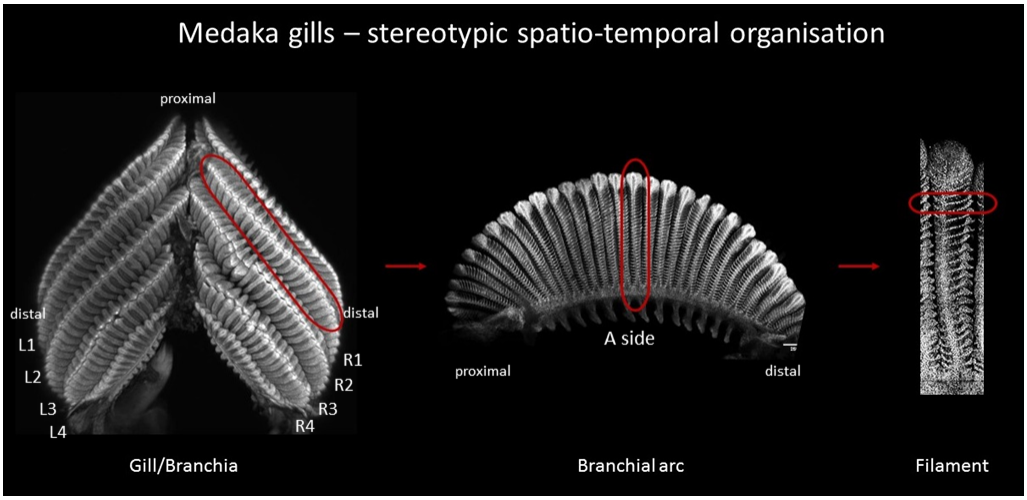
Recalling the curved shape of the branchial arches, one question which arises refers to how and why this shape is attained. Investigating the development of branchial arches into their curved shape helps uncover properties of and relations between the two types of stem cells responsible for driving growth along the two orthogonal axes. This study is further useful for understanding coordination between growth stem cells and homeostatic stem cells during post-embryonic organ development. The study is done by employing mathematical modelling methods such as ordinary differential equations describing the development of each filament, coupled *via* indicator functions to an algebraic equation determining the moments of filament generation. The mathematical and numerical results of generated *in silico* branchial arches are then compared to the experimental data. The results of the investigation suggest a decay in the proliferation of filament stem cells with filament elongation, and also in the proliferation of branchial arch stem cells dependent on the number of generated filaments.

Chapter 8
Experimental Procedure and Data Acquisition

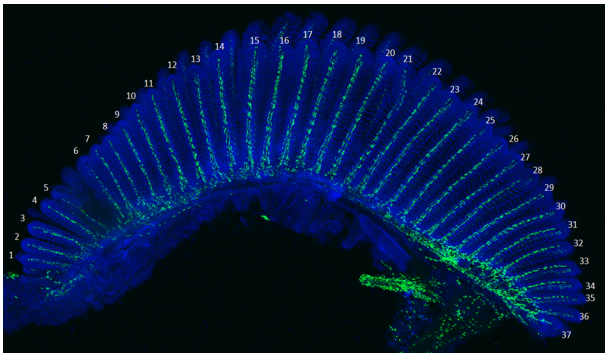
This chapter describes the experimental procedures used in acquiring the data used for the mathematical modelling.

For the purpose of studying the organ shape, the data personally gathered from the experimental images provided by the group of Prof. Dr. Lazaro Centanin (Centre for Organismal Studies, Heidelberg University) recorded lengths of filaments, at one time point, of branchial arches from multiple gills of fish of similar lengths (Fig. 8.1). For obtaining this data, gills from multiple fish were removed and stained with antibodies against GFP and DAPI. Subsequently, the branchial arches from these gills were separated and mounted on 96 circular well microplates with glycerol. Finally, they were imaged under the microscope. A total of 12 fish were imaged, which resulted in data consisting of $12 \times 8 \times 2 = 192$ sides of branchial arches.

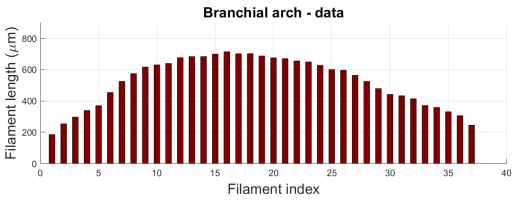
Each branchial arch is composed of two rows of intercalated filaments growing from the same base. These are referred to as sides A and B of the same branchial arch. The two sides do not necessarily have the same number of filaments, and these filaments can have different lengths. We quantified the filament lengths using the *ImageJ* software [79, 80], and recorded those by carefully keeping track of distal/proximal extremes and A/B sides. The proximal extreme is the one where the branchial arch is attached to the gill mid-axis, while at the distal side the branchial arch is free, unattached. By inspecting the data, we observed that in general for branchial arches R1, R2, L1 and L2 the A side has more and longer filaments than side B, while for branchial arches R4 and L4 the opposite is true. The two sides are more or less similar for branchial arches R3 and L3.



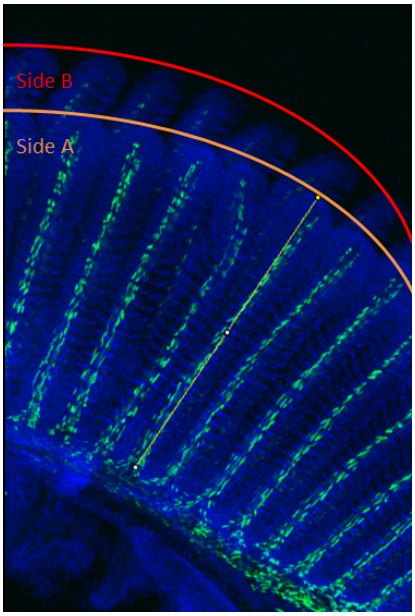
(a) Gill structure



(b) Experimental image



(c) Data



(d) A/B sides

Fig. 8.1: (a) Structure of a gill, revisited. (b) Confocal microscope image (by Julian Stolper, Centanin group, Heidelberg University) of the A side of a branchial arch with indexed filaments from proximal to distal extreme. Here, the green expression indicates that the stem cells generating Pattern 4 are labelled in this branchial arch. The GFP labelling is unnecessary when gathering the data, however the image serves as a good example. (c) Filament lengths (in μm) extracted from the image above (b), counted from proximal to distal extremes. One can observe that the distal half-arch is bigger (in number of filaments) and more linear than the proximal half-arch. (d) Each branchial arch consists in fact of two rows of filaments growing from the same base. These can have different numbers and lengths on each side (A and B). In the figure, the yellow line depicts the length of one filament from the A side, but one can observe behind the corresponding longer filament from side B. The orange curve approximately outlines the shape of side A, while the red curve that of side B.

Furthermore, under a closer inspection of the branchial arches, one can easily notice the skewness of the arch. This asymmetry comes from having more and shorter filaments in the distal half-arch compared to fewer and longer filaments in the proximal half-arch. In other words, the distal part has a bigger curvature, while the proximal side is almost linear. This is generally the case in bigger arches (R/L1-2), while the asymmetry is not as visible in smaller arches (R/L3-4).

Chapter 9
Mathematical Model Derivation

"With four parameters I can fit an elephant,
with five I can make him wiggle his trunk."
- John von Neumann

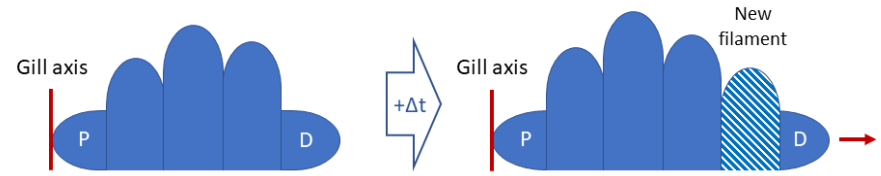
This chapter presents the various models developed for studying the behaviour of stem cells during the development of branchial arches by analysing properties of the arch shapes. The assumptions for each model are presented with various possible hypotheses suggesting these assumptions.

9.1 General considerations

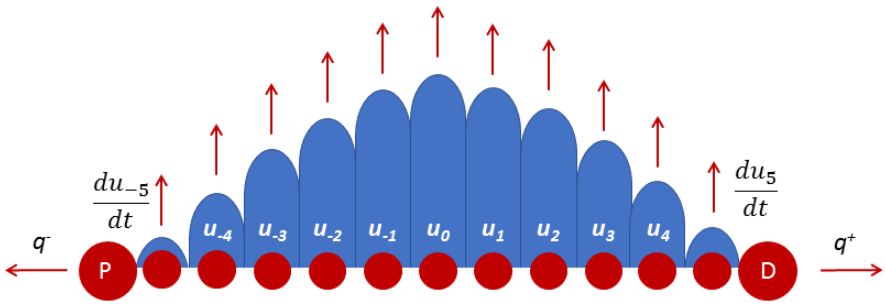
In order to study the postembryonic growth of branchial arches, various mathematical models consisting of a system of ordinary differential equations were designed. Each equation describes the growth of one filament according to the divisions of filament stem cells (*filSCs*) found at the filament tips, but since the shape of the arch is to be analysed, indicator functions were used to determine whether a new filament should be generated at a certain time point. This addition of filaments *via* the indicator functions relies on the proliferation of branchial arch stem cells (*brSCs*), found in the peripheral niches, which can either be constant or time- or space-dependent, as will be presented below. The peculiarity of the models comes from the indicator function approach used to deal with the discrete growing domain.

Recall that the proximal extreme of the branchial arches are attached to the gill mid-axis, thus having a fixed position. Therefore, when new filaments are added at each of the proximal and/or distal sides, the distal niche is pushed away from the gill axis. For simplicity reasons, in the following models the niches are assumed to move away from

each other, being pushed away from the "centre" of the arch where the oldest filament grows, instead of having one niche fixed and the other one pushed away approximately twice as much (Figure 9.1).



(a) Distal niche is pushed



(b) Niches move away from each-other

Fig. 9.1: (a) Figure explaining how the distal niche (D) is pushed away from the proximal one (P) as new filaments are added. (b) Image exemplifying some of the general model assumptions. Niches start at point 0 at time t_0 where the oldest and longest filament will grow, and move away from each-other with speeds q^+ and q^-, depositing filament-generating stem cells on their trail. Filaments are indexed with positive integers towards the distal niche (D) and with negative integers towards the proximal one (P). Each filament i grows according to its ODE $\frac{du_i}{dt}$, depending on the model.

9.2 General model

All the models presented in this chapter have the form of (9.2), with their assumptions summarised below:

General Model

Assumptions 9.2.1

- ◇ *Filaments $i \in \mathbb{Z}$ grow independently of each-other, having lengths $u_i : [0,\infty) \to \mathbb{R}$, which change in time.*
- ◇ *Filaments are indexed with integers $i < 0$ for the proximal side and $i \geq 0$ in the distal half (Figure 9.1b).*
- ◇ *Functions $g^{\pm} : [0,\infty) \to \mathbb{R}$ govern the proliferation rates $q^{\pm}$ of stem cells in the proximal(-) and distal(+) niches. They can be viewed as the velocities with which the niches move away from the origin, where the oldest filament lies: $q^-(t) < 0, q^+(t) \geq 0,\ \forall t \in [0,\infty)$.*
- ◇ *Functions $f_i^{\pm}(t) : [0,\infty) \to \{0,1\}$ indicate if filament i has been added to the arch at time t. Note $f_i^+(t) = 0$, for $i < 0, \forall t \in [0,\infty)$, and $f_i^-(t) = 0$, for $i \geq 0, \forall t \in [0,\infty)$. A filament $i \in \mathbb{Z}$ is added when the respective niche gets to that specific position, i.e. $q^{\pm}(t)t = i$. We define by $t_i = \frac{i}{q^{\pm}(t_i)}$ the time at which filament i is generated.*
- ◇ *Once a filament $i \in \mathbb{Z}$ has been added, i.e. $f_i^+(t) + f_i^-(t) = 1$, it grows at a rate $\dfrac{du_i(t)}{dt} = g(t,u_i(t))$, with $g : [0,\infty) \times \mathbb{R} \to \mathbb{R}$. Note $u_i(t) = 0$, $t < t_i$ and u_i starts growing with rate $g(t,u_i(t))$ once $t \geq t_i$.*

$$
\begin{aligned}
&\frac{du_i(t)}{dt} = g(t,u_i(t))\left(f_i^+(t) + f_i^-(t)\right), \quad u_i(t_i) = 0 \\
&\text{with} \qquad f_i^+(t) = \begin{cases} 1, & q^+(t)t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \\
&\qquad\qquad f_i^-(t) = \begin{cases} 1, & q^-(t)t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \\
&\text{and} \quad q^{\pm}(t) = g^{\pm}(t),
\end{aligned}
\tag{9.2}
$$

where $g^{\pm}$ are constant or time-dependent (directly or indirectly), giving the behaviour of velocities $q^{\pm}$.

The differences among the following models lie in the form of the functions $g(t,u_i(t))$ and $g^{\pm}(t)$. Thy can be split into two categories, with $p>0,\ \alpha,\beta^{\pm}\geq 0,\ r^{+}>0, r^{-}<0$:

- ◇ Exponential decay models: Sections 9.3-9.5 - here $g(t,u_i(t)) = pe^{-\alpha u_i(t)}$ and $g^{\pm} = \dfrac{r^{\pm}}{1+\beta^{\pm}t}$. These can be simplified by having $\alpha=\beta^{\pm}=0$ (Section 9.3) or only $\beta^{\pm}=0, \alpha>0$ (Section 9.4). They will be presented in the order of their growing complexity.
- ◇ Negative-feedback models: Sections 9.6-9.7 - here $g(t,u_i(t)) = \dfrac{p}{1+\alpha u_i(t)}$ and $g^{\pm}$ time-dependent (directly, or indirectly *via* space dependence), as we shall see.

A challenge and relative limitation of the model comes from the combination of discrete and continuous properties. We have a discrete number of filaments (compartments), each of which is added at a specific time point and grows in a continuous manner. Thus, we are dealing with a discrete growing domain, which means we cannot use partial differential equations (as growing domains are usually approached) if we want to be able to study independent filaments. So we use indicator functions for adding new compartments to the system at certain time points. In some of the models we can analytically compute the times at which new compartments are introduced in the system, but in some others the times have to be determined numerically, which means that these particular time points must lie on the grid used for differentiation.

9.3 Linear Growth Model

In this first model scenario, filaments are assumed to be generated according to the divisions of *brSCs* in the niches at the basal peripheries of the arches, which proliferate at constant rates. Filaments elongate based on the constant proliferation rate of the *filSCs* found at the tips of filaments.

Linear Growth (LG) Model

Assumptions 9.3.1

- ◇ *Filament $i \in \mathbb{Z}$ is generated at time $t_i = i/q^{\pm} \geq 0$ and grows according to the constant proliferation $p > 0$ of* filSCs.
- ◇ *The* brSCs *in the distal and proximal niches divide with constant rates $q^+ > 0$ and $|q^-|$ ($q^- < 0$), respectively.*

$$\begin{aligned} &\frac{du_i(t)}{dt} = p\left(f_i^+(t) + f_i^-(t)\right), \quad u_i(t_i) = 0 \\ &\text{with} \quad f_i^+(t) = \begin{cases} 1, & q^+ t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \\ &\qquad\quad f_i^-(t) = \begin{cases} 1, & q^- t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \end{aligned} \tag{9.3}$$

9.4 Elongation-Decay Model

In this second model a decay in the proliferation of the stem cells responsible for elongating the filaments is introduced. This decay depends on the length of the filament in the sense that the longer the filament is, the slower it grows. This decay is plausible in view of the following hypotheses:

- ◇ Filament stem cells proliferation slows down with their ageing, and for simplicity this can be formulated in terms of the filament length since older *filSCs* reside at the tips of longer filaments.
- ◇ Filaments decrease their growth because of nutrient intake. Recall that blood vessels transporting nutrients can be found along the filament axis [10, 81, 82], as seen in Figure 9.2. In this scenario, the longer the filament is, the longer it takes for the stem

cells at the tips to receive the needed amount of nutrients for them to divide. This is the case because nutrients are consumed by all the other stem cells found along the filament axis.

◇ Similarly, filaments can slow down their growth due to the transport of oxygen through the filaments and lamellae.

Elongation-Decay (ED) Model

Assumptions 9.4.1

◇ *Filament $i \in \mathbb{Z}$ is generated at time $t_i = i/q^{\pm} \geq 0$ and grows according to the proliferation $\hat{p} = pe^{-\alpha u_i(t)}$ of* filSCs*, which decays with the length of filaments.*

◇ *The* brSCs *in the distal and proximal niches divide with constant rates $q^+ > 0$ and $|q^-|$ ($q^- < 0$), respectively.*

$$\begin{aligned} \frac{du_i(t)}{dt} &= pe^{-\alpha u_i(t)} \left(f_i^+(t) + f_i^-(t) \right), \quad u_i(t_i) = 0 \\ \text{with} \quad f_i^+(t) &= \begin{cases} 1, & q^+ t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \\ f_i^-(t) &= \begin{cases} 1, & q^- t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \end{aligned} \tag{9.4}$$

and $q^{\pm}$ constant, $p, \alpha > 0$, $t_i = \frac{i}{q^{\pm}}$, based on the sign of index i.

9.5 Elongation-Generation-Decay Model

In this third model we further introduce decays in the proliferation of the *brSCs* responsible for generating filaments. This decay depends on time such that the older the arch

is, the slower it generates filaments, and is modelled *via* a negative feedback. This is of course equivalent to saying that the more filaments the arch stem cells have produced, the slower they generate new ones.

These assumptions take us a step further, by considering that, similarly to filament stem cells, also branchial arch stem cells residing in the peripheral niches either age, or slow down their proliferation due to slower nutrient uptake, or oxygen transport. Since filament generation corresponds to discrete events, another plausible hypothesis is related to the number of divisions that stem cells have performed so far. The concept of "stem cells counting" is considered but disputed in the scientific community of stem cell research, but it fits with our model assumptions since the older the stem cells are, the more filaments they have produced and thus the slower they proliferate.

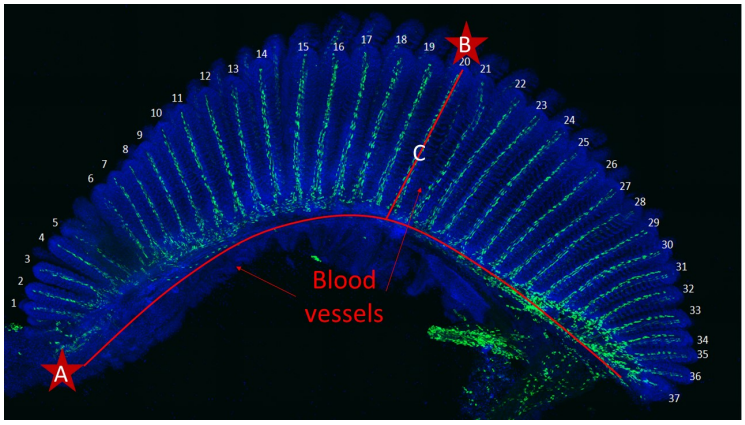


Fig. 9.2: Confocal microscope image of a branchial arch with filaments indexed from distal to proximal extreme. Blood vessels are depicted by red curves. A, B and C correspond to *brSCs* found in the peripheral niches, *filSCc* residing at the tips of filaments, and homeostatic stem cells found along the filament mid-axis, respectively. The green expression indicates that the stem cells generating Pattern 4 are labelled in this branchial arch.

Elongation-Generation-Decay (EGD) Model

Assumptions 9.5.1

- ◇ *Filament $i \in \mathbb{Z}$ is generated at time $t_i \geq 0$ and grows according to the proliferation $\hat{p} = pe^{-\alpha u_i(t)}$ of* filSCs*, which decays with the length of filament.*
- ◇ *The* brSCs *in the distal and proximal niches divide with rates $q^+(t) = \frac{r^+}{1+\beta^+ t}$ and $|q^-(t)| = \left|\frac{r^-}{1+\beta^- t}\right|$, respectively, both decreasing as time elapses.*

$$\begin{aligned} &\frac{du_i(t)}{dt} = pe^{-\alpha u_i(t)}\left(f_i^+(t) + f_i^-(t)\right), \quad u_i(t_i) = 0 \\ &\text{with} \quad f_i^+(t) = \begin{cases} 1, & q^+(t)t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \\ &\qquad\quad f_i^-(t) = \begin{cases} 1, & q^-(t)t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \end{aligned} \tag{9.5}$$

and $q^+(t) = \dfrac{r^+}{1+\beta^+ t}, \quad q^-(t) = \dfrac{r^-}{1+\beta^- t}, \ p, \alpha, \beta^\pm, r^+ > 0, \ r^- < 0,$
$t_i = \dfrac{i}{q^\pm(t_i)}$, based on the sign of the filament index i.

Remark 9.5.1 *Note that the EGD model is the general formulation of LG model ($\alpha = 0, \beta^\pm = 0, q^\pm = r^\pm$) and ED model ($\beta^\pm = 0, q^\pm = r^\pm$).*

9.6 Alternative Elongation-Generation-Decay Model

As an alternative to the EGD model, the right hand side $g(t, u_i(t))$ of the differential equation can be replaced by a negative feedback.

Alternative Elongation-Generation-Decay (AEGD) Model

Assumptions 9.6.1

- ◇ *Filament $i \in \mathbb{Z}$ is generated at time $t_i \geq 0$ and grows according to the proliferation $\hat{p} = \dfrac{p}{1+\alpha u_i(t)}$ of the* filSCs.
- ◇ *The* brSCs *in the distal and proximal niches divide with rates $q^+(t) = \frac{r^+}{1+\beta^+ t}$ and $|q^-(t)| = \left|\frac{r^-}{1+\beta^- t}\right|$, respectively.*

$$\begin{aligned} \frac{du_i(t)}{dt} &= \frac{p}{1+\alpha u_i(t)}\left(f_i^+(t)+f_i^-(t)\right), \quad u_i(t_i)=0 \\ f_i^+(t) &= \begin{cases} 1, & q^+(t)t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \\ \text{with} \quad f_i^-(t) &= \begin{cases} 1, & q^-(t)t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \end{aligned} \tag{9.6}$$

and $q^+(t) = \dfrac{r^+}{1+\beta^+ t}, \quad q^-(t) = \dfrac{r^-}{1+\beta^- t}, \ p, \alpha, \beta^\pm, r^+ > 0, \ r^- < 0,$ $t_i = \dfrac{i}{q^\pm(t_i)}$, based on the sign of the filament index i. To avoid singularities, we impose $u_i \neq -1/\alpha$, which holds - as we shall prove later.

This model can also be in agreement with ageing of stem cells or nutrient consumption, but in addition, it could result from a hypothesis of signalling molecules ($S(t)$) being produced (at a constant rate γ) and degraded (at a constant rate δ) within the arches, as well as being eliminated proportionally to the lengths of filaments (at a constant rate μ), which is equivalent to the density of cells within a filament. This process can be described [83, 28] by equation

$$\frac{dS}{dt}(t) = \gamma - \delta S(t) - \mu S u_i(t),$$

which following a quasi-state approximation resulting from the assumption that the process is fast in comparison to filament elongation (i.e. stem cells cell cycle duration) leads to:

$$s(t) = \frac{1}{1+\alpha u_i(t)},$$

where $s(t) := \frac{\mu}{\gamma} S(t)$ and $\alpha := \frac{\delta}{\mu}$. The validity of the quasi-state approximation has been rigorously proven [84].

9.7 Space-Dependent Elongation-Generation-Decay Model

Another variant of the AEGD model sets the velocities $q^{\pm}(t)$ as space-dependent instead of time-dependent. This would mean that the more filaments there are between the two niches, the slower the stem cells within proliferate. In particular, we set *brSCs* to increase their cell cycle length, the more filaments there are. Even though this is an intuitively easy concept, putting it into equations is slightly more complicated, especially because filament numbers are discrete integers. Such a concept can be described through the equations in (9.7).

The hypothesis behind this model is related to nutrient uptake. It is believed that nutrients travel from the distal towards the proximal extreme of the branchial arch through the blood vessel found along the base of the arch (Figure 9.2). This means that the more filaments exist between the two niches, the slower the stem cells residing in the proximal niche receive the needed amount of nutrients for division. The other niche, in theory, is found at the immediate entrance of the nutrients into the arch, thus in some sense, being in direct contact with the source. This would suggest that the proliferation of the proximal stem cells slows down with the number of filaments, while that of the distal stem cells should remain constant. For the moment, for generality, we keep the model with both velocities $q^{\pm}(t)$ depending on the total number of filaments.

Space-Dependent Elongation-Generation-Decay (SAEGD) Model

Assumptions 9.7.1

- ◇ *Filament $i \in \mathbb{Z}$ is generated at time $t_i \geq 0$ and grows according to the proliferation $\hat{p} = \dfrac{p}{1+\alpha u_i(t)}$ of the* filSCs.
- ◇ *The* brSCs *in the distal and proximal niches divide with rates $q^{\pm}(t)$, which decrease as more filaments are generated.*

$$
\begin{aligned}
&\frac{du_i(t)}{dt} = \frac{p}{1+\alpha u_i(t)}\left(f_i^+(t) + f_i^-(t)\right), \quad u_i(t_i) = 0 \\
&\text{with } f_i^+(t) = \begin{cases} 1, & q^+ t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \quad f_i^-(t) = \begin{cases} 1, & q^-(t)t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \\
&\qquad q^+(t) = \frac{r^+}{1+\beta^+\left(\lfloor q^+(t)t \rfloor - \lceil q^-(t)t \rceil\right)} \\
&\qquad q^-(t) = \frac{r^-}{1+\beta^-\left(\lfloor q^+(t)t \rfloor - \lceil q^-(t)t \rceil\right)} \\
&\text{and } t_i = \frac{i}{q^{\pm}(t_i)}, p, \alpha, \beta^{\pm}, r^+ \geq 0, r^- < 0
\end{aligned}
\tag{9.7}
$$

The equations for $q^{\pm}$ come from the following reasoning: as filament $i \geq 0$ on the distal side is generated when $q^+(t)t = i$ it means that at this particular moment there exist a total of i filaments in the distal half. Similarly, on the proximal side, filament $j < 0$ is generated when $q^-(t)t = j$ giving a total of j filaments in the proximal side at that time point. For times t between two filament generations, there exist $\lfloor q^+(t)t \rfloor$ filaments in the distal half, and $\lfloor |q^+(t)t| \rfloor = -\lceil q^-(t)t \rceil$ in the proximal half, resulting in a total of $\lfloor q^+(t)t \rfloor - \lceil q^-(t)t \rceil$ filaments in the entire arch.

A complication in dealing with this model stems from the implicit equations for the velocities $q^{\pm}(t)$ of the niches moving away from each-other. Not only are the equations implicit, but they contain floor and ceiling terms which together make it very difficult,

if not impossible, to analyse. Thus, the SAEGD model will only be approached through numerics and is discussed in Chapter 11.

Chapter 10
Analysis of the Models for Branchia Growth and Shape

This chapter is devoted to the analysis of the two general models (EGD and AEGD) presented in Chapter 9. The existence, uniqueness and non-negativity of the solutions is proven, the analytical solutions are determined and their regularity is discussed.

10.1 Analysis of the Elongation-Generation-Decay Model

Let $u_i : [0,\infty) \to \mathbb{R}$ be the length of filament with index $i \in \mathbb{Z}$ and parameters $\alpha, \beta^{\pm} \geq 0$, $p, r^+ > 0, r^- < 0$. The EGD model, as described in Chapter 9, is recapitulated bellow.

Elongation-Generation-Decay Model

$$\begin{aligned}
&\frac{du_i(t)}{dt} = pe^{-\alpha u_i(t)}\left(f_i^+(t) + f_i^-(t)\right), \quad u_i(t_i) = 0 \\
&\text{with } f_i^+(t) = \begin{cases} 1, & q^+(t)t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \quad f_i^-(t) = \begin{cases} 1, & q^-(t)t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \\
&\text{and } q^+(t) = \frac{r^+}{1+\beta^+ t}, \quad q^-(t) = \frac{r^-}{1+\beta^- t} \quad \text{with } t_i = \frac{i}{q^{\pm}(t_i)}
\end{aligned} \tag{9.5}$$

Proposition 10.1.1 (Existence, uniqueness and non-negativity) The system of equations described in (9.5) has a unique non-negative solution for each $i \in \mathbb{Z}$.

Proof. As previously mentioned, for each $i \in \mathbb{Z}$, the sum $f_i^+(t) + f_i^-(t) = 0$ for $t < t_i$ and $f_i^+(t) + f_i^-(t) = 1$ for $t \geq t_i$, where t_i is the time at which filament i is generated. Thus, our ordinary differential equation can be split into two:

$$\begin{cases} \dfrac{du_i(t)}{dt} = 0, & \text{for } t < t_i \\ \dfrac{du_i(t)}{dt} = pe^{-\alpha u_i(t)}, & \text{for } t \geq t_i \end{cases}$$

with initial conditions $u_i(0) = 0$, $u_i(t_i) = 0$, respectively, and parameters $p > 0, \alpha \geq 0$.

Let $g(t, u_i(t)) = g^*(u_i) = pe^{-\alpha u_i(t)}$, $g : [t_i, \infty) \times \mathbb{R} \to \mathbb{R}$, $g^* : \mathbb{R} \to \mathbb{R}$. It holds that $\dfrac{d}{dt} u_i|_{u_i=0} = p > 0$ for all $i \in \mathbb{Z}, t \geq t_i$. This implies non-negativity of solutions for non-negative initial values. In addition, g is continuous and bounded, as $|g(t, u_i(t))| \leq p$, for all $i \in \mathbb{Z}, t \geq t_i$. To prove uniform Lipschitz-continuity in u_i, we look for a uniform constant L such that

$$|g(t, u^1) - g(t, u^2)| = |g^*(u^1) - g^*(u^2)| = |pe^{-\alpha u^1} - pe^{-\alpha u^2}| \leq L|u^1 - u^2|.$$

By the Mean Value Theorem, there exists a constant $c \in [\min\{u^1, u^2\}, \max\{u^1, u^2\}] \geq 0$ such that $|g^*(u^1) - g^*(u^2)| \leq |(g^*)'(c)| \cdot |u^1 - u^2|$, i.e.

$$\left| pe^{-\alpha u^1} - pe^{-\alpha u^2} \right| \leq p \left| -\alpha e^{-\alpha c} \right| \cdot |u^1 - u^2| \leq p\alpha |e^{-\alpha c}| \cdot |u^1 - u^2| \leq p\alpha |u^1 - u^2|$$

since $|e^{-\alpha c}| \leq 1$ because $-\alpha c \leq 0$. Thus g is a globally Lipschitz-continuous function with respect to u_i, with Lipschitz constant $L = p\alpha$. By Picard-Lindelöf Theorem [47], this proves that a global integral solution of (9.4) *for $t \geq t_i$ exists and is unique. Finally, by joining the unique solution u_i for $t \geq t_i$ with $u_i = 0$ for $t < t_i$ we obtain a unique non-negative solution to* (9.4) *for all $t \geq 0$.* □

Proposition 10.1.2 (Analytical Solution) The unique analytical solution to the system of equations described in (9.4) reads as follows

$$u_i(t) = \begin{cases} 0, & \text{for } t < t_i \\ \dfrac{1}{\alpha} \ln\left(1 + \alpha p (t - t_i)\right), & \text{for } t \geq t_i \end{cases} \tag{10.1}$$

with $t_i = \frac{i}{q^+(t_i)} = \frac{i}{r^+ - i\beta^+}$ for $i \geq 0$ and $t_i = \frac{i}{q^-(t_i)} = \frac{i}{r^- - i\beta^-}$ for $i < 0$.

Proof. In order to find the general form of the solution, it is sufficient to look for it in the interval $t \in [t_i, \infty)$ for any particular $i \in \mathbb{Z}$. Thus, for simplicity, we can rename $u(t) = u_i(t)$ with t_0 corresponding to the time t_i at which the filament should be generated.

We thus wish to find the solution of the ordinary differential equation

$$\frac{du(t)}{dt} = pe^{-\alpha u(t)}, \quad u(t_0) = 0$$

By using the method of Separation of Variables, this is equivalent to

$$\int_0^u e^{\alpha \bar{u}} d\bar{u} = p \int_{t_0}^t d\bar{t} \quad \Leftrightarrow \quad \left[\frac{1}{\alpha} e^{\alpha \bar{u}}\right]_0^u = [p\bar{t}]_{t_0}^t \quad \Leftrightarrow \quad e^{\alpha u} = 1 + \alpha p(t - t_0)$$

Coming back to the original notations, the analytical solution for $t \geq t_i$ reads

$$u_i(t) = \frac{1}{\alpha} \ln\left(1 + \alpha p(t - t_i)\right) = \frac{1}{\alpha} \ln\left[1 + \alpha p\left(t - \frac{i}{r^\pm - i\beta^\pm}\right)\right]$$

and $u_i(t) = 0$ for $t < t_i$. □

Remark 10.1.1 (Regularity) *Note that each function $u_i : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous for every $i \in \mathbb{Z}$, but only u_0 is differentiable on $\mathbb{R}_+$ with respect to time. For u_i with $i \neq 0$, the slope in t_i on the left side $u_i'(t_i^-) = 0$ while on the right side $u_i'(t_i^+) = p$, since*

$$u_i'(t) = \frac{p}{1 + \alpha p(t - t_i)}, \quad \text{for } t \geq t_i$$
$$u_i'(t) = 0, \quad \text{for } t < t_i$$

with $t_0 = 0$ and $t_i > 0$ for all $i \neq 0$. Thus u_i is not differentiable in t_i for $i \neq 0$. However, each $u_i \in \mathscr{C}^\infty((t_i, \infty))$, so the solutions are smooth functions on the corresponding restricted domains on which they are positive ($t > t_i \geq 0$ for each $i \in \mathbb{Z}$).

Remark 10.1.2 (Linear-Growth and Elongation-Decay Models) *All the above results also hold for both the LG and ED models, as they are simplified versions of the EGD model, with $\alpha = \beta^\pm = 0$ (LG) and $\alpha > 0, \beta^\pm = 0$ (ED).*

10.2 Analysis of the Alternative Elongation-Generation-Decay Model

Let $u_i : [0,\infty) \to \mathbb{R}$ be the length of filament with index $i \in \mathbb{Z}$ and parameters $\alpha, \beta^{\pm} \geq 0$, $p, r^{+} > 0$, $r^{-} < 0$. The AEGD model, as described in Chapter 9, is recapitulated bellow.

Alternative Elongation-Generation-Decay Model

$$\begin{aligned}
&\frac{du_i(t)}{dt} = \frac{p}{1+\alpha u_i(t)}\left(f_i^{+}(t) + f_i^{-}(t)\right), \quad u_i(t_i) = 0 \\
&\text{with } f_i^{+}(t) = \begin{cases} 1, & q^{+}(t)t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \quad f_i^{-}(t) = \begin{cases} 1, & q^{-}(t)t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \\
&\text{and } q^{+}(t) = \frac{r^{+}}{1+\beta^{+}t}, \quad q^{-}(t) = \frac{r^{-}}{1+\beta^{-}t} \quad \text{with } t_i = \frac{i}{q^{\pm}(t_i)}
\end{aligned} \tag{9.6}$$

Proposition 10.2.1 (Existence, uniqueness and non-negativity) The system of equations described in (9.6) has a unique non-negative solution for each $i \in \mathbb{Z}$.

Proof. As previously mentioned, for each $i \in \mathbb{Z}$, the sum $f_i^{+}(t) + f_i^{-}(t) = 0$ for $t < t_i$ and $f_i^{+}(t) + f_i^{-}(t) = 1$ for $t \geq t_i$, where t_i is the time at which filament i is generated. Thus, our ordinary differential equation can be split into two:

$$\begin{cases} \dfrac{du_i(t)}{dt} = 0, & \text{for } t < t_i \\ \dfrac{du_i(t)}{dt} = \dfrac{p}{1+\alpha u_i(t)}, & \text{for } t \geq t_i \end{cases}$$

with initial conditions $u_i(0) = 0$, $u_i(t_i) = 0$, respectively, and parameters $p > 0, \alpha \geq 0$.

Let $g(t, u_i(t)) = g^*(u_i) = \frac{p}{1+\alpha u_i(t)}$, $g : [t_i, \infty) \times \mathbb{R} \to \mathbb{R}$, $g^* : \mathbb{R} \to \mathbb{R}$. It holds that $\dfrac{d}{dt}u_i|_{u_i=0} = p > 0$ for all $i \in \mathbb{Z}, t \geq t_i$. This implies non-negativity of solutions for non-negative initial values. It also confirms that no singularity appears in the system, as $u_i \neq -\frac{1}{\alpha}$ for all $i \in \mathbb{Z}$, $t \geq t_i$. In addition, g is continuous and bounded, as $|g(t, u_i(t))| \leq p$, for all $i \in \mathbb{Z}, t \geq t_i$. To prove uniform Lipschitz-continuity in u_i, we look for a uniform constant L such that

$$|g(t,u^1)-g(t,u^2)| = |g^*(u^1)-g^*(u^2)| = \left|\frac{p}{1+\alpha u^1}-\frac{p}{1+\alpha u^2}\right| \leq L|u^1-u^2|$$

By the Mean Value Theorem, there exists a constant $c \in [\min\{u^1,u^2\},\max\{u^1,u^2\}] \geq 0$ *such that* $|g^*(u^1)-g^*(u^2)| \leq |(g^*)'(c)| \cdot |u^1-u^2|$*, i.e.*

$$\left|\frac{p}{1+\alpha u^1(t)}-\frac{p}{1+\alpha u^2(t)}\right| \leq p\alpha \left|\frac{1}{(1+\alpha c)^2}\right| \cdot |u^1-u^2| \leq p\alpha |u^1-u^2|$$

since $(1+\alpha c)^2 \geq 1$ *because* $\alpha c \geq 0$*. Thus g is a globally Lipschitz-continuous function with respect to* u_i *with uniform Lipschitz constant* $L = p\alpha$*. By Picard-Lindelöf Theorem [47], this proves that a global integral solution of* (9.6) *for* $t \geq t_i$ *exists and is unique. Finally, by joining the unique solution* u_i *for* $t \geq t_i$ *with* $u_i = 0$ *for* $t < t_i$ *we obtain a unique non-negative solution to* (9.6) *for all* $t \geq 0$*.* □

Proposition 10.2.2 (Analytical Solution) The unique analytical solution to the system of equations described in (9.6) reads as follows

$$u_i(t) = \begin{cases} 0, & \text{for } t < t_i \\ \dfrac{\sqrt{1+2\alpha p(t-t_i)}-1}{\alpha}, & \text{for } t \geq t_i \end{cases} \tag{10.2}$$

with $t_i = \frac{i}{q^+(t_i)} = \frac{i}{r^+ - i\beta^+}$ for $i \geq 0$ and $t_i = \frac{i}{q^-(t_i)} = \frac{i}{r^- - i\beta^-}$ for $i < 0$.

Proof. In order to find the general form of the solution, it is sufficient to look for it in the interval $t \in [t_i,\infty)$ *for any particular* $i \in \mathbb{Z}$*. Thus, for simplicity, we again rename* $u(t) = u_i(t)$ *with* t_0 *corresponding to the time* t_i *at which the filament should be generated. We thus wish to find the solution of the ordinary differential equation*

$$\frac{du(t)}{dt} = \frac{p}{1+\alpha u(t)}, \quad u(t_0) = 0$$

By using the method of Separation of Variables, this is equivalent to

$$\int_0^u (1+\alpha\bar{u})d\bar{u} = p\int_{t_0}^t d\bar{t} \quad \Leftrightarrow \quad \left[\bar{u}+\frac{\alpha\bar{u}^2}{2}\right]_0^u = [p\bar{t}]_{t_0}^t \quad \Leftrightarrow \quad u+\frac{\alpha u^2}{2} = p(t-t_0)$$

This gives us a quadratic equation to solve. Thus by only keeping the positive root and coming back to the original notations, the analytical solution for $t \geq t_i$ reads

$$u_i(t) = \frac{\sqrt{1+2\alpha p(t-t_i)}-1}{\alpha}$$

and $u_i(t) = 0$ for $t < t_i$, with $t_i = \frac{i}{r^{\pm} - i\beta^{\pm}}$. This is real and non-negative since $p > 0$, $\alpha \geq 0$ and $t \geq t_i$. □

Remark 10.2.1 (Regularity) *Similarly to the EGD model, each function $u_i : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous for every $i \in \mathbb{Z}$, but only u_0 is differentiable on $\mathbb{R}_+$ with respect to time. For u_i with $i \neq 0$, the slope in t_i on the left side $u_i'(t_i^-) = 0$ while on the right side $u_i'(t_i^+) = p$, since*

$$u_i'(t) = \frac{p}{\sqrt{1+2\alpha p(t-t_i)}}, \quad \text{for } t \geq t_i$$
$$u_i'(t) = 0, \quad \text{for } t < t_i$$

with $t_0 = 0$ and $t_i > 0$ for all $i \neq 0$. Thus u_i is not differentiable in t_i for $i \neq 0$. However, each $u_i \in \mathscr{C}^{\infty}((t_i, \infty))$, so the solutions are smooth functions on the corresponding restricted domains on which they are positive ($t > t_i \geq 0$ for each $i \in \mathbb{Z}$).

Remark 10.2.2 (Space-Dependent Elongation-Generation-Decay Model) *The global existence, uniqueness and non-negativity of solutions hold for the SAEGD model, as well, as it is described by the same system of ordinary differential equations. Further, the analytical solution of the SAEGD model is as in* (10.2)*, the difference appearing in the times t_i at which the filaments are added to the arch. As opposed to previous models, the times t_i of the SAEGD model* (9.7) *cannot be computed analytically since they depend on implicit equations involving terms containing floor and ceiling, and these times will be found numerically as explained in the next Chapter, Section 11.3.5.*

Chapter 11
Numerical Considerations and Application of the Models for Branchia Growth and Shape

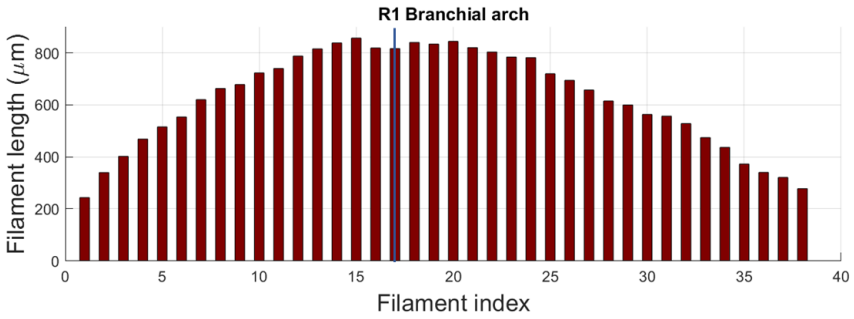
This chapter starts with an overview of the numerical procedures used for data transformation and the parameter estimation methods employed. Subsequently, each of the models presented and analysed in Chapters 9-10 is simulated and compared to data from measurements of one R1 branchial arch. Model selection is subsequently performed using the Akaike Information Criterium.

11.1 Numerical Aspects and Data Transformation

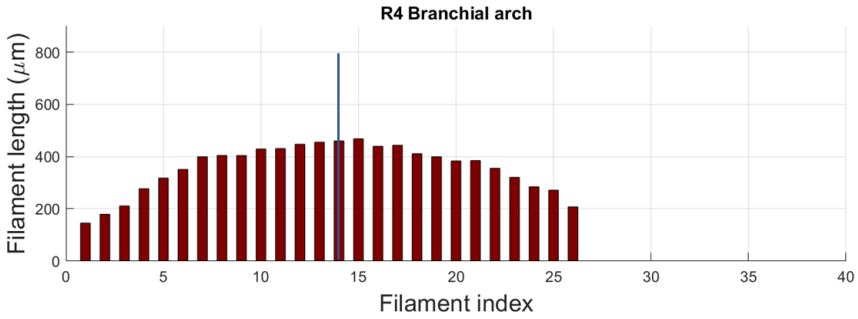
In order for the models to be properly applied to the available measurements, one needs to first "center" these data such that for each branchial arch, the oldest filament lays at position 0 on the x-axis with the rest of the filaments to the left (proximal half) and right (distal half) being labeled by negative and positive integers, respectively.

However, a difficulty arises stemming from the variability in the lengths and numbers of filaments per branchial arch. The smallest arches have up to 25 filaments of lengths up to 500 μm, while in the case of bigger arches we have up to 45 filaments of up to 900 μm in length. Considering that branchial arches are known to have a curved concave shape, one would expect that in the data of each arch the oldest filament is the longest one which is supposed to split the two "half-arches" generated by the two different peripheral niches (proximal and distal). Nevertheless, due to measurement errors this is not always the case, so instead of just looking for the filament of maximum length out of a branchial

arch, we need a different method of determining the oldest filament which we need to set at position 0 so that the data are properly "centered" (see Figure 11.1).



(a) R1 Branchial arch with 40 filaments

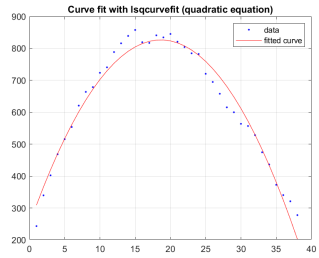


(b) R4 Branchial arch with 26 filaments

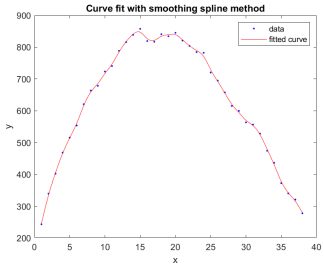
Fig. 11.1: (a) Measurements of lengths of filaments from a R1 branchial arch. This arch has 40 filaments with lengths of up to approximately 800 μm. We notice that the filaments which we would expect to be the oldest (blue vertical line) are shorter than the ones around. (b) Data from an R4 branchial arch with 26 filaments of lengths of up to approximately 500 μm. As in (a), the filaments expected to be the oldest seem to be shorter than the surrounding ones, which is in contradiction with biological information.

We need to first fit a concave asymmetric curve to the measurements for each branchial arch and then find the maximum of that curve, which would then be set at position 0. It turns out that, although a seemingly basic process, fitting a concave asymmetric curve to the data is not as straightforward using existing computational routines. Various issues appear when trying to use implemented routines in *MATLAB*®. In the case of the `lsqcurvefit` function, the issue comes from the asymmetry of the arch, for example.

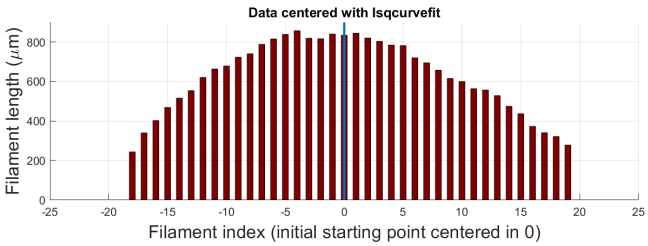
The `lsqcurvefit` offers the possibility to input a specific formula for the function one wants to fit to the data and then it estimates the best parameters for that particular function. For example, if one tries to fit a parabola given by a quadratic equation, then it will be symmetric, which is not the case in our arches (see Fig. 11.2(a,c)). For other methods, such as the '`spline`' option in the `fit` function of *MATLAB*®, other issues appear: either the curve reproduces all the inflexions in the data (thus not being concave), or if one decides to skip certain data points one might end up skipping the actual inflexion point at which the maximum should be (see Fig. 11.2(b,d)). What one needs instead is a spline curve-fitting method which has the possibility of specifying options such as the convexity of the curve.



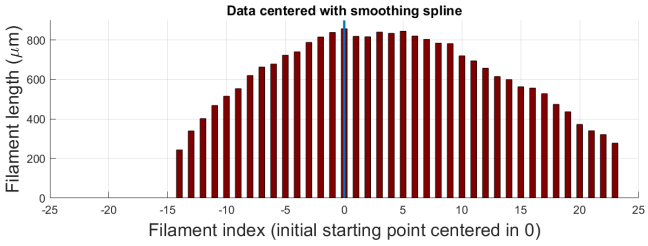
(a) R1 Branchial arch - curve fit with parabola



(b) R1 Branchial arch - curve fit with smoothing spline method



(c) R1 Branchial arch centered according to (a)

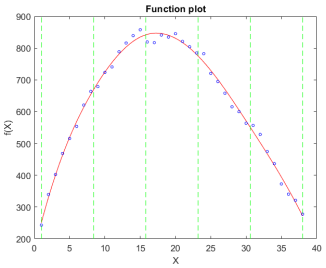


(d) R1 Branchial arch centered according to (b)

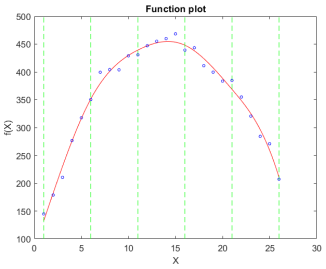
Fig. 11.2: (a) Fitted curve with a quadratic equation method by using the `lsqcurvefit` function in *MATLAB*®. Data fitted are the same as in 11.1(a). (b) Fitted curve with a smoothing spline method by using the `fit` routine in *MATLAB*®. Data fitted are the same as in 11.1(a). (c) The R1 branchial arch data centered so that the maximum of the curve from (a) is set to position 0. (d) The R1 branchial arch data centered so that the maximum of the curve from (b) is set to position 0.

The open-source `SLM curve-fitting tool` [85] for *MATLAB*® gives the possibility of passing prior knowledge about the expected shape as an option. This tool uses least squares splines for the curve fit, which gives the method enough flexibility to fit almost any kind of data, but in addition it allows the possibility of inputting prior knowledge about the model into the fitting process.

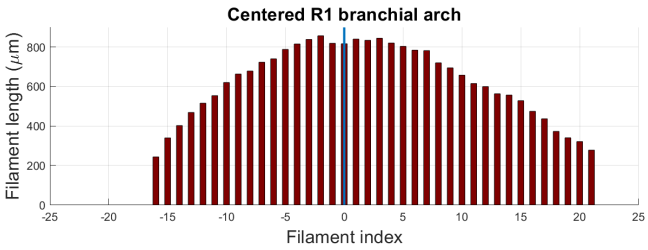
By using the `SLM curve-fitting tool`, a concave curve was fitted to each branchial arch and the maximum of the curve was determined, subsequently aligning the data accordingly.



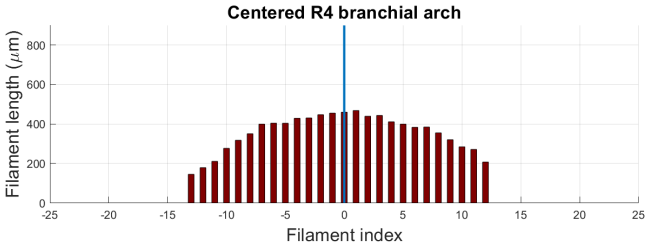
(a) R1 Branchial arch - SLM curve fit



(b) R4 Branchial arch - SLM curve fit



(c) R1 Branchial arch centered



(d) R4 Branchial arch centered

Fig. 11.3: (a) Concave curve fit with the `SLM tool` to the R1 branchial arch data from Fig. 11.1(a). (b) Concave curve fit with the `SLM tool` to the R4 branchial arch data from Fig. 11.1(b). (c) The R1 branchial arch data centered so that the maximum of the curve from (a) is set to position 0. (d) The R4 branchial arch data centered so that the maximum of the curve from (b) is set to position 0.

Once the data have been "centered", the models can be applied to individual branchial arches. Another aspect to think about in order to run the models is the end time point. When should the simulations stop?

Remark 11.1.1 *Since the models do not attain a stable state, in agreement with the fish permanent growth throughout their life, it means that the time point at which to stop the simulation is also a parameter. However, we don't have any information on the age at which fish were analysed since fish were selected based on their length and not on their age. Thus a way of approximately determining the age of fish based on their length is needed.*

One can in fact model the length-at-age relation between the size of the fish and their respective age, according to the von Bertalanffy growth equation (11.1) [86].

$$L(t) = L_\infty \left(1 - e^{-K(t-t_0)}\right) \tag{11.1}$$

where L is the length of the organism, K is the growth rate and L_∞ the asymptotic length at which the growth is zero. The "initial" time point t_0 is defined as the time at which the fish would have had zero length.

As studied in [87], the parameters for medaka model organism are as follows: $L_\infty = 49.5\ TL$ (total length), $K = 0.01\ dph^{-1}$ (days post hatch), $t_0 = 3.3\ dph$. Accordingly, a fish of approximately $30mm$ in length would have an age of circa $100dph$ (Fig. 11.4).

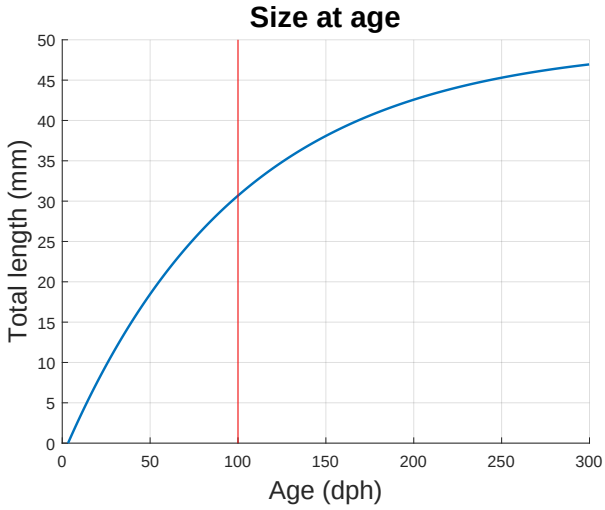


Fig. 11.4: Size-at-age relation of medaka fish based on the von Bertalanffy growth equation (11.1), with parameters taken from [86]. An $100dph$ fish would have a size of approximately $30mm$.

Thus, in the simulations hereon, the end time point has been set to 100 (as our fish had a length of approximately 30*mm*), but parameters for fish of other size and age can be easily estimated following the same procedure described in the next section.

11.2 Parameter Estimation

In order to fit the models to the data one compares the simulated lengths of filaments at the end time point to the experimental measurements of filament lengths. Note that this should be done independently for each branchial arch since they differ in number and lengths of filaments and hence in the proliferation rates of stem cells. In the following sections (11.3.1-11.3.5), data from the same branchial arch is used to fit the various models. This branchial arch consist of 40 filaments, with 17 found in the proximal half (negative axis), and 22 in the distal half (positive axis), plus the oldest one centred at 0 - Fig. 11.5. It has been measured $n = 7$ times, from which the mean and standard deviation for each filament have been computed. Further by investigating the measurement errors for each filament *via* Quantile-Quantile Plots, it was shown that these are normally distributed $\sim \mathcal{N}(0,1)$.

We thus need to find the best parameters such that the "distance" between data and simulations is minimal. For this, an objective function to be minimised is developed, whihc depends on the parameters to be estimated. This sum-of-squares-type function is not difficult to construct since we have analytical formulae for each solution $u_i(t)$ and

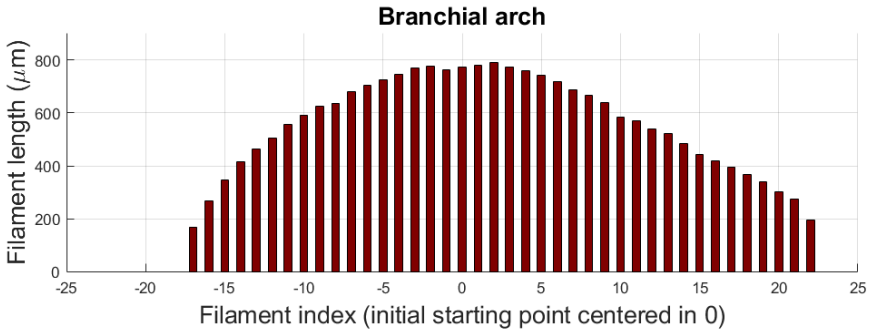
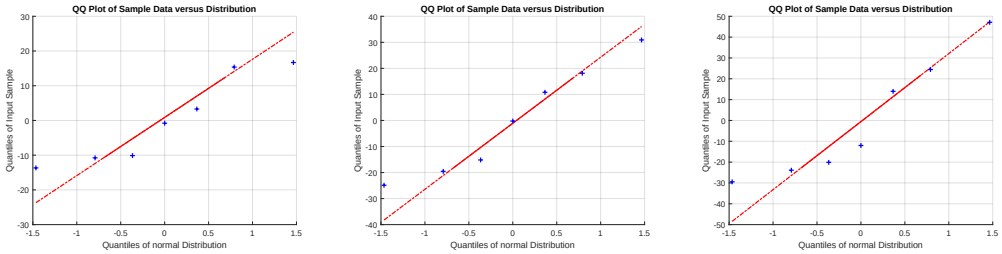


Fig. 11.5: R1 branchial arch with 40 filaments (17 proximal, 22 distal) used for parameter fitting and model comparison.



(a) QQ Plot for filament -3 (b) QQ Plot for filament 0 (c) QQ Plot for filament 3

Fig. 11.6: Quantile-Quantile Plots for filaments with indices -3 (a), 0 (b) and 3 (c). If the points can be fitted by a straight line it means that the measurement errors follow a Normal Distribution.

times t_i. In addition, the algorithm defining the objective function must internally first compute the times t_i at which filaments should be inserted in the branchial arch. So not only do the solutions u_i depend on the parameters and the times t_i, but the times t_i depend on the parameters as well.

The parameter estimation itself is implemented in the *MATLAB*® programming language and is performed by using the `lsqnonlin` routine which takes as arguments the objective function, the parameters to be estimated, and bounds of intervals in which to search for the best parameters. The `lsqnonlin` routine solves non-linear least-squares curve fitting problems of the form

$$\min_x \|h(x)\|_2^2 = \min_x \left(h_1(x)^2 + h_2(x)^2 + \cdots + h_n(x)^2\right)$$

where in our case, $h_i(x)$ is the difference between the solution of the model at the final time point (for filament i) and the length of the filament i from the centred experimental data, with x being the vector of parameters to be optimized. We chose the Trust-Region-Reflective Least Squares Algorithm for `lsqnonlin`, which has been described in Chapter 2. Having shown that the measurement errors are normally distributed, it is allowed to fit the parameters by using Least-Squares algorithms.

11.3 Model Comparison

In the following sections 11.3.1-11.3.5, data from the same branchial arch (shown in Fig. 11.5) is used to fit the various models. For model comparison, the Akaike Information Criterion (*AIC*), also described in Chapter 2, was computed. This takes into account the fit between the data and the model, but penalizes the model for the number of parameters (11.2). In other words, it tries to avoid over-fitting by penalizing models with too many parameters.

$$AIC = n\log(\hat{\sigma}^2) + 2K, \quad \text{with} \quad \hat{\sigma}^2 = \frac{\|f(x)\|_2^2}{n} \tag{11.2}$$

Here n represents the sample size, i.e. the number of repeated measurements for the same branchial arch. The K corresponds to the number of parameters used in the model. A further development of the *AIC* number, namely the corrected Akaike Information Criterion *AICc* (11.3) introduces a correction term for small sample sizes.

$$AICc = AIC + \frac{2K^2 + 2K}{n - K - 1} \tag{11.3}$$

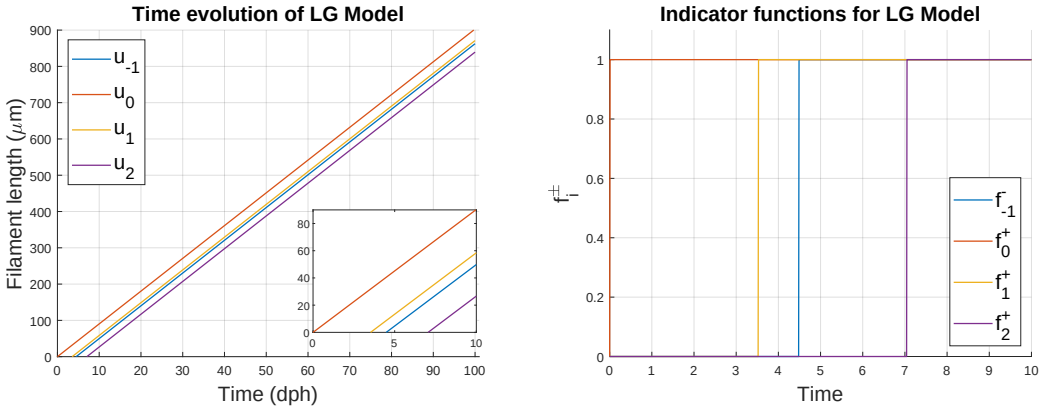
In the following sections, for comparing the various models developed, the *AICc* was used since our sample size $n = 7$ is very small. The smaller the *AICc* is, the better the model. The absolute value of the *AICc* is not important, but only the differences Δ_i between the $AICc_i$ of each model and the minimum $AICc_{min}$ resulting from one of the models is what matters.

11.3.1 Results of the Linear-Growth Model

This scenario presents the case of filaments growing at constant rates so the solutions are linear in time 11.7(a). Further, as the stem cells in the peripheral niches proliferate at constant rates, the times t_i of new filament generation are equally distributed in time.

$$u_i(t) = p(t - t_i), \ \text{for} \ t \geq t_i = \frac{i}{q^{\pm}} \quad \text{and} \quad u_i(t) = 0, \ \text{for} \ t < t_i$$

Thus the indicator functions $f_i^{\pm}$ are either equal to zero, or has the form of a step function with discontinuity at t_i otherwise, as can be seen in Fig. 11.7(b). Further note that if $q_+ = q_-$, the times $t_i = t_{-i}$.



(a) Time evolution of solutions $u_i(t)$

(b) Indicator functions $f_i^{\pm}(t)$

Fig. 11.7: (a) Plot of solutions $u_{-1}(t)$ (blue), $u_0(t)$ (red), $u_1(t)$ (orange), $u_2(t)$ (purple) with respect to time, for the Linear Growth Model, with parameters estimated to best fit the data from a branchial arch with 40 filaments (17 proximal and 22 distal) - as seen in Fig. 11.8. (b) Indicator functions $f_{-1}^{-}(t)$ (blue), $f_0^{+}(t)$ (red), $f_1^{+}(t)$ (orange), $f_2^{+}(t)$ (purple) with respect to time.

As expected, the fit of the LG model to the data from the branchial arch is very poor, as the model produces an *in silico* arch with a triangular shape which does not match the curvature of the experimental data (Fig. 11.8).

11.3.2 Results of the Elongation-Decay Model

In contrast to the LG model, in the Elongation-Decay Model the solutions follow logarithmic curves as is deduced from (10.1), and observed in Fig. 11.9(a). Also for this model, the times t_i are situated at regular intervals on each of the half-axes (positive/negative), but as the parameters $q_+ \neq q_-$, overlaps of the type $t_i = t_{-i}$ are not observed for the first few proximal/distal filaments (Fig. 11.9(b)).

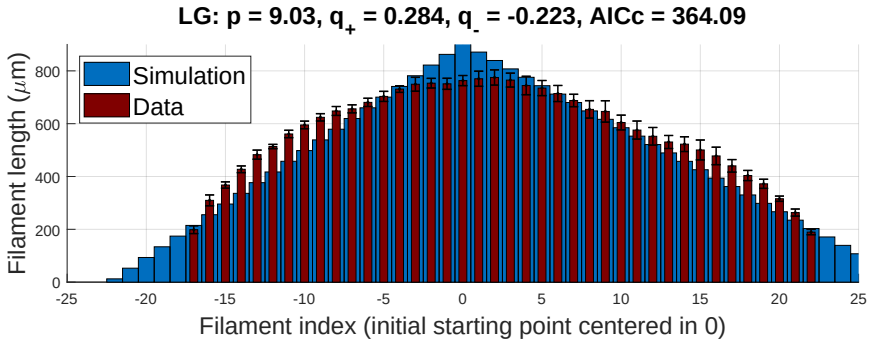
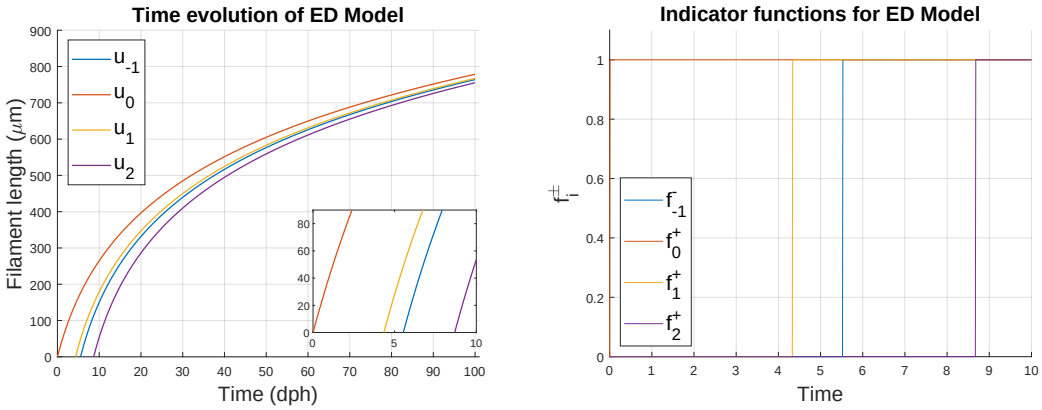


Fig. 11.8: Fit of Linear-Growth Model with optimized parameters.

$$u_i(t) = \frac{1}{\alpha}\ln\left(1+\alpha p(t-t_i)\right), \text{ for } t \geq t_i = \frac{i}{q^{\pm}} \quad \text{and} \quad u_i(t) = 0 \text{ for } t < t_i$$



(a) Time evolution of solutions $u_i(t)$

(b) Indicator functions $f_i^{\pm}(t)$

Fig. 11.9: (a) Plot of solutions $u_{-1}(t)$ (blue), $u_0(t)$ (red), $u_1(t)$ (orange), $u_2(t)$ (purple) with respect to time, for the Elongation-Decay Model, with parameters estimated to best fit the data from a branchial arch with 40 filaments (17 proximal and 22 distal) - as seen in Fig. 11.8. (b) Indicator functions $f_{-1}^{-}(t)$ (blue), $f_0^{+}(t)$ (red), $f_1^{+}(t)$ (orange), $f_2^{+}(t)$ (purple) with respect to time.

Due to its logarithmic filament growth curves, this model allows the final combination of filaments in a single arch to produce a curved shape similar to what is observed in experimental data (11.10).

In Figure 11.11, it can be seen that even if the fit is not too bad, certain issues appear. First of all, on both sides more filaments than in the data are produced. One could of course argue that, considering their length is so small, they correspond to filaments that have just been generated and could not be observed in the experimental data, or that were even cut when the branchial arches were separated. Secondly, the model produces approximately the same shape for both the proximal and distal sides which is not in agreement with

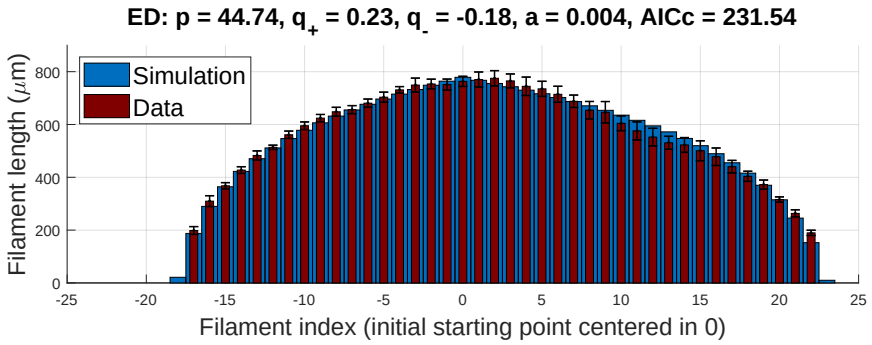
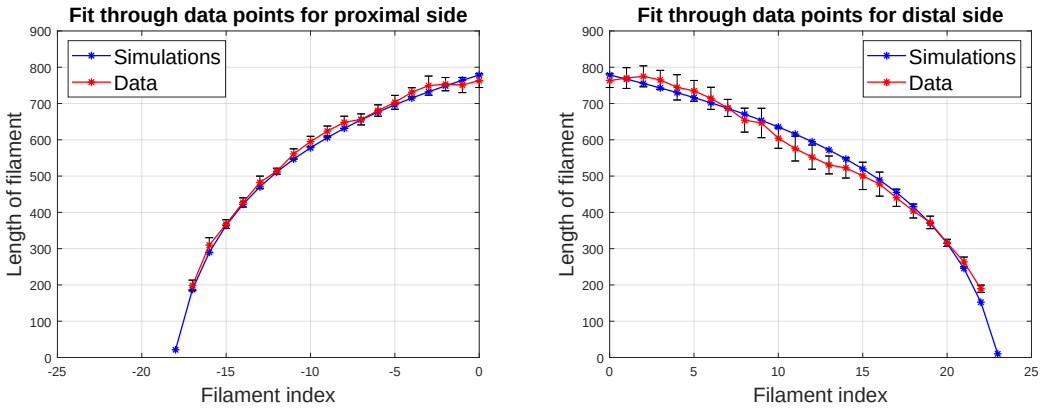


Fig. 11.10: Fit of Elongation-Decay Model with optimized parameters.



(a) Fit of simulation results to data, for the proximal half

(b) Fit of simulation results to data, for the distal half

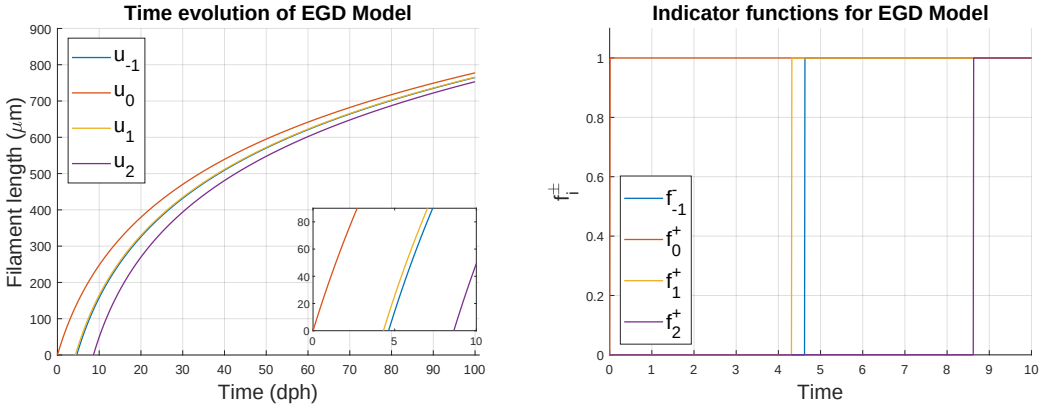
Fig. 11.11: Fit of the Elongation-Decay Model (blue) to experimental data (red) for proximal (a) and distal (b) halves.

the data. The model can reproduce the proximal side relatively well, with many model points situated within the error bars, however a higher curvature should be aimed for, i.e. a more shallow descent among the older filaments followed by a much steeper one for the younger filaments.

In conclusion, a simple logarithmic trend with a linearly decreasing argument, corresponding to decay only in the growth of filaments is not sufficiently good to reproduce the biological mechanism behind the experimental data collected.

11.3.3 Results of the Elongation-Generation-Decay Model

As in the ED model, the filament elongation curves in time follow logarithmic trends (10.1) in the case of the Elongation-Generation-Decay (EDG) Model, as well (Fig. 11.12(a)). The difference between the two models comes from the times t_i at which new filaments are added to the arch. In the current model these moments are no longer distributed at equal intervals on the time axis, since the proliferation $q_{\pm}(t)$ of stem cells in the peripheral niches depends on time.



(a) Time evolution of solutions $u_i(t)$

(b) Indicator functions $f_i^{\pm}(t)$

Fig. 11.12: (a) Plot of solutions $u_{-1}(t)$ (blue), $u_0(t)$ (red), $u_1(t)$ (orange), $u_2(t)$ (purple) with respect to time, for the Elongation-Generation-Decay Model, with parameters estimated to best fit the data from a branchial arch with 40 filaments (17 proximal and 22 distal) - as seen in Fig. 11.8. (b) Indicator functions $f_{-1}^{-}(t)$ (blue), $f_0^{+}(t)$ (red), $f_1^{+}(t)$ (orange), $f_2^{+}(t)$ (purple) with respect to time.

$$u_i(t) = \frac{1}{\alpha}\ln\left(1+\alpha p(t-t_i)\right), \text{ for } t \geq t_i = \frac{i}{r^{\pm} - i\beta^{\pm}} \quad \text{and} \quad u_i(t) = 0 \text{ for } t < t_i$$

The further away the niche (at filament i) is from the middle (zero) point, the slower its stem cells proliferate thus the slower it advances to the next filament $(i+1)$ and the longer the following time interval $[t_i, t_{i+1}]$ will be.

Once again, the combination of filaments in a single arch produces a curved shape, and the fit for our test branchial arch can be seen in Fig. 11.13. It would seem that the fit is better than for the previous model, fact also proven by the AICc value which is smaller than before. The AICc values can be seen in the title of the fit figures, for each model. Nevertheless, the difference is too small – with 5.38% smaller, not enough to state that this is a much better model.

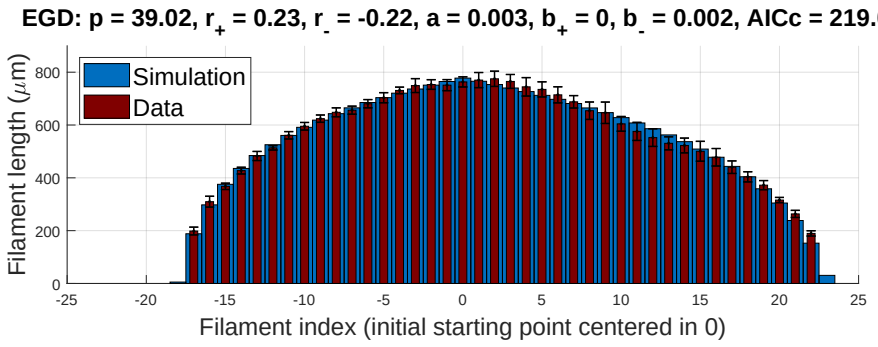
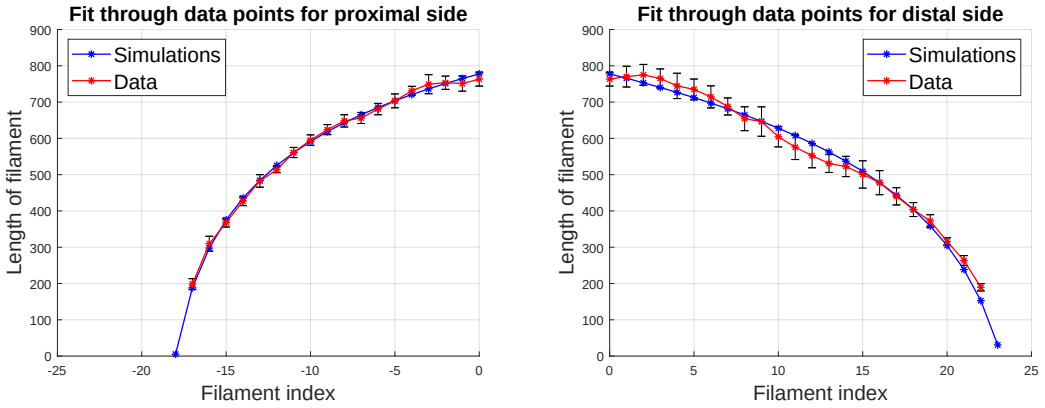


Fig. 11.13: Fit of Elongation-Generation-Decay Model with optimized parameters.

The EGD model provides a much better overlap with data on the proximal side, but still does not reproduce the distal side sufficiently well. As can be noted from the parameter values (Fig. 11.13), the distal side actually has the same behaviour as in the previous model, due to its decay parameter in the proliferation of arch stem cells being null: $\beta^+ = 0$. In fact even the initial proliferation rate of the distal niche is the same as in the previous model: $r^+ = 0.23$. Thus, one could argue that since the distal side could not be reproduced well enough by a logarithmic curve, the fitting process is governed by the proximal side which chooses parameters for its benefit and then passes the common ones (p and α) to



(a) Fit of simulation results to data, for the proximal half

(b) Fit of simulation results to data, for the distal half

Fig. 11.14: Fit of the Elongation-Generation-Decay Model (blue) to experimental data (red) for proximal (a) and distal (b) halves.

the distal side as well. Even though these have in fact been changed from their previous values in the Elongation-Decay Model, the shape of the distal side was not profoundly affected. Most of the model points on the proximal side indeed fall within the error bars of the data points (apart from those for the youngest filaments), but a majority of those are exactly at the borders of those error intervals. In addition, the issue of having one extra filament at each side remains.

11.3.4 Results of the Alternative Elongation-Generation-Decay Model

The Alternative Elongation-Generation-Decay (AEGD) Model, as the name suggests, is a variant of the EGD model, in which the function governing the evolution of solution is changed from a logarithmic into an inverse-quadratic one as in (10.2) and recapitulated below.

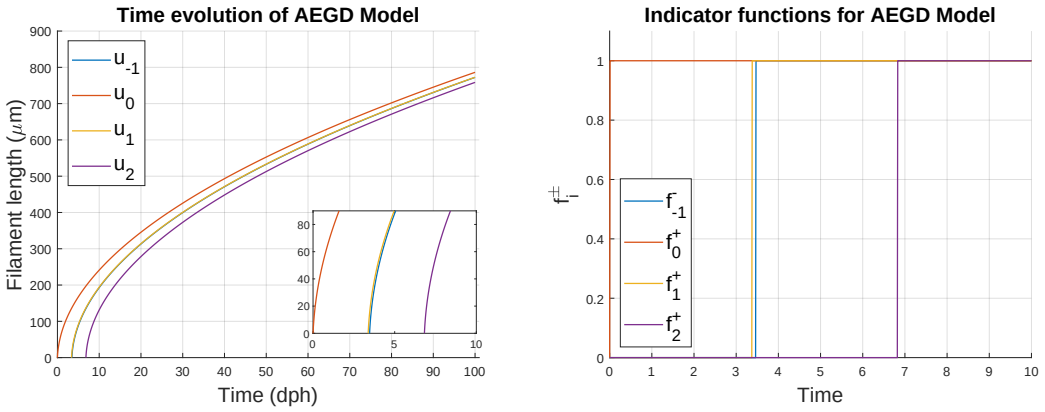
$$u_i(t) = \frac{\sqrt{1+2\alpha p(t-t_i)}-1}{\alpha}, \text{ for } t \geq t_i = \frac{i}{r^{\pm} - i\beta^{\pm}} \quad \text{and} \quad u_i(t) = 0 \text{ for } t < t_i$$

Its time evolution and the corresponding indicator functions for the fitted parameters of the AEGD model can be seen in Fig. 11.15. One notices the important difference between the solution of the AEGD model in comparison to the EGD model is the steepness of the

curve at the initial time points. The curve first climbs relatively abruptly and then slows down starting to resemble the shape of the entire EGD model curve.

Furthermore, Fig. 11.16 shows the fit to the data from the same branchial arch used throughout this chapter, and displays the corresponding estimated values for the parameters. As before, the two half-arches are linked through the proliferation p and decay α in the filaments. To be noted is that both $\beta^{\pm}$ decay parameters are non-zero suggesting that there is decay in the proliferation of both niches. Further, the $r^{\pm}$ are equal (in absolute value) so the initial proliferation in both niches is the same, which is more biologically plausible since both niches start in the same place at time $t_0 = 0$. Striking is the value of the proliferation p, which is one order of magnitude larger than in the previous models, but this is a consequence of the correlation between parameters p and α, which will be discussed in Section 11.3.6. Here the resulting fit looks extremely good, the AICc value reinforcing this assertion by having a remarkably smaller value than in the EGD model, with 74.87% smaller than before.

Probably the most striking outcome of this model is its capability to fit the data on the distal side much better than all the previous models while at the same time recapitulating



(a) Time evolution of solutions $u_i^{\pm}(t)$

(b) Indicator functions $f_i^{\pm}(t)$

Fig. 11.15: (a) Plot of solutions $u_{-1}(t)$ (blue), $u_0(t)$ (red), $u_1(t)$ (orange), $u_2(t)$ (purple) with respect to time, for the Alternative Elongation-Generation-Decay Model, with parameters estimated to best fit the data from a branchial arch with 40 filaments (17 proximal and 22 distal) - as seen in Fig. 11.8. (b) Indicator functions $f_{-1}^{-}(t)$ (blue), $f_0^{+}(t)$ (red), $f_1^{+}(t)$ (orange), $f_2^{+}(t)$ (purple) with respect to time.

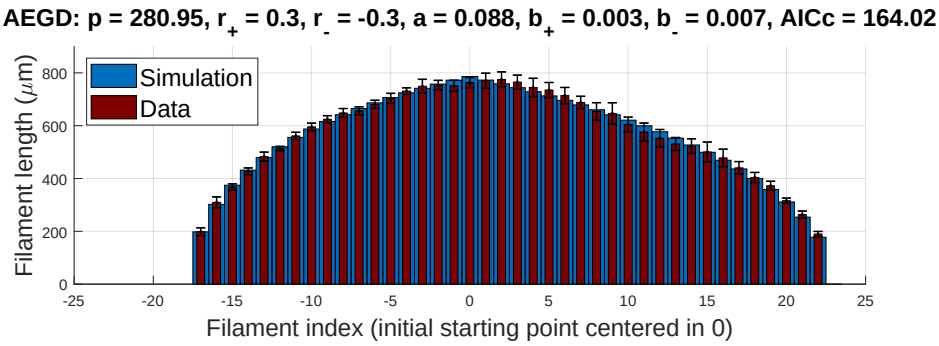
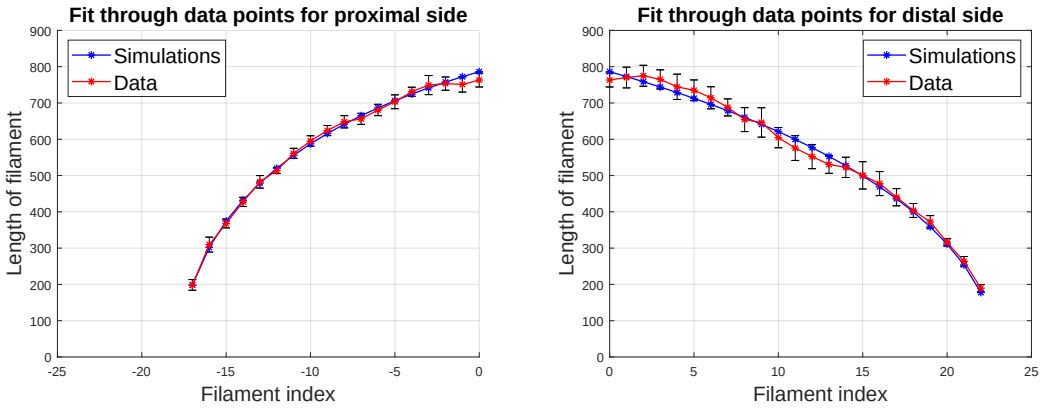


Fig. 11.16: Fit of Alternative Elongation-Generation-Decay Model with optimized parameters.

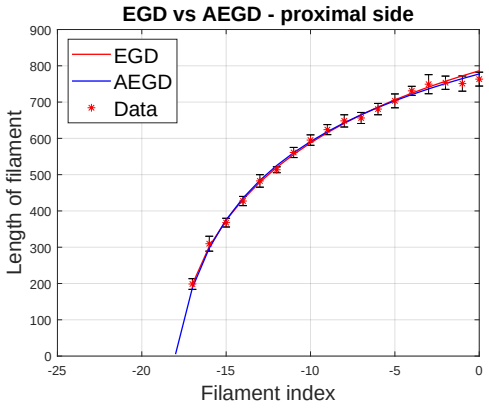


(a) Fit of simulation results to data, for the proximal half

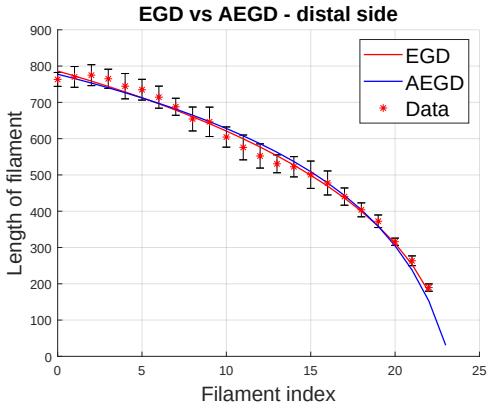
(b) Fit of simulation results to data, for the distal half

Fig. 11.17: Fit of the Elongation-Decay Model (blue) to experimental data (red) for proximal (a) and distal (b) halves.

the data on the proximal side very well. First of all, the abrupt ascent of the inverse-quadratic function for young filaments ensures that no extra filaments are added at any of two the extremes. Second, all model points fall within the error bars, with only a few lying on the boundary (see Fig. 11.17). The estimated parameters are very well suited to produce a relatively linear curve on the distal side, and a sufficiently curved one on the proximal side. Indeed, the proximal side is not reproduced as well as in the case of the EGD model, however there are only two “bad” points (those of the oldest two filaments),



(a) Fit of EGD and AEGD models to data, for the proximal half



(b) Fit of EGD and AEGD models to data, for the distal half

Fig. 11.18: Comparison between the Elongation-Generation-Decay Model (blue) and the Alternative Elongation-Generation-Decay Model (red). Data is presented by red asterisks.

which nevertheless are found on the upper borders of the error bars. Figure 11.18 presents the comparison between the two models under consideration (EGD and AEGD).

11.3.5 Results of the Space-Dependent Elongation-Generation-Decay Model

As previously presented, let us review the Space-Dependent Elongation-Generation-Decay (SAEGD) Model from (9.7). As mentioned before, this model corresponds to a hypothesis stating that nutrients are transported from the distal to the proximal extreme. For this reason, as the distal niche is placed at the source of nutrients, the proliferation of distal *brSCs* q^+ is kept constant ($\beta^+ = 0$). The proliferation rate $q^-(t)$ depends on the number of filaments existing at the time t, as the more filaments are created, the farther away is the proximal niche from the source of nutrients.

$$\begin{aligned}
&\frac{du_i(t)}{dt} = \frac{p}{1+\alpha u_i(t)}\left(f_i^+(t) + f_i^-(t)\right), \quad u_i(t_i) = 0 \\
&\text{with } f_i^+(t) = \begin{cases} 1, & q^+ t \geq i \geq 0 \\ 0, & \text{else} \end{cases}, \quad f_i^-(t) = \begin{cases} 1, & q^-(t)t \leq i \leq 0 \\ 0, & \text{else} \end{cases} \\
&\text{and } q^-(t) = r^-/(1+\beta^-\left(\lfloor q^+ t \rfloor - \lceil q^-(t)t \rceil\right)) \quad \text{with } t_i = \frac{i}{q^{\pm}(t)}
\end{aligned} \tag{11.4}$$

The complication in this model comes from the implicit equation for $q^-(t)$, which in addition contains floor and ceiling terms. As this cannot be solved analytically, we take advantage of *MATLAB*'s routine `fzero`. This routine is used to find the zeros of a function, which in this case is $z(x) = x - r^-/(1+\beta^-\left(\lfloor q^+ t \rfloor - \lceil xt \rceil\right))$, at each time point t. `fzero` relies on the algorithm created by T. Dekker [88], which uses a combination of bisection, secant and inverse interpolation methods.

The `fzero` routine is used to find $q^-(t)$, which is needed to compute the times t_i for both sides. Once we have the times at which filaments are generated, we can compute the solution. An issue in these simulations, apart from the long computational time, is that generated by rounding to integer numbers. For finding the velocity $q^-(t)$ and times t_i one needs to apply the `fzero` routine *at every time point*. This is a problem as the time points at which this computations are made depend on the size of our time step. Hence, the simulation might skip certain t_i's due to a time step which is too large, thus resulting in false conclusions. In order to simplify the simulations of the SAEGD model and to avoid the aforementioned issue, one can perform one extra approximation, by removing the floor and ceiling from the formula for $q^-(t)$. As a result, the implicit equation for the proliferation of stem cells in the proximal niche reads

$$q^-(t) = r^- / \left(1+\beta^-\left(q^+ t - q^-(t)t\right)\right),$$

and after further arithmetics this eventually gives

$$\begin{cases} q^-(t) = \dfrac{1+\beta^- q^+ t - \sqrt{(1+\beta^- q^+ t)^2 - 4\beta^- r^- t}}{2\beta^- t} \\ t_i = \dfrac{i(i\beta^- - 1)}{i\beta^- q^+ - r^-}, \quad i < 0 \end{cases} \tag{11.5}$$

This approximation thus simplifies the simulations, provides correct times t_i by using an analytical formula for their computation and speeds up the estimation of parameters. The fit of this variant of the SAEGD model (from now on referred to as SAEGD2) can be seen in Figure 11.19.

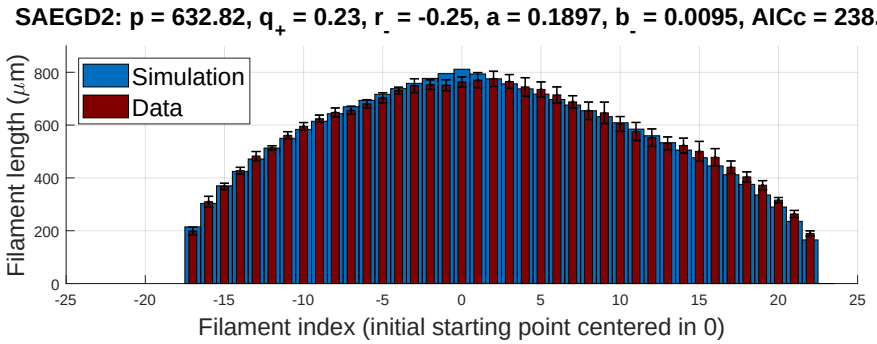


Fig. 11.19: Fit of Space-Dependent Elongation-Generation-Decay Model without floor and ceiling, for optimized parameters.

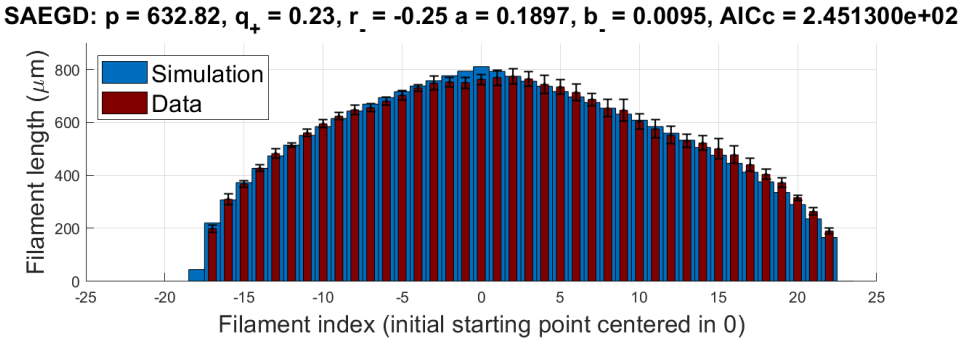


Fig. 11.20: Fit of Space-Dependent Elongation-Generation-Decay Model, for parameters used in Fig. 11.19.

11.3.6 Model Selection

By comparing the *AICc* values of the previous models (see Table 11.1), the best one out of those considered can be chosen.

Model:	LG	ED	EGD	AEGD	SAEGD	SAEGD2
AICc:	364.09	231.54	219.07	164.02	245.13	238.11

Table 11.1: Overview of *AICc* values for the models considered

It can be easily read that the Alternative-Elongation-Generation-Decay Model reproduces the data the best, while at the same time not introducing too many parameters, and thus respecting the Principle of Parsimony.

First, the "best" model suggests that filaments slow down their growth, the longer they are. Second, the stem cells in the branchial arch niches reduce their proliferation with the generation of filaments. In the following, we will focus on the "best" model AEGD, and on the SAEGD model, which will be considered because of its different assumptions and structure. The two models are based on the following two hypotheses:

- ◇ Each niche advances slower based on the amount of time that passed. This is equivalent to saying that the proliferation of the stem cells in each niche is reduced with the number of filaments it generated. This concept is suggested by the AEGD model and is supported by the reasoning that stem cells proliferate less, the older they are or the more divisions they have already had.
- ◇ The distal niche advances at a constant speed, while the proximal one decreases its speed with the total number of filaments in the branchial arch. The SAEGD/SAEGD2 model presents this option and is based on the reasoning that the proliferation of stem cells in the distal niche stays constant, while that of the stem cells in the proximal niche slows down due to its increasing distance to the nutrient source (placed at the distal niche).

A remark that holds for all of the decay models (EGD, AEGD, SAEGD, SAEGD2) is that various pairs of parameter values (p, α) result in similar fits. This is understandable as there is a tight correlation between these two parameters influencing the growth of existing filaments. Fig. 11.21 shows the correlation between parameters p and α as a

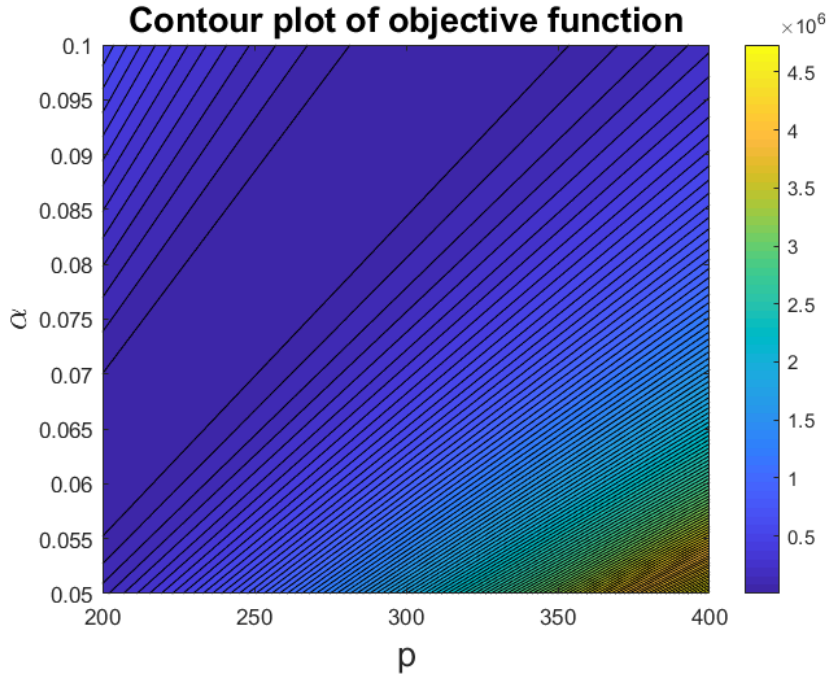


Fig. 11.21: Contour plot of the objective function with respect to parameters p and α.

contour plot of the objective function with respect to the two parameters, in the case of the AEGD model. Nevertheless, for all different sets of parameters, with various pairs (p, α), the ratio p/α stays relatively invariant. For example, for the same R1 branchial arch (the one used for comparing the models), different pairs (p, α) give an approximate ratio $p/\alpha \simeq 3200$. This holds because the height of the oldest filament (centred at 0) is approximately equal to $10\sqrt{2}\sqrt{p/\alpha}$. We accept the non-identifiability of parameters, as each parameter has a biological interpretation, but take this into account when drawing conclusions about the biological system.

11.4 Applications of the selected model

So far, the above models have been compared based on one R1 branchial arch for which seven measurements were performed. A next step is to apply the selected model to further data from various branchial arches. In this section we will use the Alternative-Elongation-

Generation-Decay Model, which has proven to be the best one out of the models under consideration, according to the AICc value.

In the following, we will focus on right branchial arches (R1-R4) and we will inspect the parameters influencing properties of the stem cells by comparing their values for arches of various sizes from different fish. In particular, we wish to get an idea about the relation between the proliferation of stem cells in arch niches compared to those in filament niches. In addition, we can investigate the reasons of why R4 arches, which have less and smaller filaments than R1 arches, for example, look more symmetric.

Out of the 12 fish imaged, not all branchial arches could be measured. Table 11.2 below shows measurements of which branchial arches are available from each fish.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
R1	×	×	-	×	×	×	×	-	-	×	×	-
R2	×	-	-	×	×	×	×	×	-	×	×	×
R3	×	-	×	-	×	×	-	×	-	-	-	-
R4	×	×	×	×	×	×	-	-	×	-	-	×

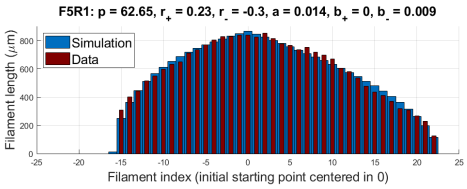
Table 11.2: Overview of the data available for each fish (F1-F12). "×" available, "-" missing.

Thus for fish F1, F5 and F6 data from all four right-branchial arches are available. Figure 11.22 shows the fit of the selected model to the four right branchial arches from the same fish (F5).

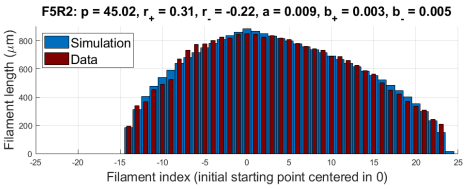
A summary of the influence of different parameters on shapes is as follows:

- ◇ As previously stated, the ratios p/α are an indication of the approximate length of the oldest filament.
- ◇ In addition, $\frac{r^+}{r^-}\frac{1+\beta^-}{1+\beta^+}$ is a measure of the symmetry around the oldest filament. The closer its value is to 1, the more symmetric the arch is, where the two halves meet.
- ◇ Finally, the values of parameters β^+ and β^- give the curvatures of each half-arch. The smaller the values are, the more linear the trend of the succession of filaments is.

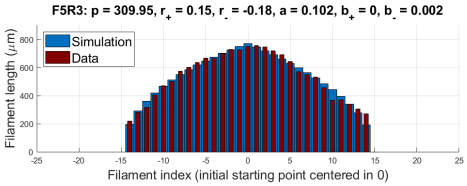
In comparison, the SAEGD2 model provides a relatively good fit, as well, for the same branchial arches from Fig. 11.22, as can be seen in Fig. 11.23. However in some cases the curvatures cannot be captured well enough.



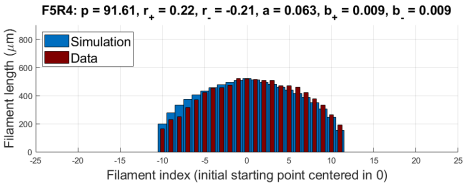
(a) R1 arch of fish F5



(c) R2 arch of fish F5

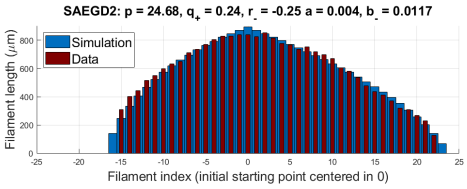


(b) R3 arch of fish F5

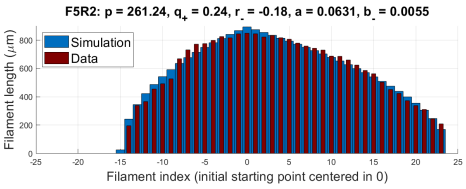


(d) R4 arch of fish F5

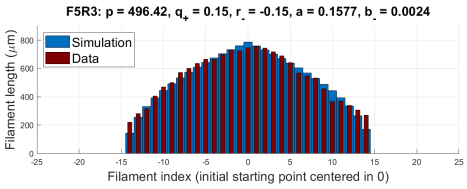
Fig. 11.22: Fit of the AEGD model to branchial arches R1-R4 from fish F5.



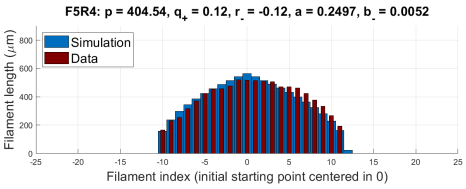
(a) R1 arch of fish F5



(c) R2 arch of fish F5



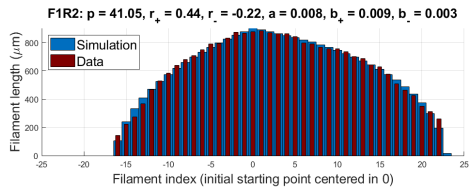
(b) R3 arch of fish F5



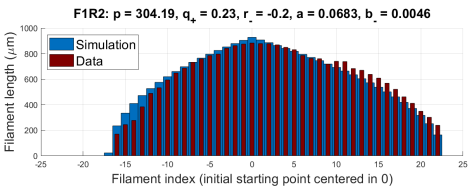
(d) R4 arch of fish F5

Fig. 11.23: Fit of the SAEGD2 model to branchial arches R1-R4 from fish F5.

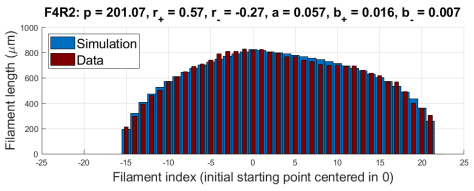
There exist few branchial arches from other fish, where the skewness is not towards the proximal side as before, but towards the distal one, instead. This could either result from a technical problem or from some underlying biological reason. In these cases we still expect the AEGD to be able to provide a good fit, while the SAEGD2 model to fail at this task, as can be seen in Fig. 11.24. From the way the SAEGD2 model was defined and implemented, it only allows for skewness to the proximal side, which prevents it from providing a more curved aspect on the distal side due to its parameter β^+ being set to 0. On the other hand, the AEGD model can easily adapt to any kind of shape, as more variability in the parameters is allowed, supporting the hypothesis of ageing stem cells in the detriment of that of nutrient transport



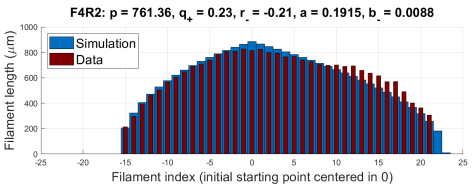
(a) Fit of AEGD model to arch R2 of fish F1



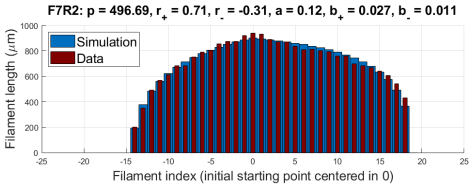
(d) Fit of SAEGD2 model to arch R2 of fish F1



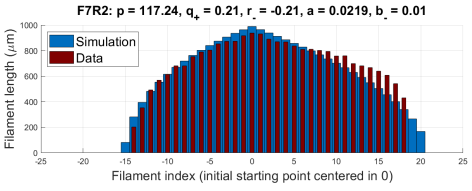
(b) Fit of AEGD model to arch R2 of fish F4



(e) Fit of SAEGD2 model to arch R2 of fish F4



(c) Fit of AEGD model to arch R2 of fish F7



(f) Fit of SAEGD2 model to arch R2 of fish F7

Fig. 11.24: Comparison of the AEGD and SAEGD2 models on their ability to fit branchial arches R2 from fish F1, F4 and F7.

Returning to the AEGD model, Fig. 11.25 presents “screen shots” at various time points of a simulation of the R1 arch from fish F1 previously used for model selection, with

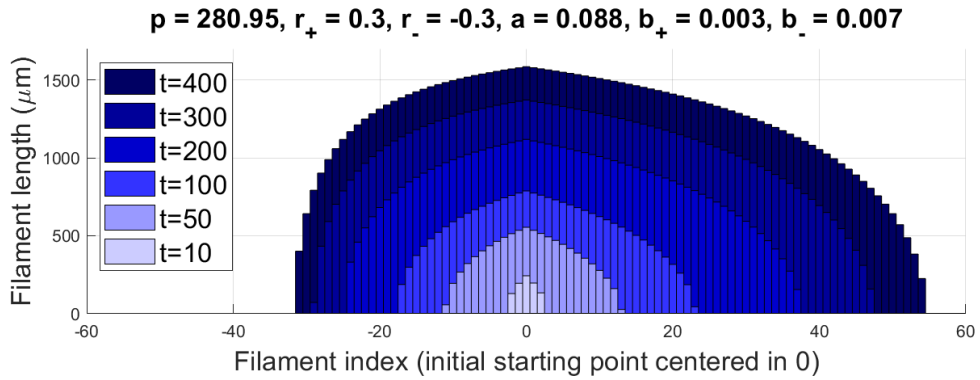
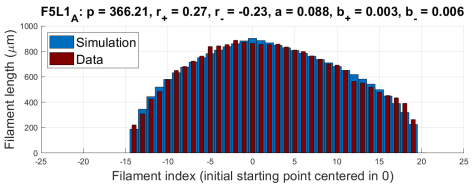


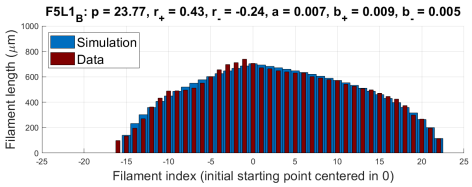
Fig. 11.25: Bar plot of the branchial arch R1 from fish F1 simulated by the AEGD model at multiple time points: 10, 50, 100, 200, 300 and 400.

the estimated parameters. As a previous remark stated, when the arch is small it is very symmetric and relatively "triangular". Subsequently, as it grows bigger in both number of filaments and length of filaments it becomes more curved and asymmetric, such that at $t = 400$ the distal side has almost twice as many filaments as the proximal side. This is a consequence of the decays α and $\beta^{\pm}$.

Further similar investigations can be performed on filaments on the B side of the arches, as well as on the left branchial arches. Figures 11.26 and 11.27 show fits of the AEGD model to left arch L1 (side A and B) and the arches of side B (R1-R4), all from fish F5. One can easily observe that the AEGD model is able to nicely reproduce the shapes of all the arches, which have similar profiles to those seen on the right part of side A.

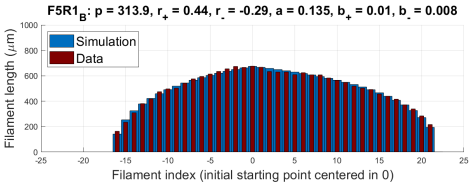


(a) Arch L1, side A, from fish F5

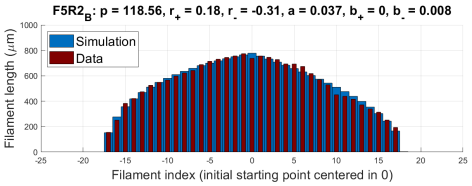


(b) Arch L1, side B, from fish F5

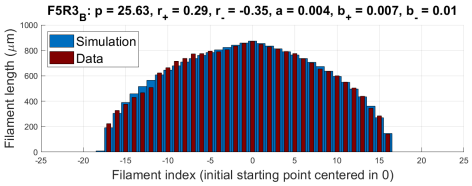
Fig. 11.26: Application of the Alternative-Elongation-Generation-Decay Model to fit data from L1 branchial arches on the A (a) and B (b) sides, from fish F5.



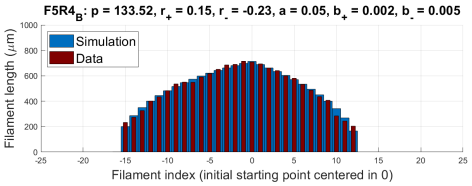
(a) Arch R1, side B, from fish F5



(b) Arch R2, side B, from fish F5



(c) Arch R3, side B, from fish F5



(d) Arch R4, side B, from fish F5

Fig. 11.27: Application of the Alternative-Elongation-Generation-Decay Model to fit data from branchial arches on the B side: R1 (a), R2 (b), R3 (c), R4 (d), all from fish F5.

Chapter 12
Conclusion and Discussion for the Organ Growth and Shape

This part of the thesis has dealt with the study of properties of various stem cells and their relation to each-other, within the respiratory organs of *Oryzias latipes* Japanese rice fish. The goal was to gather information about dynamics of filament stem cells (*filSCs*) and branchial arch stem cells (*brSCs*) by observing and modelling the shapes of branchial arches. In addition, models of various hypotheses were compared in order to speculate which is more plausible. The data for this project recorded the lengths of filaments from multiple branchial arches. A total of 12 fish were imaged, resulting in a total of 192 branchial arches, out of which we focused on the right (R1-R4) A-side ones (30 arches measured). One particular R1 arch was measured multiple times, resulting in an average value and a standard deviation per filament.

We started by developing various compartmental mathematical models, gradually increasing in complexity, consisting of a system of ordinary differential equations coupled to algebraic equations *via* indicator functions, in order to deal with the discrete growing domain, i.e. the increasing number of compartments. Next, we fitted each model to the R1 branchial arch mentioned before, and compared the models *via* the Akaike Information Criterion (AICc) for model selection.

- ◇ The Linear-Growth model, as the name suggests, assumed a constant proliferation in both *filSCs* and in *brSCs*, and produced a triangular shape, thus resulting in a very poor fit.
- ◇ The Elongation-Decay Model assumed a length-dependent decaying proliferation of *filSCs*, while the *brSCs* have a constant division rate, as before. The decay was mod-

elled *via* an exponential function with negative exponent. This model was able to capture the curved aspect of the arches, but the fit was quite poor, nevertheless.

- ◇ The Elongation-Generation-Decay Model added one extra level of complexity, by introducing a time-dependent decay in the proliferation of *brSCs*, in addition to the aforementioned length-dependent decay in the *filSCs* growth rate. This new decay was modelled by a time-dependent negative feedback. The EGD model produced a much better fit, but could not reproduce the initial fast growth of small filaments.
- ◇ The Alternative-Elongation-Generation-Decay Model is an alternative to the previous model, which used a negative feedback function for the decay in proliferation of *filSCs*. This decay depended on the length of filaments and was able to recapitulate the initial rapid growth of young filaments, as desired. By choosing different values for the parameters, one could not only reproduce the curved shape of the arch but also its asymmetry and the skewness given by the aspect of each half-arch: relatively linear on the distal side and highly curved on the proximal one. Such model corresponds to the idea of stem cells ageing, thus affecting their proliferation in time. Another hypothesis that could equally well be modelled by the AEGD model is that of decay of stem cells proliferation due to signalling molecules produced and degraded in the arches and consumed by the stem cells.
- ◇ The Space-dependent Elongation-Generation-Decay Model changed the formula for the proliferation of *brSCs* from having a time-dependent decay to one dependent on the total number of filaments existing at each time point. This model is inspired by the hypothesis of nutrients transport through the blood vessel at the base of filaments, from the distal towards the proximal side, meaning that the larger the branchial arch is, the longer it is needed by the *brSCs* in the proximal niche to receive the required amount of nutrients for their division, while the distal niche is located at the source of nutrients, in some sense.

The results of the mathematical modelling indicate the AEGD model as the best out of the models under consideration, suggesting that the aging hypothesis is more plausible than the nutrients one. This is also supported by the variability in the sizes and shapes of branchial arches, which could be nicely reproduced by the AEGD model, while the SAEGD model fails to provide a good fit in some cases.

Nevertheless, a further aspect that remains to be investigated is the difference in size between the arches within the same fish, considering that they are all studied at the same moment in the life of the fish. The difference in the asymmetry between R1 and R4 arches, for example, can be understood *via* the models presented, as the decay in proliferation influences less the branchial arches with few shorter filaments than those with more longer filaments. However, why are the parameters controlling the proliferation of stem cells so different between large and small arches, considering that they are analysed at the same time point? What biological factors influence those parameters? A possible reason could be that of space constraints, as branchial arches are hidden under the operculum, possibly exerting higher pressure on the growth of more ventral branchial arches. In order to test this hypothesis, our collaborators have grown fish which are not constrained on one side of their gills, i.e. the "bone cover" on one side of the gills (the operculum) was removed. This will allow the comparison between the size and shape of the two sides of the gills, which might shed some light on these aspects.

Furthermore, a better validation of the models could be performed when time-series data are available. With an appropriate gill cells marker, live imaging could be performed, since branchial arches could be observed through the transparent operculum at juvenile stages, but such a marker still needs to be discovered.

The study of stem cell behaviour during post-embryonic gill growth could aid in understanding mammalian embryonic development, as fish adult stem cells behave similarly to higher vertebrates embryonic stem cells. Mammals, and in particular humans, possess so-called pharyngeal arches (also known as branchial arches) during embryonic stages, which are similar to those found in fish. They are organised in the same manner as in fish, but are higher in number, specifically six pairs of arches, out of which the fifth regresses rapidly. The remaining ones develop into important facial features: the first branchial arch gives rise to the jaw, while the second one becomes the hyoid bone and jaw support, for example. Dysfunction in the development of human pharingeal arches can result in diseases such as the Pierre-Robin syndrome or the DiGeorge syndrome. These cause airway obstruction or speech and language impairments, in addition to facial features abnormalities, to name a few.

Therefore, our studies can have further applications in stem cell systems driving growth, in particular in early developmental stages or in pathological settings in which growth is uncoordinated or disrupted.

Chapter 13
Summary of the Thesis

This thesis has been devoted to the study of stem cells behaviour governing the post-embryonic growth of fish by developing mathematical models and applying them to experimental data provided by our collaborators. The focus has been on the respiratory organ of medaka fish, which due to its structured organisation enables a smooth approach *via* mathematical modelling. The thesis has been split into two parts corresponding to two projects which investigated different but related properties of stem cells driving the growth of medaka gills. A variety of mathematical methods has been used for the studies, which were adapted to the aims of the projects and to the available data.

The aim of the first project was to determine how many stem cells are needed to build an organ, specifically the gills of medaka fish. This basic question is, in fact, difficult to approach biologically as specific stem cell markers are unknown. Mathematical modelling comes to aid in this situation and two different methods have been developed and applied to clonal data, leading to agreeing results. The two approaches, using Gillespie-type stochastic simulations and Markov chains, provided agreeing results, which enforces their validity. Our question of how many stem cells are needed to build the respiratory organ of fish has a surprisingly simple answer: only few stem cells (2-3 depending on cell type) are needed in each niche, but more importantly, the stem cells must have a very high probability to divide again (~ 0.9). The later indicates that stem cells possess long activation and quiescence phases, and that they take turns in becoming activated in order to give rise to filaments, thus explaining the long stretches of consecutive filaments carrying the same label, which are observed in the data.

A relative issue appears in the case of the fourth type of stem cells, which due to their small labelling efficiency result in a non-identifiable probability of division. This happens because the pattern produced by these stem cells can be concealed beneath other patterns, when mixed patterns are observed. The non-identifiability could be avoided if specific stem cell markers for each cell type existed, allowing to label each cell type (hence pattern) with a different colour so that patterns are no longer missed, thus allowing to compute the labelling efficiencies more accurately.

The second project studied properties of gill stem cells driving growth of the organ along two orthogonal axes by inspecting the shape of branchial arches. Stem cells found in niches at the basal peripheries of the arches generate filaments, extending the arches horizontally, while stem cells located at the tips of filaments drive their elongation, thus extending the arches vertically.

Mathematical modelling *via* ordinary differential equations (ODEs) was used to reproduce the shapes of branchial arches thus allowing to draw conclusions about the properties of stem cells and to speculate on hypotheses leading to those properties. Five different ODE models were developed and compared to data recording filament lengths of branchial arches from various fish. The best out of these models were selected, by fitting to one representative branchial arch and computing the resulting Akaike Information Criterion for model comparison. All models are composed of a system of ODEs describing the change of filament lengths in time, coupled to two indicator functions indicating whether a certain filament exists at that time, and two algebraic equations governing the proliferation of branchial arch stem cells found in the peripheral niches. The differences between models stem from changing the mathematical functions which describe the proliferation of the two stem cell types. For all models, the existence, uniqueness and non-negativity of solutions is proven and their analytical solution is computed. Subsequently, numerical simulations are performed and parameters are estimated.

The insights gathered from all models presented indicate that both filament stem cells and branchial arch stem cells slow down their proliferation in time, thus slowing the growth along the two orthogonal axes. Various hypotheses about why this cell cycle lengthening happens are formulated and the mathematical modelling was able to speculate against one of them. It is believed that filament stem cells slow down their proliferation due to “starvation” of nutrients, but the same hypothesis cannot hold for the branchial arch

stem cells. Other plausible hypotheses are ageing of stem cells or signals from signalling molecules.

Further more accurate investigations could be performed if time-series data were available. Unfortunately, in the absence of a specific gill stem cells marker such data cannot be gathered as fish need to be sacrificed. Were such a marker to be available, fish could be live-imaged as labelled branchial arches could be observed through the operculum under a microscope. Our collaborators are currently searching for such a stem cell marker, and once time-series data would be available, the mathematical models could be compared not only to the final "snapshot" of the branchial arches, but at various time points, thus being better validated and possibly adjusted based on new insights from such a type of data.

In addition, a further question arises from the differences in shapes of large versus small branchial arches. Large arches are highly asymmetric and curved compared to smaller ones. The mathematical models do explain this by showing that in small arches the decays do not act as strongly as in large ones, but this is accomplished by big differences in parameters, which are needed for observing such different shapes at the same time point. The question here is why do the small arches grow much slower? What influences their parameters driving them to grow so slow in comparison to the other ones? One hypothesis considered is based on constraints from the operculum covering the gills. To test whether this is the case, our collaborators have been growing fish which had their operculum removed on one side, in order to inspect whether the arches on that side grow much bigger than the ones that are still covered. The fish have not yet been sacrificed and imaged but if this is the case, new extended mathematical models including the constraints from the operculum should be developed.

References

1. J. K. Percus and S. Childress, *Mathematical models in developmental biology*, vol. 26. American Mathematical Soc., 2015.
2. P. K. Maini and R. E. Baker, "Developmental biology: Mathematical modelling of development," *e LS*, 2001.
3. O. Diekmann, H. Heesterbeek, and T. Britton, *Mathematical tools for understanding infectious disease dynamics*, vol. 7. Princeton University Press, 2012.
4. J. Müller and C. Kuttler, *Methods and models in mathematical biology*. Springer, 2015.
5. K. P. Hadeler, *Topics in mathematical biology*. Springer, 2017.
6. M. Iannelli and A. Pugliese, *An Introduction to Mathematical Population Dynamics: Along the Trail of Volterra and Lotka*, vol. 79. Springer, 2015.
7. J. Milton and T. Ohira, *Mathematics as a Laboratory Tool*. Springer, 2016.
8. E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, *Systems biology in practice: concepts, implementation and application*. John Wiley & Sons, 2008.
9. J. Wittbrodt, A. Shima, and M. Schartl, "Medaka - a model organism from the far east," *Nature Reviews Genetics*, vol. 3, no. 1, p. 53, 2002.
10. D. H. Evans, P. M. Piermarini, and K. P. Choe, "The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste," *Physiological reviews*, vol. 85, no. 1, pp. 97–177, 2005.
11. P. Laurent, "Gill internal morphology," in *Fish physiology*, vol. 10, pp. 73–183, Elsevier, 1984.

12. I. Leguen, "Gills of the medaka (oryzias latipes): A scanning electron microscopy study," *Journal of morphology*, vol. 279, no. 1, pp. 97–108, 2018.
13. D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The journal of physical chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
14. T. Székely Jr and K. Burrage, "Stochastic simulation in systems biology," *Computational and structural biotechnology journal*, vol. 12, no. 20-21, pp. 14–25, 2014.
15. C. L. Vestergaard and M. Génois, "Temporal Gillespie algorithm: Fast simulation of contagion processes on time-varying networks," *PLoS computational biology*, vol. 11, no. 10, p. e1004579, 2015.
16. F. Ziebell, S. Dehler, A. Martin-Villalba, and A. Marciniak-Czochra, "Revealing age-related changes of adult hippocampal neurogenesis using mathematical models," *Development*, vol. 145, no. 1, p. dev153544, 2018.
17. M. Lachowicz, "Individually-based markov processes modeling nonlinear systems in mathematical biology," *Nonlinear Analysis: Real World Applications*, vol. 12, no. 4, pp. 2396–2407, 2011.
18. M. C. Gibson, A. B. Patel, R. Nagpal, and N. Perrimon, "The emergence of geometric order in proliferating metazoan epithelia," *Nature*, vol. 442, no. 7106, p. 1038, 2006.
19. G. De Vries, T. Hillen, M. Lewis, B. Schõnfisch, *et al.*, *A course in mathematical biology: quantitative modeling with mathematical and computational methods*, vol. 12. Siam, 2006.
20. J. Stolper, "Post embryonic growth and fate restriction of stem cells *in vivo*," Master's thesis, Ruperto-Carola University of Heidelberg, Germany, 2015.
21. J. Stolper, E. M. Ambrosio, D.-P. Danciu, D. Elliott, K. Naruse, A. Marciniak-Czochra, and L. Centanin, "Hierarchical stem cell topography splits growth and homeostatic functions in the fish gill," *eLife*, vol. 6, p. e43747, 2019.
22. C. Gaillochet, T. Stiehl, C. Wenzl, J.-J. Ripoll, L. J. Bailey-Steinitz, L. Li, A. Pfeiffer, A. Miotk, J. P. Hakenjos, J. Forner, *et al.*, "Control of plant cell fate transitions by transcriptional and hormonal signals," *Elife*, vol. 6, p. e30135, 2017.
23. G. Oster, J. Murray, and P. Maini, "Forces and pattern in limb morphogenesis," 1984.
24. T. Stiehl, A. Ho, and A. Marciniak-Czochra, "The impact of CD34+ cell dose on engraftment after SCTs: personalized estimates based on mathematical modeling," *Bone marrow transplantation*, vol. 49, no. 1, p. 30, 2014.

25. M. Mohr, D. Hose, A. Seckinger, and A. Marciniak-Czochra, "Quantification of plasma cell dynamics using mathematical modelling," *Royal Society open science*, vol. 5, no. 1, p. 170759, 2018.
26. T. Stiehl and A. Marciniak-Czochra, "Stem cell self-renewal in regeneration and cancer: insights from mathematical modeling," *Current Opinion in Systems Biology*, vol. 5, pp. 112–120, 2017.
27. M. V. Barbarossa, M. Polner, and G. Rost, "Stability switches induced by immune system boosting in an sirs model with discrete and distributed delays," *SIAM Journal on Applied Mathematics*, vol. 77, no. 3, pp. 905–923, 2017.
28. T. Stiehl and A. Marciniak-Czochra, "Mathematical modeling of leukemogenesis and cancer stem cell dynamics," *Mathematical Modelling of Natural Phenomena*, vol. 7, no. 1, pp. 166–202, 2012.
29. T. Stiehl, N. Baran, A. D. Ho, and A. Marciniak-Czochra, "Clonal selection and therapy resistance in acute leukaemias: mathematical modelling explains different proliferation patterns at diagnosis and relapse," *Journal of The Royal Society Interface*, vol. 11, no. 94, p. 20140079, 2014.
30. R. E. Baker, C. A. Yates, and R. Erban, "From microscopic to macroscopic descriptions of cell migration on growing domains," *Bulletin of mathematical biology*, vol. 72, no. 3, pp. 719–762, 2010.
31. M. J. Simpson, "Exact solutions of linear reaction-diffusion processes on a uniformly growing domain: Criteria for successful colonization," *PloS one*, vol. 10, no. 2, p. e0117949, 2015.
32. E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, and R. Herwig, *Systems biology: a textbook*. John Wiley & Sons, 2016.
33. G. An, Q. Mi, J. Dutta-Moscato, and Y. Vodovotz, "Agent-based models in translational systems biology," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 1, no. 2, pp. 159–171, 2009.
34. E. Tsingos, B. Hoeckendorf, T. Suetterlin, S. Kirchmaier, N. Grabe, L. Centanin, and J. Wittbrodt, "Retinal stem cells modulate proliferative parameters to coordinate post-embryonic morphogenesis in the eye of fish," *eLife*, vol. 8, p. e42646, 2019.
35. T. Stiehl, C. Lutz, and A. Marciniak-Czochra, "Emergence of heterogeneity in acute leukemias," *Biology direct*, vol. 11, no. 1, p. 51, 2016.

36. S. K. Hahl and A. Kremling, "A comparison of deterministic and stochastic modeling approaches for biochemical reaction systems: On fixed points, means, and modes," *Frontiers in genetics*, vol. 7, p. 157, 2016.
37. T. Stiehl, A. D. Ho, and A. Marciniak-Czochra, "Mathematical modeling of the impact of cytokine response of acute myeloid leukemia cells on patient prognosis," *Scientific reports*, vol. 8, no. 1, p. 2809, 2018.
38. D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of computational physics*, vol. 22, no. 4, pp. 403–434, 1976.
39. M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *The journal of physical chemistry A*, vol. 104, no. 9, pp. 1876–1889, 2000.
40. C. J. Morton-Firth, *Stochastic simulation of cell signalling pathways*. PhD thesis, University of Cambridge, 1998.
41. J. R. Norris, *Markov chains*. No. 2, Cambridge University Press, 1998.
42. G. Grimmett, D. Stirzaker, *et al.*, *Probability and random processes*. Oxford University Press, 2001.
43. L. J. Allen, *An introduction to stochastic processes with applications to biology*. Chapman and Hall/CRC, 2010.
44. A. J. Lotka, "Analytical note on certain rhythmic relations in organic systems," *Proceedings of the National Academy of Sciences*, vol. 6, no. 7, pp. 410–415, 1920.
45. V. Volterra, "Fluctuations in the abundance of a species considered mathematically," 1926.
46. W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927.
47. P. Hartman, *Ordinary differential equations*. John Wiley and Sons, Inc. New York-London-Sydney, 1964.
48. H. G. Bock, E. Kostina, H. X. Phu, and R. Rannacher, *Modeling, simulation and optimization of complex processes*. Springer, 2005.

49. M. Peifer and J. Timmer, "Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting," *IET Systems Biology*, vol. 1, no. 2, pp. 78–88, 2007.
50. K. Jaqaman and G. Danuser, "Linking data to models: data regression," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 11, p. 813, 2006.
51. J. R. Banga and E. Balsa-Canto, "Parameter estimation and optimal experimental design," *Essays in biochemistry*, vol. 45, pp. 195–210, 2008.
52. D. R. Cox and D. V. Hinkley, *Theoretical statistics*. Chapman and Hall/CRC, 1979.
53. J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.
54. W. Research, "Mathematica 8.0 documentation," 2015. Wolfram Research Inc., Champaign, Illinois, USA.
55. M. R. Celis, J. Dennis, and R. Tapia, "A trust region strategy for nonlinear equality constrained optimization," *Numerical optimization*, vol. 1984, pp. 71–82, 1985.
56. MathWorks, "Matlab optimization toolbox," 2018. The MathWorks, Natick, MA, USA.
57. J. J. Moré and D. C. Sorensen, "Computing a trust region step," *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 3, pp. 553–572, 1983.
58. M. A. Branch, T. F. Coleman, and Y. Li, "A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems," *SIAM Journal on Scientific Computing*, vol. 21, no. 1, pp. 1–23, 1999.
59. R. H. Byrd, R. B. Schnabel, and G. A. Shultz, "Approximate solution of the trust region problem by minimization over two-dimensional subspaces," *Mathematical programming*, vol. 40, no. 1-3, pp. 247–263, 1988.
60. T. F. Coleman and A. Verma, "A preconditioned conjugate gradient approach to linear equality constrained minimization," *Computational Optimization and Applications*, vol. 20, no. 1, pp. 61–72, 2001.
61. D. C. Sorensen, "Minimization of a large-scale quadratic functionsubject to a spherical constraint," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 141–161, 1997.
62. D. Anderson and K. Burnham, "Model selection and multi-model inference," *Second. NY: Springer-Verlag*, 2004.

63. K. P. Burnham and D. R. Anderson, "Multimodel inference: understanding aic and bic in model selection," *Sociological methods & research*, vol. 33, no. 2, pp. 261–304, 2004.
64. H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*, pp. 199–213, Springer, 1998.
65. H. Akaike, "A new look at the statistical model identification," in *Selected Papers of Hirotugu Akaike*, pp. 215–222, Springer, 1974.
66. H. Akaike, "Canonical correlation analysis of time series and the use of an information criterion," in *Mathematics in Science and Engineering*, vol. 126, pp. 27–96, Elsevier, 1976.
67. H. Akaike, "On the likelihood of a time series model," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 27, no. 3-4, pp. 217–235, 1978.
68. G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
69. N. Sugiura, "Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's," *Communications in Statistics-Theory and Methods*, vol. 7, no. 1, pp. 13–26, 1978.
70. M. Furutani-Seiki and J. Wittbrodt, "Medaka and zebrafish, an evolutionary twin study," *Mechanisms of development*, vol. 121, no. 7-8, pp. 629–637, 2004.
71. Y. Wakamatsu, S. Pristyazhnyuk, M. Kinoshita, M. Tanaka, and K. Ozato, "The see-through medaka: a fish model that is transparent throughout life," *Proceedings of the National Academy of Sciences*, vol. 98, no. 18, pp. 10046–10050, 2001.
72. L. Centanin, J.-J. Ander, B. Hoeckendorf, K. Lust, T. Kellner, I. Kraemer, C. Urbany, E. Hasel, W. A. Harris, B. D. Simons, *et al.*, "Exclusive multipotency and preferential asymmetric divisions in post-embryonic neural stem cells of the fish retina," *Development*, pp. dev–109892, 2014.
73. L. Centanin, B. Hoeckendorf, and J. Wittbrodt, "Fate restriction and multipotency in retinal stem cells," *Cell stem cell*, vol. 9, no. 6, pp. 553–562, 2011.
74. T. Kage, H. Takeda, T. Yasuda, K. Maruyama, N. Yamamoto, M. Yoshimoto, K. Araki, K. Inohaya, H. Okamoto, S. Yasumasu, *et al.*, "Morphogenesis and regionalization of the medaka embryonic brain," *Journal of Comparative Neurology*, vol. 476, no. 3, pp. 219–239, 2004.

75. T. Watanabe, S. Asaka, D. Kitagawa, K. Saito, R. Kurashige, T. Sasado, C. Morinaga, H. Suwa, K. Niwa, T. Henrich, *et al.*, "Mutations affecting liver development and function in medaka, oryzias latipes, screened by multiple criteria," *Mechanisms of development*, vol. 121, no. 7-8, pp. 791–802, 2004.
76. D. E. Hinton, Y. Wakamatsu, K. Ozato, and S. Kashiwada, "Imaging liver development/remodeling in the see-through medaka fish," in *Comparative hepatology*, vol. 3, p. S30, BioMed Central, 2004.
77. T. Iwamatsu, "Stages of normal development in the medaka oryzias latipes," *Mechanisms of development*, vol. 121, no. 7-8, pp. 605–618, 2004.
78. P. Laurent, G. Goss, and S. F. Perry, "Proton pumps in fish gill pavement cells?," *Archives internationales de physiologie, de biochimie et de biophysique*, vol. 102, no. 1, pp. 77–79, 1994.
79. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, p. 676, 2012.
80. C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri, "Imagej2: Imagej for the next generation of scientific image data," *BMC bioinformatics*, vol. 18, no. 1, p. 529, 2017.
81. K. R. Olson, "Vascular anatomy of the fish gill," *Journal of Experimental Zoology*, vol. 293, no. 3, pp. 214–231, 2002.
82. K. R. Olson, "Secondary circulation in fish: anatomical organization and physiological significance," *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, vol. 275, no. 2-3, pp. 172–185, 1996.
83. A. Marciniak-Czochra, T. Stiehl, A. D. Ho, W. Jäger, and W. Wagner, "Modeling of asymmetric cell division in hematopoietic stem cellsregulation of self-renewal is essential for efficient repopulation," *Stem cells and development*, vol. 18, no. 3, pp. 377–386, 2009.
84. A. Marciniak-Czochra, A. Mikelić, and T. Stiehl, "Renormalization group second-order approximation for singularly perturbed nonlinear ordinary differential equations," *Mathematical Methods in the Applied Sciences*, vol. 41, no. 14, pp. 5691–5710, 2018.

85. J. D'Errico, "SLM - Shape Language Modeling." http://www.mathworks.com/matlabcentral/fileexchange/24443-slm-shape-language-modeling, 2012.
86. G. M. Cailliet, W. D. Smith, H. F. Mollet, and K. J. Goldman, "Age and growth studies of chondrichthyan fishes: the need for consistency in terminology, verification, validation, and growth function fitting," *Environmental Biology of Fishes*, vol. 77, no. 3-4, pp. 211–228, 2006.
87. R. T. Leaf, Y. Jiao, B. R. Murphy, J. I. Kramer, K. M. Sorensen, and V. G. Wooten, "Life-history characteristics of japanese medaka oryzias latipes," *Copeia*, vol. 2011, no. 4, pp. 559–565, 2011.
88. T. Dekker, "Finding a zero by means of successive linear interpolation," *Constructive aspects of the fundamental theorem of algebra*, pp. 37–51, 1969.

Acronyms

AIC Akaike's Information Criterion - a criterion used for model comparison

BIC Bayes' Information Criterion - a criterion used for model comparison

TIC Takeuchi's Information Criterion - a criterion used for model comparison

AICc Corrected Akaike's Information Criterion - an adjusted criterion used for model comparison, for small data sample

KL Kullback-Leiber information - a measure of the distance between the model and the reality

MLE Maximum Likelihood Estimator

LS Least Squares

dpf Days post-fertilization

mpf Months post-fertilization

dph Days post-hatch

IdU Iodo-deoxyuridine - a chemical used for labelling mitotic cells

GFP Green Fluorescent Protein

RFP Red Fluorescent Protein

probLab Labelling Efficiency - probability of labelling a cell in the niche

probDiv Probability of the cell which has just divided to be the next one to divide again

LG Linear-Growth Model

ED Elongation-Decay Model

EGD Elongation-Generation-Decay Model

AEGD Alternative Elongation-Generation-Decay Model

SAEGD Space-dependent Elongation-Generation-Decay Model

filSCs Filament stem cells, located at filament tips

brSCs Branchial arch stem cells, located in niches at the basal peripheries of arches

ODE Ordinary Differential Equation

PDE Partial Differential Equation

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Prof. Anna Marciniak-Czochra for all her support ever since she took me under her wing for the internship through the HGS MathComp Post-Bachelor Programme. She initiated this collaborative project which ended up being surprisingly diverse in terms of mathematical methods needed, thus having the opportunity to learn and develop techniques from such varied branches of applied mathematics, being a perfect fit to my scientific curiosity and personality. I would like to especially thank her for indirectly teaching me through her insightful questions the importance of viewing a problem from multiple angles, of always looking for additional aspects that can be investigated and for always believing in me even when I did not, thus making me more confident in my own abilities. Further, I highly appreciated her invaluable mathematical support but even more the unexpected advice related to research, personal development and the scientific community matters.

Gratitude goes to my biology supervisor, Prof. Lázaro Centanin who supported this project from the beginning, making this journey an amazing quest for knowledge through interesting questions related to developmental biology. He made me part of his joyful group and always paused to explain the biological aspects that I could not understand even in the middle of heated biological debates which were very difficult to pause. I have always enjoyed the lab meetings, where I could not only gain insights essential to my project, but further learn so much about organ development, experimental approaches and fish in general - coming out smarter and with more interesting questions than before and looking forward to the next meeting. I especially appreciated his willingness to switch gears when modelling of available data was not possible and even more his incessant

ideas generation, always coming up with new interesting things to investigate. He could always make himself speak my language and taught me how to speak his, thus instructing me how to present my methods to scientists from various fields.

I thank HGS MathComp for the travel support which facilitated my participation in various international conferences and workshops and for organizing such interesting soft-skills courses in collaboration with the Graduate Academy which proved extremely valuable during my project.

I sincerely thank all present and former members of the “Applied Analysis and Modelling in Biosciences” group for creating a warm and vibrant research atmosphere. In particular, I thank Dr. Frederik Ziebell for being a good listener and interlocutor thus enabling me to come up with new ideas by speaking out loud. Special thanks go to Chris Kowall, who welcomed and maintained a friendly atmosphere in our office and did not mind me unexpectedly starting to speak to him, when maybe he was in the middle of something. I would further like to thank Dr. Maria Barbarossa who was always a good discussion partner and friend and made me come up with the Markov Chains approach idea for computing stem cell numbers. She also trusted me to help with tutoring even when I did not speak German good enough, thus developing teaching skills. Thanks extend to Dr. Thomas Stiehl who did not mind me spontaneously stopping by his office with interdisciplinary questions and who always had fun interesting medical facts to share. In addition, I thank Dr. Nikos Sfakianakis who helped me with some numerics aspects and who expressed great enthusiasm for the project. Last but not least, heartfelt thanks go to Dr. Marcel Mohr and Dr. Jan-Erik Busse who warmly welcomed me in the group from the very beginning during the internship, and who helped me get acquainted with and accommodated in Heidelberg city and scientific community through essential advice, German-speaking days, and from whose experiences I had a lot to learn.

Special thanks go to Julian Stolper who not only had all the patience to teach me about fish development and answer all my questions on general biological aspects, but even agreed to quench my curiosity for hands-on experiments, thus taking the time to guide me through my very first dissections. Of course, I am very grateful for his thoroughness in performing the experiments and especially in gathering the data essential for my mathematical models - I dare say we made a great team.

I conclude with gratitude towards my family and friends who supported me throughout this journey. My parents especially, because they discovered my passion for mathematics and fought hard to give me the best education possible, even through adversities such as nonsupporting teachers, material difficulties, distance, and they always advised me through difficult decisions, shaping me into an ambitious, moral person - they are my personal role models. To my grandparents, who were my first teachers from the age of three - starting with reading, writing and extending to arithmetics, all before finishing kinder-garden. Thanks extend to my friends from various corners of the world who believe in me and in particular to my boyfriend who complements me, keeps me sane and safe, makes me a stronger person, keeps my laptop “healthy” after all the intense simulations and constantly reminds me to backup my work so that I stay sane.