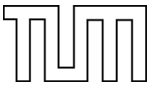

Technische Universität München

Department of Mathematics

Master's Thesis

Cell type-structured population models for the analysis of CFSE-data with continuous and discrete age structure

Anna K. Fiedler

Supervisor: Prof. Dr. Dr. Fabian Theis

Advisor: Dr. Jan Hasenauer

Submission Date: 29 November 2013

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

München,

Acknowledgements

First of all I want to express my gratitude to Prof. Dr. Dr. Fabian Theis for providing me with the intriguing topic of my master's thesis and valuable advice. I especially want to thank my supervisor Dr. Jan Hasenauer for the great guidance and advice throughout the whole writing process of my thesis, the assistance and the encouragement. Furthermore, I would like to thank Prof. Dr. med Irmela Jeremias and Sebastian Tiedt from the Helmholtz Zentrum for providing me with the proliferation data and the insightful and very interesting meetings.
I want to thank my friends for the company, the support and the great times we had together. Special thanks go to my boyfriend Moritz for his encouragement and patience. Lastly, I want to thank my family. I am so grateful to my parents for giving me the opportunity to study and for their support throughout the years.

Abstract

Cell proliferation is an important biological process and is involved in cancer progression and immune response. The study of proliferation on cell population level can be pursued by the analysis of proliferation assays of cells labeled with the fluorescent marker carboxyfluorescein succinimidyl ester (CFSE). To obtain further insight into proliferation dynamics based on experimental data, mathematical models and parameter estimation are used. Among the existing mathematical models a common approach is to structure the cell population according to certain features that can influence the cells behavior or are otherwise important in the experimental process. For example, there are models that take division number-, age- and label-structured populations into account as well as models that regard division number-, cell type- and label-structured populations. So far, no model considers population structures that incorporate all of the previously mentioned features. Hence, in this thesis we want to extend the existing models to obtain age-, label-, division number- and cell type-structured population models. We will present two models that differ in the approach to model cell age in that the first incorporates a continuous and the second a discrete age structure. Both models are a system of partial differential equations. We will assess the solution of these systems by decomposing them into a system governing the label dynamics and a system describing division number and age dynamics and the fluxes between cell types. The decomposition is used to develop numerical schemes for simulation and parameter estimation. Thereafter, these numerical schemes are deployed to analyze in vivo proliferation dynamics of leukemic cells.

Contents

Chapter 1

Introduction

1.1 Cell proliferation

In biological processes cell proliferation plays an essential role. Proliferation and subsequent increase of cellular material constitute the growth of bacterial cultures as well as the growth of multicellular organisms and the renewal of dying cells. Understanding cell proliferation is especially important in cancer research, immunology or stem cell induced tissue remodeling [2, 5]. The study of proliferation can be pursued on the scale of single cells or populations.
On the single cell level, the field of attention are the complex intrinsic and extrinsic regulatory systems that are in place to control the processes of cell division, cell growth and apoptosis and to ensure synchronised behavior of neighboring cells. If these regulatory systems fail, a possible consequence is uncontrolled tissue or population growth leading to cancer. Various cells can differ in their respective response to extrinsic stimuli thus creating heterogeneous cell populations [2, 5].
On the cell population level, which will be the focus of this thesis, the research interest lies in the growth dynamics, which are governed by cell proliferation and cell death. Hence, statistics of division and death rates are of particular interest. Both can depend on the number of divisions a cell underwent (also referred to as generation of a cell), cell age and – in a population that consist of a mixture of cell types, e.g. stem cells and more differentiated cells or a cell type that exhibits various phenotypes regarding proliferation behavior – on the cell type [2, 5, 8].

1.2 CFSE-based proliferation experiments

To track the proliferation dynamics of cell populations, proliferation assays can be used. Therefor the cells of the population are stained with a fluorescent marker, e.g. carboxyfluorescein succinimidyl ester (CFSE) or Bromodeoxy uridine (BrdU). Since commonly mostly CFSE is used, we will focus on this particular marker. To label cells with CFSE, they are kept in a medium rich of an CFSE precursor molecule (CFDA-SE) that diffuses into the cells. In the cell, the precursor is converted into the fluorescent CFSE which attaches to proteins in the cell and becomes unable to pass through the cell membrane. After the cells incorporated CFSE, the medium is changed to one not including the CFSE precursor.

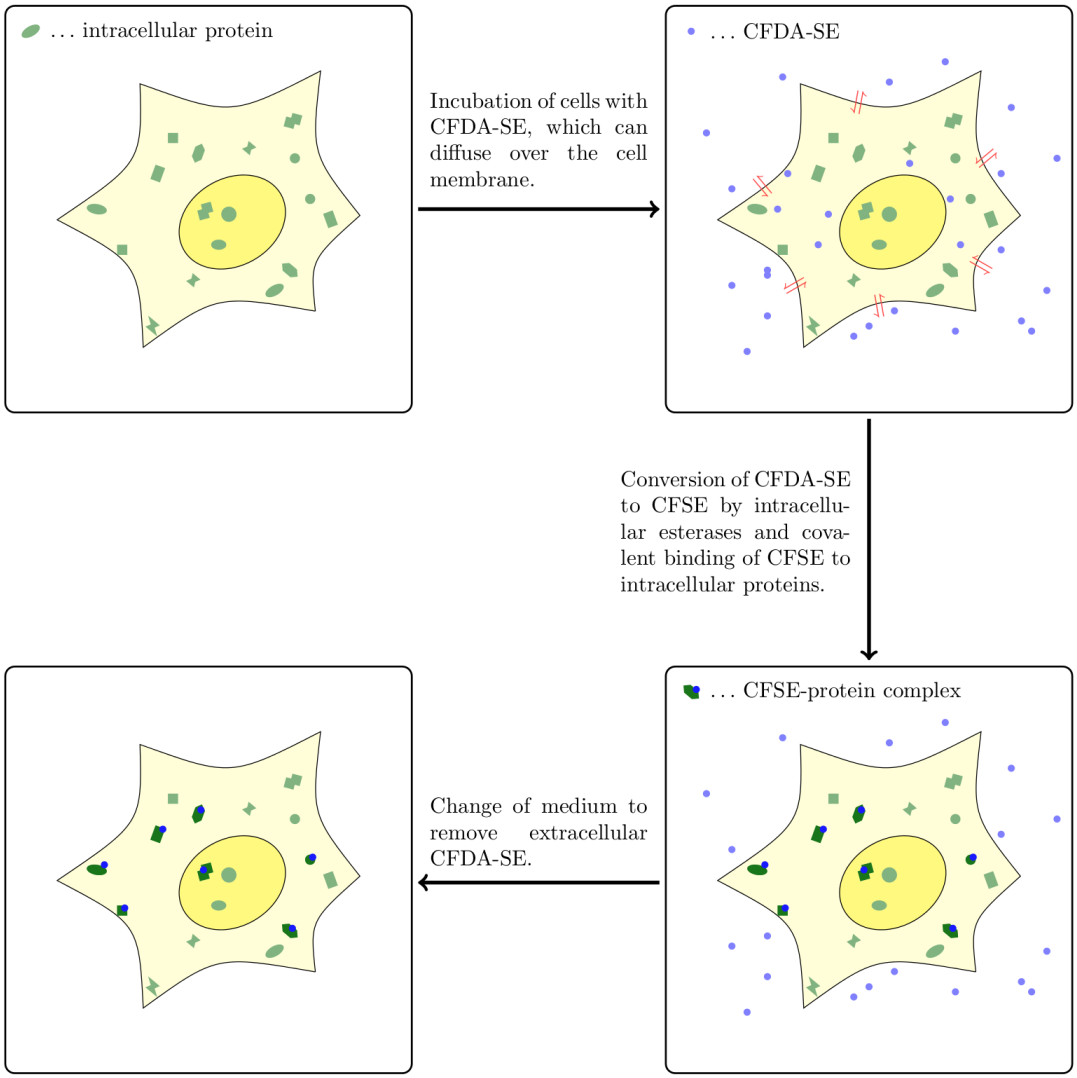

Figure 1.1: Schematic illustration of the labeling process. The CFSE precursor molecule CFDA-SE diffuses into the cell and binds to proteins in the cell forming CFSE. CFSE is unable to diffuse out of the cell. In the last step the medium is changed to one not containing CFDA-SE. The figure has been adopted from [3].

The labeling process is depicted in Figure 1.1. At sequential time points the population is analyzed via flow cytometry. For this, the fluorescence activated cell sorter (FACS) measures the fluorescence intensity of individual cells. This data can then be used to study the proliferation dynamics [1, 2, 4, 5].
Two quantities contribute to the total fluorescence intensity, the CFSE-induced fluorescence which is proportional to the amount of label and autofluorescence (also called background fluorescence) due to natural fluorescence of the cell. There are two processes that account for decline of CFSE-induced fluorescence in the individual cell. The CFSE-induced fluorescence can decreases due to protein degradation and due to cell division. If the protein CFSE is attached to is degraded, CFSE is degraded, as well. If a cell divides, the label is roughly equally distributed to the daughter cells (see Figure 1.2). If a cell dies,

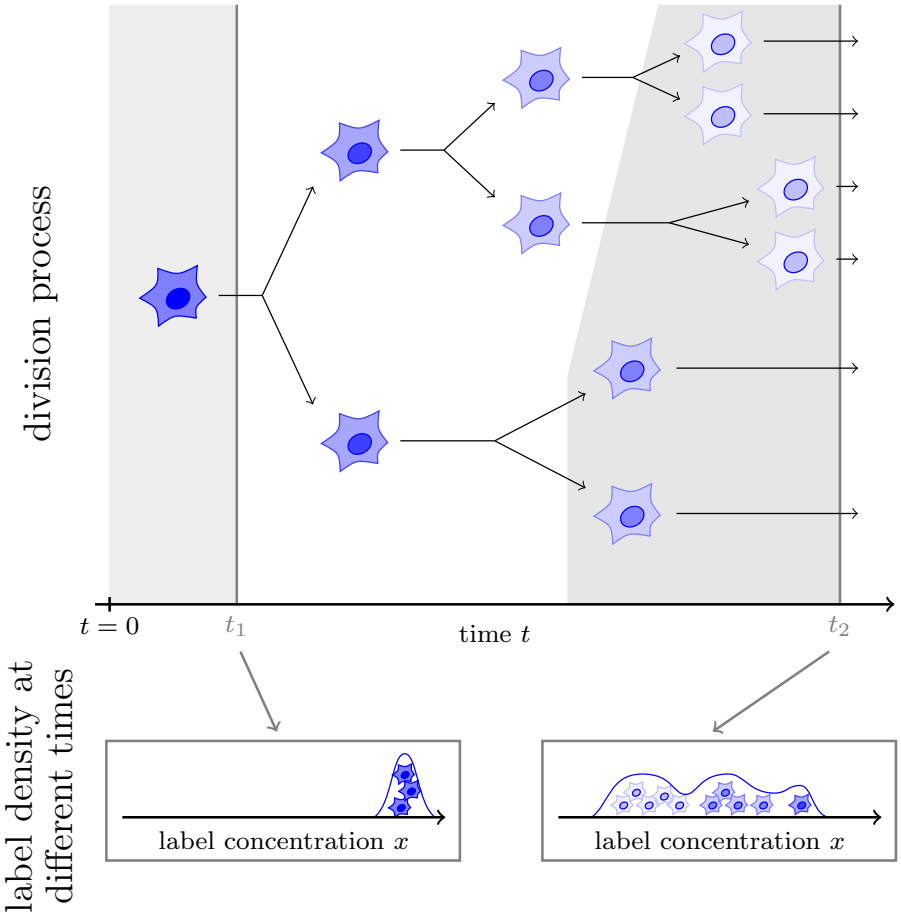

Figure 1.2: Schematic representation of label dynamics with respect to cell division showing the proliferation assays at different times. The label intensity is characterized by the transparency of color. The more transparent the color is, the more the label of the cell got diluted. Two time points are depicted by vertical lines. The grey areas indicate the cells that would be measured at these time points. In the bottom of the figure the observed fluorescence distributions for the two exemplary time points are shown. In the left one the cells did not yet divide yielding one peak. The right one shows a mixture of different generations. This illustration has been adopted from [2].

the label is degraded and the dead cell no longer contributes to the fluorescence. After a certain number of divisions the CFSE-induced fluorescence is undistinguishable from the background fluorescence. The small amount of CFSE used is not harmful for the cell and does not alter its behavior [1, 2, 4, 5]. For each considered time point the measured fluorescence data is binned and converted to histograms showing the number of cells in a certain bin of fluorescence intensity. The peaks in the histogram represent cells that underwent a certain number of divisions [2, 5].

If biological data for cell population growth is obtained by the experimental method described above, not all factors can be observed. Here mathematical models are tools to explain data, estimate missing parameters and gain further insight into the underlying biological processes.

1.3 Overview of existing proliferation models

There are several approaches to model proliferation dynamics of cell populations each focusing on different aspects [5]. Since cell populations contain a large number of cells, stochastic differences between individual cells tend to have less effect on the dynamics of the total population and can be neglected. Hence, deterministic ordinary differential equation (ODE) and partial differential equation (PDE) models are used to describe population dynamics. The simplest approach models the growth of the whole population using an ODE model with one ODE. Models using ODEs are for example exponential growth (EG) or logistic growth (LG) models. To account for differences between generations, division-structured population (DSP) models are employed. Here, a set of ODEs is applied, each ODE modeling the behavior of one generation. Age-structured population (ASP) models focus on the age dynamics of cells resulting in a single PDE in age and time. They describe the change of the age distribution in the whole population with respect to time and allow for age-dependent division and death rates. The previous models only focus on population dynamics and have therefore no relation to the proliferation assay data collected using flow cytometry. Label-structured population (LSP) models on the other hand focus solely on the distribution of label in the population. This is modeled by a single PDE in time and label concentration, yielding the change of label distribution in the population over time and enabling comparison with flow cytometry data [5].
To include more aspects and to make a more precise analysis of the data possible, the LSP and DSP model have been combined to the division- and label-structured population (DLSP) model. The DLSP model is hence capable of describing the label dynamics of subpopulations that differ in their biological properties [5].
Starting from the DLSP model, additional models were developed. The division number-, age- and label-structured population (DALSP) model developed in [5] is able to describe an age structure within the generations. Hence, it allows for age-dependent biological rates additional to division number-dependent ones. The division number-, cell type- and label-structured population (DCLSP) model introduced in [8] considers populations that consist of a mixture of different cell types. Biological rates can vary from cell type to cell type. The DCLSP model is capable of describing this variability while also accounting for differences in the behavior of generations. However, this model does not account for age structure and age-dependent rates, which can help to describe proliferation dynamics more detailed, as shown in [5]. The explicit evolution equations for the DALSP model and the DCLSP model are presented in Section 2.3.

1.4 Contribution of this thesis

None of the existing models listed in the previous section can treat both cell types and age structure. However, there are a lot of biological examples of populations with mixed cell types and age-dependent rates. Additionally, the previously mentioned models were commonly used for the analysis of data that observed only few generations. In this thesis we want to introduce novel model classes using cell type structured population models that account for age structure of the population as well, where the age structure is modeled with different approaches.

In the Chapter 2 we introduce the age-, label-, division number- and cell type-structured population (ALDC) model. This new model is a combination of the DALSP and the DCLSP model discussed in Section 1.3. It consists of a system of coupled PDEs, where each PDE describes the dynamics of label and age with respect to time for a particular generation. We extend the concepts used to analyze the DALSP model in [5] to decompose this system of PDEs into a less complex system and a set of PDEs. The first subsystem governs the label dynamics, the second subsystem describes the generation and age structure resulting from the biological processes. For these systems of PDEs the solutions are presented. We show in Section 2.3 that it is possible to relate the ALDC model to the DCLSP and DALSP model by marginalization over the respective quantities. Furthermore, a numerical scheme is outlined and compared to a gold standard that is derived from the DCLSP model.
A simulation study using the ALDC model reveals that, at least with the current implementation of the model, a consideration of *in vivo* processes with a large number of generations is not possible. Therefore, a new model class is introduced in Chapter 3 that considers a discrete age structure rather than a continuous one as in the ALDC. We thereto introduce discrete cell states motivating the name cell state-, label-, division number- and cell type-structured population (SLDC) model. It consists of a coupled system of PDEs describing the population dynamics with respect to label and time for every combination of generation, cell type and cell state. The PDE system can be decomposed into a PDE describing the label dynamics and an ODE system governing population dynamics. We further relate the SLDC model to the DCLSP and DALSP model analogously to the ALDC model. The ODE system can be solved numerically building the basis of a much more efficient numerical scheme when compared to the one developed in Chapter 2. In Chapter 4 the SLDC model is used to estimate parameters of real life proliferation data. The data displays the population dynamics of acute lymphatic leukemia (ALL) cells *in vivo*. ALL is a bone marrow and blood cancer. As treatment is started immediately after dectetion, there is little knowledge about the cancer progression *in vivo*. However, this knowledge could help to improve the prospects of patients greatly. The data is produced by extracting cancer cells from patients, labeling the cancer cells with CFSE and injecting them into mice. At certain days the cells are extracted, the mouse cells are sorted out and a fluorescence spectrogram is created. A novel preprocessing developed by the biological researchers minimizes sorting errors and achieves a new quality in the data. This sort of *in vivo* data has not yet been analyzed with the mathematical tools introduced in this thesis.

Chapter 2

Age-, label-, division number- and cell type-structured population model for CFSE-data

In this chapter the age-, label-, division number- and cell type-structured population (ALDC) model is discussed. In Section 2.1 we introduce the ALDC model, which takes into account the number of divisions a cell has undergone, i, the time that past since staining t, the label concentration x, and cell age a as in the DALSP model from [5] as well as cell types j as the DCLSP model [8]. In Section 2.2 we describe the decomposition of the ALDC model into two parts and present the solution of these. The relation of the ALDC model to existing models is studied in Section 2.3. In Section 2.4 the numerical scheme for the simulation of the model is presented and applied to a small example before we summarize the chapter in Section 2.5.

2.1 Formulation of the ALDC model

In this section the ALDC model will be established. We are going to take into account the label concentration in a certain cell $x \in \mathbb{R}_+$, the age of a cell $a \in \mathbb{R}_+$ and the time that passed since the begining of the experiment $t \in \mathbb{R}_+$. Furthermore we want to cosider the division number of a cell, $i \in \mathbb{N}_0$. The division number denotes how many times a cell divided. After each cell division the division number of a cell is increased by one and the age of the cell a is set to zero. Lastly, we take into consideration the cell type of a cell j. We distinguish the cell types by numbering them consecutively from one to the total number of cell types J, hence, $j \in \{1, \dots, J\}$. To account for all factors the joint number density for generation i, $n_i(a, x, j|t)$, is introduced. Number densities are closely related to probabilty densities in that they only attain values in $\mathbb{R}_+$. Different from probabilty densities the number density describes the concentration of particles instead of probabilities. Therefore, the integral of a number density over a certain interval yields the number of cells in this interval. Hence, the number of cells of cell type j and in generation i at a time t that exhibit an age $a \in [a_1, a_2]$ and a label concentration $x \in [x_1, x_2]$ is calculated by integrating over a and x, $\int_{a_1}^{a_2} \int_{x_1}^{x_2} n_i(a, x, j|t) dx da$.
The model we are going to introduce accounts for

- the change of the number density due to aging of cells, $\frac{\partial n_i(a,x,j|t)}{\partial a}$,
- the change of the number density with respect to time, $\frac{\partial n_i(a,x,j|t)}{\partial t}$,
- the change of the number density due to label degradation with dilution rate $\nu(t,x)$, $\frac{\partial(\nu(t,x)n_i(a,x,j|t))}{\partial x}$,
- the change of the number of cells in the subpopulation with type j and division number i due to cell death with rate $\beta_i^j(t,a)$ and cell division with rate $\alpha_i^j(t,a)$, $-(\alpha_i^j(t,a)+\beta_i^j(t,a))n_i\,(a,x,j|t)$,
- the change of the number of cells in the subpopulation with division number i and cell type j due to the division of cells with division number $i-1$ and cell type $\tilde{j}$ with division rate $\alpha_{i-1}^{\tilde{j}}(t,a)$ with simultaneous change of cell type from type $\tilde{j}$ to j with probability $\omega_{i-1}^{j\tilde{j}}(t,a)$,
  $2\gamma \int\limits_{\mathbb{R}_+} \alpha_{i-1}^{\tilde{j}}(t,a)\omega_{i-1}^{j\tilde{j}}(t,a)n_{i-1}\left(a,\gamma x,\tilde{j}|t\right)da$, and
- the change of the number of cells in the subpopulation with division number i and cell type j due to spontaneous change of cell type from type $\tilde{j}$ to j with transition rate $\delta_i^{j\tilde{j}}(t,a)$,
  $\delta_i^{j\tilde{j}}\,(t,a)\,n_i\left(a,x,\tilde{j}|t\right)$.

The model includes the following rates for the description of population dynamics like cell division, cell death or change of cell type

- The division rate of cells with division number i cell type j is denoted by $\alpha_i^j(t,a):\mathbb{R}_+\times\mathbb{R}_+\to\mathbb{R}_+$.
- The rate at which cells in generation i with type j die is denoted by $\beta_i^j(t,a):\mathbb{R}_+\times\mathbb{R}_+\to\mathbb{R}_+$.
- For a cell with division number i and cell type the transition rate from type $\tilde{j}$ to type j for spontaneous change of cell type $\tilde{j}$ is denoted by $\delta_i^{j\tilde{j}}(t,a):\mathbb{R}_+\times\mathbb{R}_+\to\mathbb{R}_+$.
- The transition probability for a cell in generation i from type $\tilde{j}$ to type j during cell division is denoted by $\omega_i^{j\tilde{j}}(t,a):\mathbb{R}_+\times\mathbb{R}_+\to[0,1]$.

All these rates can depend on the time $t\in\mathbb{R}_+$ and the age of the cell $a\in\mathbb{R}_+$. A label dependence for these rates is not considered as the amount of CFSE used is not harmful for the cell and its functions. Regarding the rate for spontaneous change of cell type, the transition rate for $\tilde{j}=j$ is set to $\delta_i^{jj}=0$ for all cell types j. For the cell type change during cell division it holds that

$$\sum_{\tilde{j}=1}^{J}\omega_i^{j\tilde{j}}(t,a)=1 \tag{2.1.1}$$

as $\omega_i^{j\tilde{j}}(t,a)$ is a probability.
Additionally, the model includes a rate that describes the dilution of label due to degradation processes in the cell.

- The rate $\nu(t,x) : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ models the label dilution due to cellular processes.

These cellular processes are assumed to depend only on time t and amount of label x. To ensure existence and uniqueness of $n_i(a,x,j|t)$ it is assumed that all rates are at least C^0. The combination of the aforementioned the processes yields a set of PDEs

$$\begin{aligned}&\frac{\partial n_i(a,x,j|t)}{\partial t} + \frac{\partial n_i(a,x,j|t)}{\partial a} + \frac{\partial(\nu(t,x)n_i(a,x,j|t))}{\partial x} = \\ &-(\alpha_i^j(t,a)+\beta_i^j(t,a))n_i(a,x,j|t) - \sum_{\tilde{j}=1}^{J}\delta_i^{\tilde{j}j}(t,a)n_i(a,x,j|t) + \sum_{\tilde{j}=1}^{J}\delta_i^{j\tilde{j}}(t,a)n_i\left(a,x,\tilde{j}|t\right)\end{aligned} \tag{2.1.2}$$

with initial conditions

$$\begin{aligned}i=0: &\qquad n_0(a,x,j|0) = n_0(a,j)p_0(x)\\ i\geq 1: &\qquad n_i(a,x,j|0)\equiv 0\end{aligned} \tag{2.1.3}$$

and boundary conditions

$$\begin{aligned}i=0: &\qquad n_0(0,x,j|t)\equiv 0\\ i\geq 1: &\qquad n_i(0,x,j|t) = 2\gamma\sum_{\tilde{j}=1}^{J}\int_{\mathbb{R}_+}\alpha_{i-1}^{\tilde{j}}(t,a)\omega_{i-1}^{j\tilde{j}}(t,a)n_{i-1}\left(a,\gamma x,\tilde{j}|t\right)da.\end{aligned} \tag{2.1.4}$$

The initial age distribution of cells with cell type j that did not undergo any cell division is denoted by $n_0(a,j)$. Since cell division can only be observed after staining, the number of cells with division numbers $i \geq 1$ is assumed to be zero. The function $p_0(x)$ describes the initial label distribution by giving the probability of a cell to have label concentration x after staining. The product form $n_0(a,x,j|0) = n_0(a,j)p_0(x)$ is assumed due to the independence of label distribution and cell number density in the beginning. This is required for the decomposition of the model presented in Section 2.2.
The fluorescence distribution of the whole population at time t, which in the end is the quantity measured, can be calculated from $n_i(a,x,j|t)$ by summing over the division numbers i and the cell types j and integrating out the cell age a.
To ensure notational simplicity, we will in the following exploit that $n_i(a,x,j|t)$ can be interpreted as the j-th entry of an vector-function

$$\mathbf{n_i}(a,x|t) = \left(n_i\left(a,x,1|t\right), n_i\left(a,x,2|t\right), \ldots, n_i\left(a,x,J|t\right)\right)^T.$$

With this formulation summation becomes a matrix-vector-multiplication. The dynamics are reformulated to

$$\frac{\partial \mathbf{n_i}(a,x|t)}{\partial t} + \frac{\partial \mathbf{n_i}(a,x|t)}{\partial a} = -\left(\boldsymbol{\alpha_i}(t,a) + \boldsymbol{\beta_i}(t,a) - \boldsymbol{\delta_i}(t,a)\right)\mathbf{n_i}(a,x|t) \tag{2.1.5}$$

where $\boldsymbol{\alpha_i}(t,a)$ and $\boldsymbol{\beta_i}(t,a)$ are diagonal matrices with entries $\alpha_i^j(t,a)$ and $\beta_i^j(t,a)$ on the diagonal,

$$\boldsymbol{\alpha_i}(t,a) = \begin{pmatrix}\alpha_i^1(t,a) & 0 & \ldots & 0\\ 0 & \alpha_i^2(t,a) & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \alpha_i^J(t,a)\end{pmatrix} \tag{2.1.6}$$

$$\boldsymbol{\beta_i}(t,a) = \begin{pmatrix} \beta_i^1(t,a) & 0 & \dots & 0 \\ 0 & \beta_i^2(t,a) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_i^J(t,a) \end{pmatrix} \tag{2.1.7}$$

and $\boldsymbol{\delta_i}(t,a)$ is the matrix consisting of $\delta_i^{j\tilde{j}}(t,a)$,

$$\boldsymbol{\delta_i}(t,a) = \begin{pmatrix} -\sum_{\tilde{j}=1}^{J} \delta_i^{\tilde{j}1}(t,a) & \delta_i^{12}(t,a) & \dots & \delta_i^{1J}(t,a) \\ \delta_i^{21}(t,a) & -\sum_{\tilde{j}=1}^{J} \delta_i^{\tilde{j}2}(t,a) & \dots & \delta_i^{2J}(t,a) \\ \vdots & \vdots & \ddots & \vdots \\ \delta_i^{J1}(t,a) & \delta_i^{J2}(t,a) & \dots & -\sum_{\tilde{j}=1}^{J} \delta_i^{\tilde{j}J}(t,a) \end{pmatrix} \tag{2.1.8}$$

In the following $\mathbf{0}$ and $\mathbf{1}$ denote the vector or matrix consisting of zeros and ones respectively in the appropriate dimension.
The initial conditions are reformulated to

$$\begin{aligned} i = 0: &\qquad \mathbf{n_0}(a,x|0) = \mathbf{n_0}(a)p_0(x) \\ i \geq 1: &\qquad \mathbf{n_i}(a,x|0) \equiv \mathbf{0}. \end{aligned} \tag{2.1.9}$$

As the boundary conditions consist of integrals, a notation for a component-wise matrix-/vector-integral is needed. Hence, in the following an integral applied on a matrix/vector denotes component-wise integration:

$$\int_{t_1}^{t_2} F\left(\tilde{t}\right) d\tilde{t} = \begin{pmatrix} \int_{t_1}^{t_2} f_{11}\left(\tilde{t}\right) d\tilde{t} & \int_{t_1}^{t_2} f_{12}\left(\tilde{t}\right) d\tilde{t} & \dots & \int_{t_1}^{t_2} f_{1n}\left(\tilde{t}\right) d\tilde{t} \\ \int_{t_1}^{t_2} f_{21}\left(\tilde{t}\right) d\tilde{t} & \int_{t_1}^{t_2} f_{22}\left(\tilde{t}\right) d\tilde{t} & \dots & \int_{t_1}^{t_2} f_{2n}\left(\tilde{t}\right) d\tilde{t} \\ \vdots & \vdots & \ddots & \vdots \\ \int_{t_1}^{t_2} f_{m1}\left(\tilde{t}\right) d\tilde{t} & \int_{t_1}^{t_2} f_{m2}\left(\tilde{t}\right) d\tilde{t} & \dots & \int_{t_1}^{t_2} f_{mn}\left(\tilde{t}\right) d\tilde{t} \end{pmatrix} \tag{2.1.10}$$

where $F : \mathbb{R} \to \mathbb{R}^{m\times n}$ maps t to a $(m \times n)$-matrix.
Accordingly, the boundary conditions are reformulated to

$$\begin{aligned} i = 0: &\qquad \mathbf{n_0}(0,x|t) \equiv \mathbf{0} \\ i \geq 1: &\qquad \mathbf{n_i}(0,x|t) = 2\gamma \int_{\mathbb{R}_+} \boldsymbol{\omega_{i-1}}(t,a)\boldsymbol{\alpha_{i-1}}(t,a)\mathbf{n_{i-1}}(a,\gamma x|t)da, \end{aligned} \tag{2.1.11}$$

with $\boldsymbol{\omega_i}$ being the matrix of the rates $\omega^{j\tilde{j}}$ which can be written as

$$\boldsymbol{\omega_i}(t,a) = \begin{pmatrix} 1-\sum_{\tilde{j}=1}^{J} \omega_i^{\tilde{j}1}(t,a) & \omega_i^{12}(t,a) & \dots & \omega_i^{1J}(t,a) \\ \omega_i^{21}(t,a) & 1-\sum_{\tilde{j}=1}^{J} \omega_i^{\tilde{j}2}(t,a) & \dots & \omega_i^{2J}(t,a) \\ \vdots & \vdots & \ddots & \vdots \\ \omega_i^{J1}(t,a) & \omega_i^{J2}(t,a) & \dots & 1-\sum_{\tilde{j}=1}^{J} \omega_i^{\tilde{j}1}(t,a) \end{pmatrix} \tag{2.1.12}$$

by plugging in Property (2.1.1). The reformulation of the ALDC model will allow for simple reformulation and the analysis of solution.

2.2 Analysis of ALDC model

The ALDC model is a coupled system of PDEs with initial and boundary conditions. To simplify the analysis of the ALDC model, we will decompose the PDE system in two parts, one part describing the size of the subpopulations and their age structure, and one describing the label distribution. These easier systems can then be solved, e.g. by the method of characteristics.

2.2.1 Decomposition

In this section the ALDC model will be decomposed into two simpler PDE system. Those new systems correspond to the dynamics of the distribution of the label on the one hand and the dynamics of the number densities of cells with a certain age, cell type and division number on the other. This decomposition is only possible, if certain assumptions on the dilution rate of CFSE in the cell due to cellular processes are fulfilled. As mentionend in Section 1.2 the dilution of CFSE is caused by the degradation of the CFSE-protein complex.

Theorem 2.2.1. *If the dilution rate has the form $\nu(t,x) = -k(t)x$ and the initial label distribution $p_0(x)$ is independent of the cell type j, the number density $n_i(a,x,j|t)$ can be decomposed into two independent parts: The number density of cells with a certain age a, division number i and cell type j at time t, $n_i(a,j|t)$, and the distribution of the label x for cells with division number i at time t, $p(x|i,t)$.*

$$n_i(a,x,j|t) = n_i(a,j|t)p(x|i,t) \tag{2.2.1}$$

where $n_i(a,j|t)$ solves

$$\begin{aligned}&\frac{\partial n_i(a,j|t)}{\partial t} + \frac{\partial n_i(a,j|t)}{\partial a} = \\ &-(\alpha_i^j(t,a) + \beta_i^j(t,a))n_i(a,j|t) - \sum_{\tilde{j}=1}^{J} \delta_i^{\tilde{j}j}(t,a)n_i(a,j|t) + \sum_{\tilde{j}=1}^{J} \delta_i^{j\tilde{j}}(t,a)n_i\left(a,\tilde{j}|t\right)\end{aligned} \tag{2.2.2}$$

with initial conditions

$$\begin{aligned} i = 0: &\qquad n_0(a,j|0) = n_0(a,j) \\ i \geq 1: &\qquad n_i(a,j|0) \equiv 0 \end{aligned} \tag{2.2.3}$$

and boundary conditions

$$\begin{aligned} i = 0: &\qquad n_0(0,j|t) \equiv 0 \\ i \geq 1: &\qquad n_i(0,j|t) = 2\sum_{\tilde{j}=1}^{J} \int_{\mathbb{R}_+} \alpha_{i-1}^{\tilde{j}}(t,a)\omega_{i-1}^{j\tilde{j}}(t,a)n_{i-1}\left(a,\tilde{j}|t\right) da. \end{aligned} \tag{2.2.4}$$

$p(x|i,t)$ has to fulfill

$$\frac{\partial p(x|i,t)}{\partial t} + \frac{\partial \left(\nu(x,t)p(x|i,t)\right)}{\partial x} = 0 \tag{2.2.5}$$

with initial conditions

$$p(x|i,0) = \gamma^i p_0(\gamma^i x). \tag{2.2.6}$$

For the corresponding matrix-vector-notation for $n_i(a,j|t)$ we consider the vector $\mathbf{n_i}(a|t)$,

$$\mathbf{n_i}(a|t) = \left(n_i(a,1|t),\ldots,n_i(a,J|t)\right)^T. \tag{2.2.7}$$

The decompotition then amounts to

$$\mathbf{n_i}(a,x|t) = \mathbf{n_i}(a|t)p(x|i,t) \tag{2.2.8}$$

with $p(x|i,t)$ a scalar function following the dynamics (2.2.5) and with initial condition (2.2.6) as above.
The number of cells $\mathbf{n_i}(a|t)$ follows the dynamics

$$\frac{\partial \mathbf{n_i}(a|t)}{\partial t} + \frac{\partial \mathbf{n_i}(a|t)}{\partial a} = -(\boldsymbol{\alpha_i}(t,a) + \boldsymbol{\beta_i}(t,a) - \boldsymbol{\delta_i}(t,a))\mathbf{n_i}(a|t) \tag{2.2.9}$$

with $\boldsymbol{\alpha_i}$, $\boldsymbol{\beta_i}$ and $\boldsymbol{\delta_i}$ defined as in (2.1.6), (2.1.7) and (2.1.8) respectively. The initial conditions are

$$\begin{aligned} i = 0: &\qquad \mathbf{n_0}(a|0) = \mathbf{n_0}(a) \\ i \geq 1: &\qquad \mathbf{n_i}(a|0) \equiv \mathbf{0} \end{aligned} \tag{2.2.10}$$

and the boundary conditions become

$$\begin{aligned} i = 0: &\qquad \mathbf{n_0}(0|t) \equiv \mathbf{0} \\ i \geq 1: &\qquad \mathbf{n_i}(0|t) = 2\int_{\mathbb{R}_+} \boldsymbol{\omega_{i-1}}(t,a)\boldsymbol{\alpha_{i-1}}(t,a)\mathbf{n_{i-1}}(a|t)da =: 2\boldsymbol{\psi_{i-1}}(t) \end{aligned} \tag{2.2.11}$$

with $\boldsymbol{\omega_{i-1}}$ as before.

Proof: To prove Theorem 2.2.1 it is shown, that plugging in the ansatz (2.2.8) on both sides of (2.1.5) yields the same dynamics as the original equation and that the initial and boundary conditions are consistent, similarly to the proof done in [5]. Inserting the ansatz (2.2.8) on the left hand side of (2.1.5) we obtain

$$\begin{aligned} &\frac{\partial \mathbf{n_i}(a,x|t)}{\partial t} + \frac{\partial \mathbf{n_i}(a,x|t)}{\partial a} + \frac{\partial(\nu(t,x)\mathbf{n_i}(a,x|t))}{\partial x} \\ &\quad = \frac{\partial(\mathbf{n_i}(a|t)p(x|i,t))}{\partial t} + \frac{\partial(\mathbf{n_i}(a|t)p(x|i,t))}{\partial a} + \frac{\partial(\nu(t,x)(\mathbf{n_i}(a|t)p(x|i,t)))}{\partial x} \\ &\quad = \frac{\partial \mathbf{n_i}(a|t)}{\partial t}p(x|i,t) + \frac{\partial p(x|i,t)}{\partial t}\mathbf{n_i}(a|t) + \frac{\partial \mathbf{n_i}(a|t)}{\partial a}p(x|i,t) + \frac{\partial(\nu(t,x)p(x|i,t))}{\partial x}\mathbf{n_i}(a|t) \\ &\quad = \left(\frac{\partial \mathbf{n_i}(a|t)}{\partial t} + \frac{\partial \mathbf{n_i}(a|t)}{\partial a}\right)p(x|i,t) + \left(\frac{\partial p(x|i,t)}{\partial t} + \frac{\partial(\nu(t,x)p(x|i,t))}{\partial x}\right)\mathbf{n_i}(a|t) \\ &\quad = -\left(\boldsymbol{\alpha_i}(t,a) + \boldsymbol{\beta_i}(t,a) - \boldsymbol{\delta_i}(t,a)\mathbf{n_i}(a|t)\right)p(x|i,t). \end{aligned}$$

From line two to three we used the product rule for differentiation applied to multiplication of a vector with a scalar. In the next step $p(x|i,t)$ and its partial derivatives can be factorized since they are scalar functions. In the last step we inserted the dynamics of the decomposed model (2.2.9) and (2.2.5). Inserting the ansatz (2.2.8) on the right hand side of (2.1.5) yields

$$\begin{aligned} &-\left(\boldsymbol{\alpha_i}(t,a)+\boldsymbol{\beta_i}(t,a)-\boldsymbol{\delta_i}(t,a)\right)\mathbf{n_i}(a,x|t) \\ &\quad=-\left(\boldsymbol{\alpha_i}(t,a)+\boldsymbol{\beta_i}(t,a)-\boldsymbol{\delta_i}(t,a)\right)\mathbf{n_i}(a|t)p(x|i,t). \end{aligned}$$

Hence, inserting the ansatz on the left hand side yields the same equation as inserting it on the right hand side. The ansatz thus fulfills the PDE.
To check the initial conditions for consistency we distinguish the cases $i=0$ and $i\geq 1$. For $i\geq 1$ inserting the ansatz and the original equation both yield

$$\begin{aligned} &\mathbf{n_i}(a,x|0)\equiv\mathbf{0} \\ &\mathbf{n_i}(a|0)p(x|i,0)=\mathbf{0}\gamma^i p_0(\gamma^i x)\equiv\mathbf{0}. \end{aligned}$$

For $i=0$ inserting the ansatz on the left hand side of (2.1.9) yields

$$\mathbf{n_0}(a,x|0)=\mathbf{n_0}(a|0)p_0(x|0,0)=\mathbf{n_0}(a)p_0(x)$$

which is the same as the right hand side of (2.1.9).
Inserting the ansatz in the boundary condition (2.1.11) for $i=0$ yields zero on both sides. For $i\geq 1$ inserting the ansatz on the right hand side of the equation (2.1.11) yields

$$\begin{aligned} \mathbf{n_i}(0,x|t)&=2\gamma\int_{\mathbb{R}_+}\boldsymbol{\omega_{i-1}}(t,a)\boldsymbol{\alpha_{i-1}}(t,a)\mathbf{n_{i-1}}(a,\gamma x|t)da \\ &=2\gamma\int_{\mathbb{R}_+}\boldsymbol{\omega_i}(t,a)\boldsymbol{\alpha_{i-1}}(t,a)\mathbf{n_{i-1}}(a|t)p(\gamma x|i-1.t)da. \end{aligned}$$

Inserting it on the left hand side of (2.1.11) we obtain

$$\mathbf{n_i}(0|t)p(x|i,t)=\left(2\int_{\mathbb{R}_+}\boldsymbol{\omega_{i-1}}(t,a)\boldsymbol{\alpha_{i-1}}(t,a)\mathbf{n_{i-1}}(a|t)da\right)p(x|i,t).$$

Both sides are equal iff

$$p(x|i,t)=\gamma p(\gamma x|i-1,t).$$

For $\nu(x,t)=k(t)x$ the solution of (2.2.5) with (2.2.6) fulfills this. This can be proven by looking at the solution derived in (2.2.2). See also [5]. Hence, the boundary conitions are consistent. This concludes the proof. □

As the initial label distribution $p_0(x)$ is assumed to be independent of the cell type j, $p(x|i,t)$ is also independent of the cell type j. This is biologically reasonable. To generalize this assumption one could later use the superposition principle to construct solutions for cell type-dependent initial label distributions $p_0(x,j)$.
With the decomposition, the initial problem of solving the coupled PDE system (2.1.5) has been transformed into two easier problems, namely solving the PDE (2.2.5) and PDE system (2.2.9). This can be done by the method of characteristics.

2.2.2 Solution of the ALDC model

After decomposing the PDE system (2.1.5) the solution for the decomposed systems is presented. We then proceed to show that the solutions indeed fulfill the PDE system.

Theorem 2.2.2 ([2,5]). *The solution for* (2.2.5) *and* (2.2.6) *is*

$$p(x|i,t) = \gamma^i exp\left(\int_0^t k(\tilde{t})d\tilde{t}\right) p_0\left(\gamma^i \exp\left(\int_0^t k(\tilde{t})d\tilde{t}\right) x\right) \tag{2.2.12}$$

Theorem 2.2.3. *In the case of* $\boldsymbol{\delta_i}(t,a) = \mathbf{0}$ *for all* $i \in \mathbb{N}_0$, $a,t \in \mathbb{R}_+$ *the solution of* (2.2.9), (2.2.10) *and* (2.2.11) *is*

$$\begin{aligned}
\mathbf{n_0}(a|t) &= \begin{cases} \mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right) & ,t \le a \\ 0 & ,t > a \end{cases} \\
\mathbf{n_i}(a|t) &= \begin{cases} 2\boldsymbol{\psi_{i-1}}(t-a)\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right) & ,t > a \\ 0 & ,t \le a \end{cases}
\end{aligned} \tag{2.2.13}$$

with $\boldsymbol{\psi_i}(t-a) = \int_{\mathbb{R}_+} \boldsymbol{\omega_i}\left(t-a,\tilde{a}\right)\boldsymbol{\alpha_i}\left(t-a,\tilde{a}\right)\mathbf{n_i}(\tilde{a}|t-a)d\tilde{a}$.

The scalar form of (2.2.13) is

$$\begin{aligned}
n_0(a,j|t) &= \begin{cases} n_0(a-t,j)\exp\left(-\int_0^t \alpha_0^j\left(\tilde{t},\tilde{t}+a-t\right)+\beta_0^j\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right) & ,t \le a \\ 0 & ,t > a \end{cases} \\
n_i(a,j|t) &= \begin{cases} 2\sum_{\tilde{j}=1}^{J}\int_{\mathbb{R}_+}\alpha_{i-1}^{\tilde{j}}\left(t-a,\tilde{a}\right)\omega_{i-1}^{j\tilde{j}}\left(t-a,\tilde{a}\right)n(\tilde{a},i-1,\tilde{j}|t-a)d\tilde{a} & \\ \qquad \exp\left(-\int_0^a \alpha_i^j\left(\tilde{a}+t-a,\tilde{a}\right)+\beta_i^j\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right) & ,t > a\,. \\ 0 & ,t \le a \end{cases}
\end{aligned} \tag{2.2.14}$$

Since $\boldsymbol{\alpha_i}$ and $\boldsymbol{\beta_i}$ are diagonal matrices their exponential is again diagonal. This holds also for the exponential of the component-wise integral

$$\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)$$

and

$$\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right).$$

In other words the exponential function acts component-wise on diagonal matrices. Furthermore, partial derivatives in direction of a scalar t of vector- and matrix-functions $\frac{\partial F(t)}{\partial t}$ can also be considered as acting component-wise. Hence, in this case with diagonal matrices as above all operations in consideration (i.e. integration, differentiation and the exponential function) can be considered to work only on the matrix components. Therefore, the matrices can in this case be treated like scalars and all computational rules for real scalar analysis can be applied.

Proof: We show in the following that $\mathbf{n_i}(a|t)$ as defined in (2.2.13) solves the PDE system (2.2.9). We first consider the partial derivation in time direction for $i = 0$. We obtain

$$\begin{aligned}\frac{\partial \mathbf{n_0}(a|t)}{\partial t} =& \frac{\partial}{\partial t}\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)\\ &+\mathbf{n_0}(a-t)\underbrace{\frac{\partial}{\partial t}\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)}_{(*)}\end{aligned} \tag{2.2.15}$$

by applying the product rule. To make the calculation more structured the derivation of the second factor $(*)$ is considered on its own. To differentiate the chain-rule and the Leibnitz-rule for differentiation of integrals with parameter-dependent limits are applied

$$\begin{aligned}(*) =& \exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)\\ &\left(-\left(\boldsymbol{\alpha_0}\left(t,a\right)+\boldsymbol{\beta_0}\left(t,a\right)\right)-\int_0^t \frac{\partial}{\partial t}\boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\frac{\partial}{\partial t}\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right).\end{aligned}$$

Next we plug $(*)$ back into (2.2.15) and use (2.2.13) to replace $\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)$ with $\mathbf{n_0}(a|t)$ in the second summand. Hence, the time derivative is

$$\begin{aligned}\frac{\partial \mathbf{n_0}(a|t)}{\partial t} =& \frac{\partial}{\partial t}\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)-\mathbf{n_0}(a|t)\\ &\left(\boldsymbol{\alpha_0}\left(t,a\right)+\boldsymbol{\beta_0}\left(t,a\right)+\int_0^t \frac{\partial}{\partial t}\boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\frac{\partial}{\partial t}\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right).\end{aligned} \tag{2.2.16}$$

The derivation in age direction is again calculated by applying product- and chain-rule

$$\begin{aligned}\frac{\partial \mathbf{n_0}(a|t)}{\partial a} =& \frac{\partial}{\partial a}\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)\\ &+\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)\\ &\left(-\frac{\partial}{\partial a}\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right).\end{aligned}$$

Assuming that $\boldsymbol{\alpha_0}(t,a)$ and $\boldsymbol{\beta_0}(t,a)$ are continuously differentiable with respect to a, Theorem 10 in [9, p. 352] can be used to interchange the integral and the differentiation. Again (2.2.13) can be plugged in yielding

$$\begin{aligned}\frac{\partial \mathbf{n_0}(a|t)}{\partial a} =& \frac{\partial}{\partial a}\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)\\ &-\mathbf{n_0}(a|t)\int_0^t \frac{\partial}{\partial a}\boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\frac{\partial}{\partial a}\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}.\end{aligned} \tag{2.2.17}$$

Taking a closer look at the remaining derivatives in (2.2.16) and (2.2.17), one finds that

$$\begin{aligned}\int_0^t \frac{\partial}{\partial t}\boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\frac{\partial}{\partial t}\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\\ =-\int_0^t \frac{\partial}{\partial a}\boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\frac{\partial}{\partial a}\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\end{aligned}$$

as well as

$$\frac{\partial}{\partial a}\mathbf{n_0}(a-t)=-\frac{\partial}{\partial t}\mathbf{n_0}(a-t)$$

because of the inverted sign of a and t in these expressions. Therefore, it follows that

$$\begin{aligned}&\frac{\partial}{\partial t}\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)\\ &+\frac{\partial}{\partial a}\mathbf{n_0}(a-t)\exp\left(-\int_0^t \boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\right)=0.\end{aligned}$$

Also

$$\begin{aligned}&\mathbf{n_0}(a|t)\int_0^t \frac{\partial}{\partial t}\boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\frac{\partial}{\partial t}\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\\ &+\mathbf{n_0}(a|t)\int_0^t \frac{\partial}{\partial a}\boldsymbol{\alpha_0}\left(\tilde{t},\tilde{t}+a-t\right)+\frac{\partial}{\partial a}\boldsymbol{\beta_0}\left(\tilde{t},\tilde{t}+a-t\right)d\tilde{t}\end{aligned}$$

cancels out. Hence, the sum of (2.2.16) and (2.2.17) becomes

$$\begin{aligned}\frac{\partial\mathbf{n_0}(a|t)}{\partial t}+\frac{\partial\mathbf{n_0}(a|t)}{\partial a}&=-\mathbf{n_0}(a|t)\left(\boldsymbol{\alpha_0}(t,a)+\boldsymbol{\beta_0}(t,a)\right)\\ &=-\left(\boldsymbol{\alpha_0}(t,a)+\boldsymbol{\beta_0}(t,a)\right)\mathbf{n_0}(a|t)\end{aligned}$$

and $\mathbf{n_0}(a|t)$ fulfills (2.1.5). The order of the factors can be changed because diagonal matrices are commutative. The initial condition (2.1.9) holds as well

$$\mathbf{n_0}(a|0)=e^{\mathbf{0}}\mathbf{n_0}(a)=\mathbf{n_0}(a).$$

In the case $a=0$ and $t>0$ it holds that $t>a$, hence, $\mathbf{n_0}(0|t)$ is equal to zero and thus fulfills the boundary condition for $i=0$.
The calculations for the case $i\geq 1$ proceed analogously with the roles of t and a interchanged. Again the derivation in age and time is considered separately and as before the matrices are treated like scalars. The derivation with respect to time is calculated by using the product- and chain-rule

$$\begin{aligned}\frac{\partial\mathbf{n_i}(a|t)}{\partial t}=&2\frac{\partial}{\partial t}\boldsymbol{\psi_{i-1}}(t-a)\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right)\\ &+2\boldsymbol{\psi_{i-1}}(t-a)\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right)\\ &\left(-\frac{\partial}{\partial t}\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right).\end{aligned}$$

Again, Theorem 10 in [9, p. 352] is utilized to interchange the integral and the differentiation with the assumption that $\boldsymbol{\alpha_i}(t,a)$ and $\boldsymbol{\beta_i}(t,a)$ are continuously differentiable with respect to t. Furthermore, (2.2.13) is plugged in for $\mathbf{n_i}(a|t)$

$$\begin{aligned}\frac{\partial \mathbf{n_i}(a|t)}{\partial t} =& 2\frac{\partial}{\partial t}\boldsymbol{\psi_{i-1}}(t-a)\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right)\\ &-\mathbf{n_i}(a|t)\int_0^a \frac{\partial}{\partial t}\boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\frac{\partial}{\partial t}\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}.\end{aligned}$$

Derivation of $\mathbf{n_i}(a|t)$ with respect to age follows the same rules as derivation of $\mathbf{n_0}(a|t)$ with respect to time. The result is the same with interchanged roles of t and a

$$\begin{aligned}\frac{\partial \mathbf{n_i}(a|t)}{\partial a} =& 2\frac{\partial}{\partial a}\boldsymbol{\psi_{i-1}}(t-a)\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right)\\ &-2\boldsymbol{\psi_{i-1}}(t-a)\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right)\\ &\left(\boldsymbol{\alpha_i}\left(t,a\right)+\boldsymbol{\beta_i}\left(t,a\right)+\int_0^a \frac{\partial}{\partial a}\boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\frac{\partial}{\partial a}\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right)\\ =& 2\frac{\partial}{\partial a}\boldsymbol{\psi_{i-1}}(t-a)\exp\left(-\int_0^a \boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right)-\mathbf{n_i}(a|t)\\ &\left(\boldsymbol{\alpha_i}\left(t,a\right)+\boldsymbol{\beta_i}\left(t,a\right)+\int_0^a \frac{\partial}{\partial a}\boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\frac{\partial}{\partial a}\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\right).\end{aligned}$$

Again the inverted sign of t and a in the expressions is used to conclude that

$$\begin{aligned}&\int \frac{\partial}{\partial t}\boldsymbol{\alpha}_i\left(\tilde{a}+t-a,\tilde{a}\right)+\frac{\partial}{\partial t}\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\\ &\quad=-\int \frac{\partial}{\partial a}\boldsymbol{\alpha_i}\left(\tilde{a}+t-a,\tilde{a}\right)+\frac{\partial}{\partial a}\boldsymbol{\beta_i}\left(\tilde{a}+t-a,\tilde{a}\right)d\tilde{a}\end{aligned}$$

and

$$\frac{\partial}{\partial t}\boldsymbol{\psi_{i-1}}(t-a)=-\frac{\partial}{\partial a}\boldsymbol{\psi_{i-1}}(t-a).$$

As before these terms cancel out when the derivations in time- and age direction are added. Hence, we obtain

$$\begin{aligned}\frac{\partial \mathbf{n_i}(a|t)}{\partial t}+\frac{\partial \mathbf{n_i}(a|t)}{\partial a} &= -\mathbf{n_i}(a|t)\left(\boldsymbol{\alpha_i}\left(t,a\right)+\boldsymbol{\beta_i}\left(t,a\right)\right)\\ &= -\left(\boldsymbol{\alpha_i}\left(t,a\right)+\boldsymbol{\beta_i}\left(t,a\right)\right)\mathbf{n_i}(a|t).\end{aligned}$$

This proves that the solution (2.2.13) for $\mathbf{n_i}(a|t)$ fulfills (2.1.5). Since the $\mathbf{n_i}(a|t)$ is equal to zero if $t \leq a$, which is the case for $t = 0$, the initial condition (2.1.9) for $\mathbf{n_i}(a|0)$ is fulfilled, as well. The boundary conditions are also fulfilled

$$\mathbf{n_i}(0|t)=e^{\mathbf{0}}\cdot 2\int_{\mathbb{R}_+}\boldsymbol{\omega_{i-1}}\left(t,\tilde{a}\right)\boldsymbol{\alpha_{i-1}}\left(t,\tilde{a}\right)\mathbf{n_{i-1}}(\tilde{a}|t)d\tilde{a}.$$

This concludes the proof. □

The analytical solution presented in (2.2.13) and (2.2.12) can be used to implement a numerical scheme to simulate the model. An example for these simulations is given in Section 2.4. However, the solution can only be computed iteratively for each i since $\mathbf{n_i}$ depends on $\mathbf{n_{i-1}}$. Furthermore, the solution for all cell types in one generation has to be known to compute the solution for the following generation.
For $\boldsymbol{\delta_i}(t,a) \neq \mathbf{0}$ there exists, in general, no closed form for solving (2.2.9). Spontaneous change of cell type however is biologically rather rare. For the case that spontaneous change of cell type has to be considered, numerical approximations to the solution of (2.2.9) can be employed.

2.3 Relation of the ALDC model to existing models

As the ALDC model consideres age-, label-, division number- and cell type-structured populations a obvious question is, how it relates to the existing models which consider a subset of the cellular properties. We will focus on the relation to the division number-, label- and age-structured population (DLASP) model [5] and the division number-, cell type- and label- strucutured population (DCLSP) model [8] discussed in Section 1.3, as the division number- and label-structured population (DLSP) model, the label-structured population model (LSP), the division number-structured population (DSP) model and the age-structured population (ASP) model mentioned in Section 1.3 are each special cases of the DCLSP model or DALSP model.

2.3.1 DALSP model

The division number-, age- and label-structured population (DALSP) model is governed by the following evolution equations, see [5]. The notation is adapted to fit the one used in this thesis

$$\frac{\partial n_i(a,x|t)}{\partial t} + \frac{\partial n_i(a,x|t)}{\partial a} + \frac{\partial(\nu(t,x)n_i(a,x|t))}{\partial x} = -(\alpha_i(t,a) + \beta_i(t,a))n_i(a,x|t) \quad (2.3.1)$$

with initial conditions

$$\begin{aligned} i = 0: &\qquad n_0(a,x|0) = n_0(a)p_0(x) \\ i \geq 1: &\qquad n_i(a,x|0) \equiv 0 \end{aligned} \quad (2.3.2)$$

and boundary conditions

$$\begin{aligned} i = 0: &\qquad n_0(0,x,|t) \equiv 0 \\ i \geq 1: &\qquad n_i(0,x|t) = 2\gamma \int_{\mathbb{R}_+} \alpha_{i-1}(t,a)n_{i-1}(a,\gamma x|t)da. \end{aligned} \quad (2.3.3)$$

To establish the link between the DALSP model and ALDC model, we can consider the case of a population with one cell type, $J = 1$. In this case, both models are exactly identical. Alternatively, we can marginalize over the cell type j. This corresponds to summing the individual equations for the cell types j and thus only considering the age distribution of the total population in each generation i.

Theorem 2.3.1. *If the rates $\alpha_i^j(t,a) = \alpha_i(t,a)$ and $\beta_i^j(t,a) = \beta_i(t,a)$ are independent of the cell type j, marginalization of the equations (2.1.2) or rather (2.1.5) and the accompanying conditions over j, i.e. summing over the cell types j, gives the DALSP model.*

Proof: It has to be shown that marginalizing the equation (2.1.5) over j yields the evolution equations governing the DALSP. In matrix-vector-form summing over j means forming the $(\mathbb{R}^J)$-scalar product with the J-dimensional vector consisting of ones: $\mathbf{1} = (1,\dots,1)^T$. The scalar product of $\mathbf{n_i}(a,x|t)$ with this vector yields the number density considered in the DALSP model and shall be denoted by

$$\sum_{j=1}^{J} n_i(a,x,j|t) = \langle \mathbf{n_i}(a,x|t), \mathbf{1}\rangle =: n_i(a,x|t). \tag{2.3.4}$$

Because the rates $\boldsymbol{\alpha_i}$ and $\boldsymbol{\beta_i}$ are independent of j they can be viewed as scalar functions times the J-dimensional unit matrix I: $\boldsymbol{\alpha_i} = \alpha_i(t,a)\mathrm{I}$ and $\boldsymbol{\beta_i} = \beta_i(t,a)\mathrm{I}$, respectively. The variables a, x and t will be omitted in the calculations and n_i will denote the number density considered in the DALSP.
For the governing PDEs summing over j results in

$$\begin{aligned}\left\langle \frac{\partial \mathbf{n_i}}{\partial t} + \frac{\partial \mathbf{n_i}}{\partial a} + \frac{\partial(\nu \mathbf{n_i})}{\partial x}, \mathbf{1}\right\rangle &= \frac{\partial\langle \mathbf{n_i},\mathbf{1}\rangle}{\partial t} + \frac{\partial\langle \mathbf{n_i},\mathbf{1}\rangle}{\partial a} + \frac{\partial(\nu\langle \mathbf{n_i},\mathbf{1}\rangle)}{\partial x} \\ &= \frac{\partial n_i}{\partial t} + \frac{\partial n_i}{\partial a} + \frac{\partial(\nu n_i)}{\partial x}\end{aligned}$$

by using the bilinearity of the scalar product and the linearity of partial differentiation in the first and the definition of $\langle \mathbf{n_i},\mathbf{1}\rangle$ in the second step. The marginalization of the right hand side can again be transformed using bilinearity of the scalar product and (2.3.4)

$$\begin{aligned}\langle -(\alpha_i\mathrm{I} + \beta_i\mathrm{I} - \boldsymbol{\delta_i})\mathbf{n_i},\mathbf{1}\rangle &= \langle -\alpha_i\mathrm{I}\mathbf{n_i},\mathbf{1}\rangle - \langle \beta_i\mathrm{I}\mathbf{n_i},\mathbf{1}\rangle + \langle \boldsymbol{\delta_i}\mathbf{n_i},\mathbf{1}\rangle \\ &= -\alpha_i\langle \mathbf{n_i},\mathbf{1}\rangle - \beta_i\langle \mathbf{n_i},\mathbf{1}\rangle + \langle \boldsymbol{\delta_i}\mathbf{n_i},\mathbf{1}\rangle \\ &= -\alpha_i n_i - \beta_i n_i + \langle \boldsymbol{\delta_i}\mathbf{n_i},\mathbf{1}\rangle .\end{aligned}$$

With a basic result from linear algebra it holds that

$$\langle \boldsymbol{\delta_i}\mathbf{n_i},\mathbf{1}\rangle = \left\langle \mathbf{n_i}, \boldsymbol{\delta_i}^T\mathbf{1}\right\rangle . \tag{2.3.5}$$

Considering the structure of $\boldsymbol{\delta_i}$

$$\boldsymbol{\delta_i} = \begin{pmatrix} -\sum\limits_{\tilde{j}=1}^{J}\delta_i^{\tilde{j}1} & \delta_i^{12} & \dots & \delta_i^{1J} \\ \delta_i^{21} & -\sum\limits_{\tilde{j}=1}^{J}\delta_i^{\tilde{j}2} & \dots & \delta_i^{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_i^{J1} & \delta_i^{J2} & \dots & -\sum\limits_{\tilde{j}=1}^{J}\delta_i^{\tilde{j}J} \end{pmatrix}$$

one can see, that summing the element of each column of $\boldsymbol{\delta_i}$ yields zero. This is reasonable, as these columns contain the fluxes in and out of this cell type under mass conservation. Therefore, it holds that ${\boldsymbol{\delta_i}}^T\mathbf{1} = \mathbf{0}$. Hence, summing (2.1.5) over j yields

$$\frac{\partial n_i}{\partial t} + \frac{\partial n_i}{\partial a} + \frac{\partial(\nu n_i)}{\partial x} = -(\alpha_i + \beta_i)n_i.$$

This is the PDE (2.3.1) governing the DALSP. The initial conditions of the corresponding DALSP model can be derived from the new model. For $i = 0$ the left hand side is by Definition (2.3.4)

$$\langle \mathbf{n_0}(a, x|0), \mathbf{1}\rangle = n_0(a, x|0).$$

With defining $\langle n_0(a, j), \mathbf{1}\rangle =: n_0(a)$ and using bilinearity of the scalar product the right hand side becomes

$$\langle n_0(a, j)p_0(x), \mathbf{1}\rangle = n_0(a)p_0(x).$$

In the case $i \geq 1$ the outcome is trivial

$$n_i(a, x|0) = \langle \mathbf{n_i}(a, x|0), \mathbf{1}\rangle \equiv \langle 0, \mathbf{1}\rangle \equiv 0.$$

Regarding the boundary conditions the structure of $\boldsymbol{\omega_i}$ is exploited. As the columns of $\boldsymbol{\omega_i}$ contain the probability to end up in cell type $1, \ldots, J$ starting from a certain type j they have to add up to one. For $i = 0$ the scalar product with $\mathbf{1}$ is equal to

$$\langle \mathbf{n_0}(0, x|t), \mathbf{1}\rangle \equiv 0.$$

The situation for $i \geq 1$ is a little more complicated. With the definition of $n_i(0, x|t)$ (2.3.4) the left hand side is

$$\langle \mathbf{n_i}(0, x|t), \mathbf{1}\rangle = n_i(0, x|t).$$

Summation of the right hand side yields

$$\left\langle 2\gamma \int_{\mathbb{R}_+} \boldsymbol{\omega_i}(t, a)\boldsymbol{\alpha_{i-1}}(t, a)\mathbf{n_{i-1}}(a, \gamma x|t)da, \mathbf{1}\right\rangle \tag{2.3.6}$$

$$= 2\gamma \int_{\mathbb{R}_+} \langle \boldsymbol{\omega_i}(t, a)\alpha_{i-1}(t, a)\mathbf{I}\mathbf{n_{i-1}}(a, \gamma x|t), \mathbf{1}\rangle \, da. \tag{2.3.7}$$

Here the scalar product and integration have been interchanged by using the linearity of the integral, since forming the scalar product with $\mathbf{1}$ results in adding up the component-wise integrals. Again the property of the scalar product used in (2.3.5) is exploited to further simplify (2.3.7)

$$\begin{aligned} 2\gamma \int_{\mathbb{R}_+} \langle \boldsymbol{\omega_i}(t, a)\alpha_{i-1}(t, a)\mathbf{I}\mathbf{n_{i-1}}(a, \gamma x|t), \mathbf{1}\rangle \, da &= 2\gamma \int_{\mathbb{R}_+} \left\langle \alpha_{i-1}(t, a)\mathbf{n_{i-1}}(a, \gamma x|t), {\boldsymbol{\omega_i}}^T\mathbf{1}\right\rangle da \\ &= 2\gamma \int_{\mathbb{R}_+} \alpha_{i-1}(t, a) \left\langle \mathbf{n_{i-1}}(a, \gamma x|t), \mathbf{1}\right\rangle da \\ &= 2\gamma \int_{\mathbb{R}_+} \alpha_{i-1}(t, a)n_{i-1}(a, \gamma x|t)da. \end{aligned}$$

For the last two steps we used that ${\boldsymbol{\omega_i}}^T\mathbf{1} = \mathbf{1}$ and the definition of $\langle \mathbf{n_{i-1}}(a, \gamma x|t), \mathbf{1}\rangle$ (2.3.4). □

This shows that the ALDC model is a generalization of the DALSP model. All information obtained by the latter can also be gained by analyzing the ALDC model. As the DALSP model is itself a generalization of the DLSP, LSP, ASP and DSP model, these models are also special cases of the ALDC model.

2.3.2 DCLSP-model

The division number-, cell type- and label-structured population (DCLSP) model is governed by the following evolution equations, see [8]. Again the notation has been changed to fit the one used in this thesis

$$\begin{aligned}&\frac{\partial n_i(x,j|t)}{\partial t}+\frac{\partial(\nu(t,x)n_i(x,j|t))}{\partial x}\\&\quad=-(\alpha_i^j(t)+\beta_i^j(t))n_i(x,j|t)-\sum_{\tilde{j}=1}^{J}\delta_i^{\tilde{j}j}(t)n_i(x,j|t)+\sum_{\tilde{j}=1}^{J}\delta_i^{j\tilde{j}}(t)n_i(x,\tilde{j}|t)\\&\qquad+\begin{cases}0 & ,i=0\\ 2\gamma\sum_{\tilde{j}=1}^{J}\alpha_{i-1}^{\tilde{j}}(t)\omega_{i-1}^{j\tilde{j}}(t)n_{i-1}(\gamma x,\tilde{j}|t) & ,i\geq 1\end{cases}\end{aligned}\tag{2.3.8}$$

with initial conditions

$$\begin{aligned}&i=0: && n_0(x,j|0)=n_0(j)p_0(x)\\&i\geq 1: && n_i(x,j|0)=0.\end{aligned}\tag{2.3.9}$$

Analogously to the previous section, the DCLSP model can be derived from the ALDC model by marginalizing over the cell age a. By integrating over the cell age a, the age structure of the population is neglected and only the number of cells with a certain cell type j in a certain generation i is considered.

Theorem 2.3.2. *For rates $\boldsymbol{\alpha_i}(t,a)=\boldsymbol{\alpha_i}(t)$, $\boldsymbol{\beta_i}(t,a)=\boldsymbol{\beta_i}(t)$, $\boldsymbol{\delta_i}(t,a)=\boldsymbol{\delta_i}(t)$ and $\boldsymbol{\omega_i}(t,a)=\boldsymbol{\omega_i}(t)$ that are independent of a, the marginalization of the evolution equations of the ALDC model over a, i.e. integrating over a, yields the DCLSP model from [8]. The marginalization of $\mathbf{n_i}(a,x|t)$ over a is denoted by*

$$\int_0^\infty \mathbf{n_i}(a,x|t)da =: \mathbf{n_i}(x|t).\tag{2.3.10}$$

The vector-function $\mathbf{n_i}(x|t)$ consist of the functions $(n_i(x,1|t),\ldots,n_i(x,J|t))$ where $n_i(x,j|t)$ is the function that is considered in the DCLSP model formulation.

Proof: It has to be shown that marginalizing over a yields the DCLSP model. In the following it will often be used that

$$\int_{a_1}^{a_2} M(t)v(t,\tilde{a})d\tilde{a}=M(t)\int_{a_1}^{a_2} v(t,\tilde{a})d\tilde{a}\tag{2.3.11}$$

where $M(a)\in\mathbb{R}^{m\times n}$ is a matrix for all $a\in\mathbb{R}$ and $v(t,a):\mathbb{R}\times\mathbb{R}\to\mathbb{R}^n$ is a vector-function. This is true because the multiplication with $M(t)$ results in a weighted summation of the component functions of $v(t,a)$. Hence, multiplication with rates that are independent of the integration variable and integration can be interchanged.

Marginalizing of the left hand side of (2.1.5) over a yields

$$\int_0^\infty \frac{\partial \mathbf{n_i}(a,x|t)}{\partial t}da + \int_0^\infty \frac{\partial \mathbf{n_i}(a,x|t)}{\partial a}da + \int_0^\infty \frac{\partial(\nu(t,x)\mathbf{n_i}(a,x|t))}{\partial x}da.$$

With the assumption of uniform convergence of

$$\frac{\partial \mathbf{n_i}(a,x|t)}{\partial t} \text{ and } \frac{\partial(\nu(t,x)\mathbf{n_i}(a,x|t))}{\partial x}$$

the integral and the partial derivative can be interchanged by applying Theorem 14 [9, p. 358]. With the fundamental theorem of calculus applied to $\int_0^\infty \frac{\partial \mathbf{n_i}(a,x|t)}{\partial a}da$ and using the fact that $\nu(t,x)$ is independent of a we get

$$\begin{aligned}
&\int_0^\infty \frac{\partial \mathbf{n_i}(a,x|t)}{\partial t}da + \int_0^\infty \frac{\partial \mathbf{n_i}(a,x|t)}{\partial a}da + \int_0^\infty \frac{\partial(\nu(t,x)\mathbf{n_i}(a,x|t))}{\partial x}da \\
&= \frac{\partial \int_0^\infty \mathbf{n_i}(a,x|t)da}{\partial t} + \lim_{a\to\infty} \mathbf{n_i}(a,x|t) - \mathbf{n_i}(0,x|t) + \frac{\partial(\nu(t,x)\int_0^\infty \mathbf{n_i}(a,x|t)da)}{\partial x} \\
&= \frac{\partial \mathbf{n_i}(x|t)}{\partial t} + \mathbf{0} - \mathbf{n_i}(0,x|t) + \frac{\partial(\nu(t,x)\mathbf{n_i}(x|t))}{\partial x}.
\end{aligned}$$

The equality holds since all component functions of $\mathbf{n_i}$ have finite support regarding the age, hence, the limit at infinity is $\lim_{a\to\infty} \mathbf{n_i}(a,x|t) = \mathbf{0}$. Also the Definition (2.3.10) was used. When inserting the boundary condition for $\mathbf{n_i}(0,x|t)$ the cases $i = 0$ and $i \geq 1$ have to be distinguished. The case $i = 0$ is quite simple as $\mathbf{n_0}(0,x|t) \equiv \mathbf{0}$. Plugging this in the marginalization of the left hand side of (2.1.5) over a is

$$\frac{\partial \mathbf{n_i}(x|t)}{\partial t} - \mathbf{0} + \frac{\partial(\nu(t,x)\mathbf{n_i}(x|t))}{\partial x}.$$

In the case $i \geq 1$ the boundary condition can be plugged in for $\mathbf{n_i}(0,x|t)$, as well. The marginalization of the left hand side of (2.1.5) over a then becomes

$$\begin{aligned}
&\frac{\partial \mathbf{n_i}(x|t)}{\partial t} - 2\gamma \int_{\mathbb{R}_+} \boldsymbol{\omega_{i-1}}(t)\boldsymbol{\alpha_{i-1}}(t)\mathbf{n_{i-1}}(a,\gamma x|t)da + \frac{\partial(\nu(t,x)\mathbf{n_i}(x|t))}{\partial x} \\
&\quad = \frac{\partial \mathbf{n_i}(x|t)}{\partial t} - 2\gamma\boldsymbol{\omega_{i-1}}(t)\boldsymbol{\alpha_{i-1}}(t)\mathbf{n_{i-1}}(\gamma x|t) + \frac{\partial(\nu(t,x)\mathbf{n_i}(x|t))}{\partial x}.
\end{aligned}$$

Here it was used that the rates ω_{i-1} and α_{i-1} are independent of a and we plugged in the defintion (2.3.10). In both cases $i = 0$ and $i \geq 1$ it holds for the right hand side that

$$\begin{aligned}
\int_0^\infty -(\boldsymbol{\alpha_i}(t) + \boldsymbol{\beta_i}(t) - \boldsymbol{\delta_i}(t))\mathbf{n_i}(a,x|t)da &= -(\boldsymbol{\alpha_i}(t) + \boldsymbol{\beta_i}(t) - \boldsymbol{\delta_i}(t)) \int_0^\infty \mathbf{n_i}(a,x|t)da \\
&= -(\boldsymbol{\alpha_i}(t) + \boldsymbol{\beta_i}(t) - \boldsymbol{\delta_i}(t))\mathbf{n_i}(x|t).
\end{aligned}$$

because of the independence of α_i and β_i of a and the definition (2.3.10). Hence, in the case $i = 0$ marginalizing returns the following equation

$$\frac{\partial \mathbf{n_i}(x|t)}{\partial t} + \frac{\partial(\nu(t,x)\mathbf{n_i}(x|t))}{\partial x} = -(\boldsymbol{\alpha_i}(t) + \boldsymbol{\beta_i}(t) - \boldsymbol{\delta_i}(t))\mathbf{n_i}(x|t).$$

In the case $i \geq 1$ marginalizing equations (2.1.5) yields

$$\frac{\partial \mathbf{n_i}(x|t)}{\partial t} - 2\gamma \int_{\mathbb{R}_+} \boldsymbol{\omega_{i-1}}(t)\boldsymbol{\alpha_{i-1}}(t)\mathbf{n_{i-1}}(a, \gamma x|t)da + \frac{\partial(\nu(t,x)\mathbf{n_i}(x|t))}{\partial x} = -(\boldsymbol{\alpha_i}(t) + \boldsymbol{\beta_i}(t) - \boldsymbol{\delta_i}(t))\mathbf{n_i}(x|t)$$

or

$$\frac{\partial \mathbf{n_i}(x|t)}{\partial t} + \frac{\partial(\nu(t,x)\mathbf{n_i}(x|t))}{\partial x} = -(\boldsymbol{\alpha_i}(t) + \boldsymbol{\beta_i}(t) + \boldsymbol{\delta_i}(t))\mathbf{n_i}(x|t) + 2\gamma \int_{\mathbb{R}_+} \boldsymbol{\omega_{i-1}}(t)\boldsymbol{\alpha_{i-1}}(t)\mathbf{n_{i-1}}(a, \gamma x|t)da.$$

This reformulated to scalar form is exactly the governing PDE system of the DCLSP (2.3.8).
It remains to show the statement for the initial conditions. For $i = 0$ the marginalitsation of the left hand side of the initial condition (2.1.9) is with (2.3.10)

$$\int_0^\infty \mathbf{n_0}(a, x|0)da = \mathbf{n_0}(x|0).$$

The marginalization of the left hand side of (2.1.9) is

$$\int_0^\infty \mathbf{n_0}(a)p_0(x)da = \mathbf{n_0}p_0(x).$$

This results in

$$\mathbf{n_0}(x|0) = \mathbf{n_0}p_0(x)$$

or in scalar form

$$n_0(x, j|0) = n_0(j)p_0(x).$$

For $i \geq 1$ the the initial conditions agree as well

$$\mathbf{n_i}(x|0) = \int_0^\infty \mathbf{n_i}(a, x|0)da = \int_0^\infty \mathbf{0}da = \mathbf{0}.$$

□

Hence, the DCLSP model is a special case of the ALDC model and all the information inherent to the DCLSP model is contained in the ALDC model. This can for example be used to test the numerical implementation of the ALDC model.

2.4 Numerical example

Equations (2.2.13) and (2.2.12) are used to extend the existing MATLAB-algorithm for the DALSP model. The algorithm computes the solution starting from $i = 0$ and iteratively calculates the solution for $i = 1, \ldots, S$ until a maximal division number S. The solution is calculated on a fixed grid of points in age- and time-direction. To compute the solution for a certain division number i, the solution for all $j = 1, \ldots, J$ with division number

Table 2.1: Table of rates and parameters used to produce the numerical example. The initial number of cells is denoted by $n_0(j)$.

parameter	$\log(\alpha_i^1)$ $i=0,\dots,5$	α_6^1	$\log(\alpha_i^2)$ $i=0,\dots,5$	α_6^2	$\log(\beta_i^1)$ $i=0,\dots,6$	$\log(\beta_i^2)$ $i=0,\dots,6$
value	$1+0.9i$	0	$0+0.5i$	0	$-1-0.3i$	-2-0.125i
parameter	ω^{12}	ω^{21}	$\log(n_0(1))$	$\log(n_0(2))$	$\log(k)$	$\log(c)$
value	0.6	0.7	8.1	8.0	-1.8	-3.8
parameter	μ_x	$\log(\sigma_x)$	μ_n	$\log(\sigma_n)$		
value	6.3	-1.8	2.2	-1.0		

$i-1$ have to be known. Integrals that occur during calculation are approximated by trapezoidal rule to make the scheme more time efficient. The runtime depends largely on the number of grid points and on the maximal number of divisions considered S. However, to achieve accurate results the number of necessary grid points might be large.

As shown in Section 2.3.2 integrating over a yields the DCLSP-model from [8]. Hence, for rates $\alpha_i^j(t,a) = \alpha_i^j(t)$ and $\beta_i^j(t,a) = \beta_i^j(t)$ independent of age integrating the solution of the ALDC model over age should yield the solution of the corresponding DCLSP model. The latter can be considered as a good approximation of the real solution. Comparing the evolution of cell number of cells with a certain cell type and in a certain generation in time for both models can give some indication of the accuracy of the numerical implementation of the ALDC model.

In the following three different grid sizes will be compared regarding runtime and accuracy. The number of divisions considered are six so $i = 0,\dots,6$ and the number of cell types is set to two, $J = 2$. There is no change of cell type without cell division, so $\delta_i^j(t,a) = 0$ for all $i \in \{0,\dots,6\}$, $j \in \{1,2\}$ and $t, a \in \mathbb{R}_+$. The rates $\alpha_i^j(t,a) = \alpha_i^j$ and $\beta_i^j(t,a) = \beta_i^j$ depend on cell type j and division number i, but not on time t or age a. The probability to change cell type during cell division $\omega_i^{j\tilde{j}}(t,a) = \omega^{j\tilde{j}}$ is independent of division number i, time t and age a. We consider cells of age $a \in [0,4]$, where we start with $n_0(j)$ cells with age zero $a = 0$, and a time period $t \in [0,3]$. The age- and time-scale have the same units, say for example days. The initial label distribution is assumed to be a log-normal distribution with parameters μ_x and σ_x and the dilution function is modeled by a Gompertz decay process with parameters k and c, $\nu(x,t) = -k\exp(-ct)x$ [5] and we set $\gamma = 2$. We model the background fluorescence by a log-normally distribution with parameters μ_n and σ_n. The rates and parameters are listed in Table 2.1.

In the first case (shown in Figure 2.1) the difference between two neighboring points, the step size, is set to be $\Delta = 0.01$ in both age and time direction which yields approximately 120 000 grid points. The Figure 2.1 shows the numbers of cells with a certain division number with respect to time. The straight line represents the simulation produced by the DCLSP model solved by the standard MATLAB solver ode45 as a line and the one obtained by using the numerical scheme to simulate the ALDC model as dots. For each cell type three exemplary division numbers are shown: $i = 0$, $i = 2$ and $i = 5$. The curves agree quite well for cells with few divisions $i = 0$ and $i = 2$. However the simulations for generation five $i = 5$ differ greatly. The runtime is approximately 0.6 seconds.

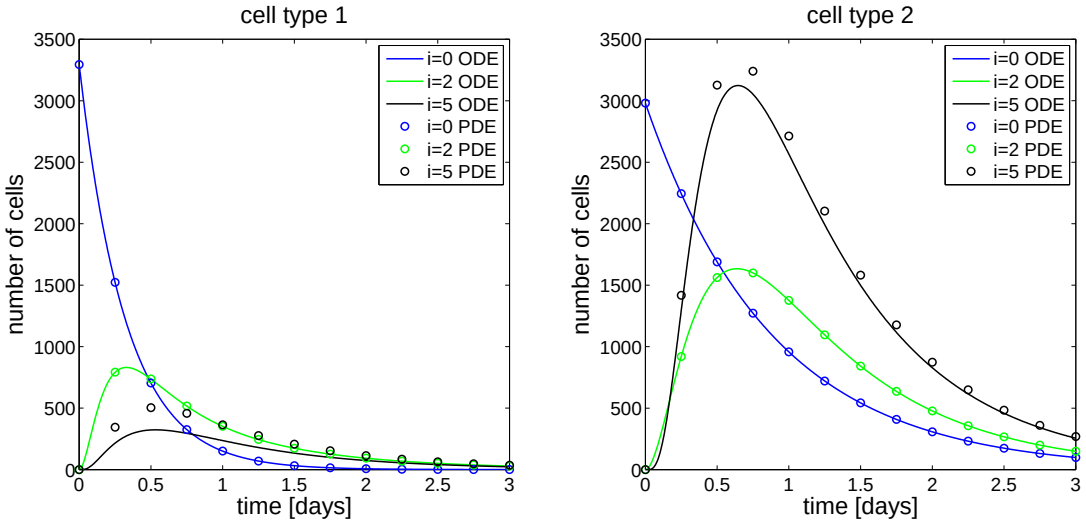

Figure 2.1: Comparison of the numerical solution of the ALDC model (PDE) represented as dots with the solution for the DCLSP (ODE) represented by a straight line for step size $\Delta = 0.01$. The figure depicts the time-dependent cell number for different generations. There is a noticeable difference between both simulations for generation $i = 5$.

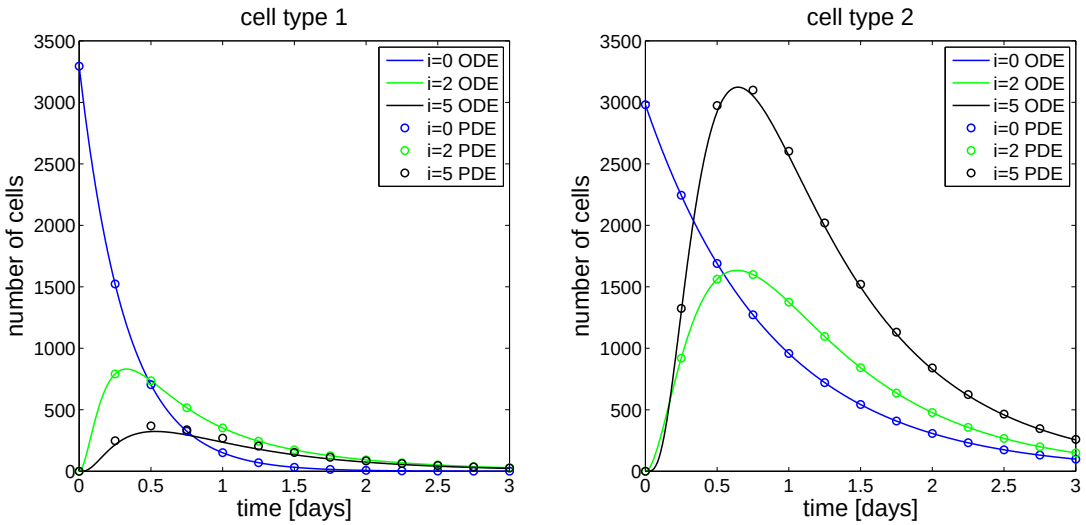

Figure 2.2: Comparison of the numerical solution of the ALDC model (PDE) represented as dots with the solution for the DCLSP (ODE) represented by a straight line for step size $\Delta = 0.005$. The figure depicts the time-dependent cell number for different generations. There is still a slight discrepancy between both curves for the fifth generation.

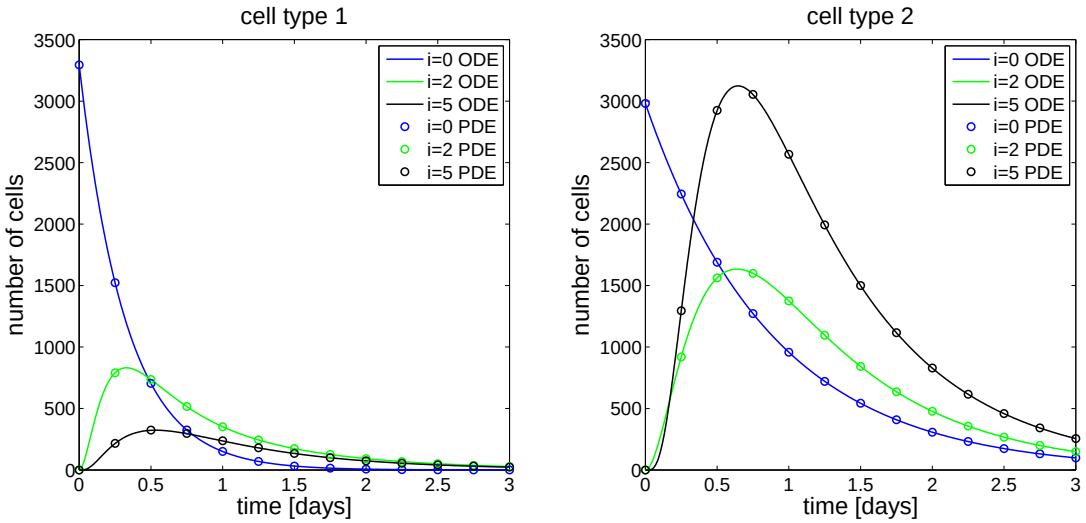

Figure 2.3: Comparison of numerical solution of the ALDC model (PDE) represented as dots with the solution for the DCLSP (ODE) represented by a straight line for step size $\Delta = 0.001$. The figure depicts the time-dependent cell number for different generations. The two curves are indistinguishable.

In the second case, shown in Figure 2.2, the step size is halved, $\Delta = 0.005$, corresponding to a grid size of around 480 000 points. Again the numbers of cells with division number $i = 0$ and $i = 2$ show little to no discrepancy. The difference in the case $i = 5$ is reduced in comparison to Figure 2.1, but it is still noticeable. The run time for this step size is roughly four seconds that is seven times the time needed in the first case.
For the third case the original step size is a tenth of the first one, i.e. $\Delta = 0.001$. The number of grid points considered now is hence 12 000 000. In this case the implementation for all division numbers seems to fit nicely to the curves produced by the ODE-solver as can be seen in Figure 2.3. However, the computational time is now around 230 seconds which is more than 350 times the time used for the simulation for $\Delta = 0.01$. Since parameter estimation needs to evaluate this model at every step of optimization, 230 seconds is much to long to be used, especially, since the number of divisions considered can exceed the number used for these simulations making the calculation even more time consuming.
To illustrate the properties of the ALDC model, we will now use the stepsize $\Delta = 0.001$ and simulate the model with the parameters given in Table 2.1 for the time interval $t \in [0, 1]$. In Figure 2.4 the composition of the total population regarding generation is illustrated. Over time the cells accumulate in generation six, because $\alpha_6^j = 0$ for all cell types j. Considering a larger time frame will not yield more insight as this trend will continue. The left subfigure of Figure 2.5 depicts the probabiltiy distributions for the measured fluorescence of a cell, i.e. the sum of CFSE-induced and background fluorescence, to lie in a certain bin of the histograms for different time points. The timepoints are distinguished by different colors. In this case six peaks can be identified in the histograms which correspond to the individual generations i. For real life data, however, this is effect appears rarely. Again the cells seem to accumualte in generation six. In the right subfigure the total number of cells is shown. The population growth is

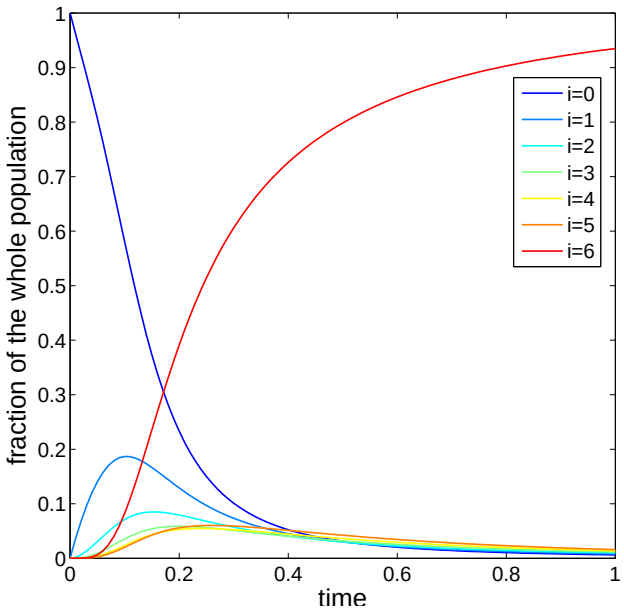

Figure 2.4: This figure depicts the percentage cells in generation i contribute to the total amount of cells at a time t. The different generations are represented by different colors.

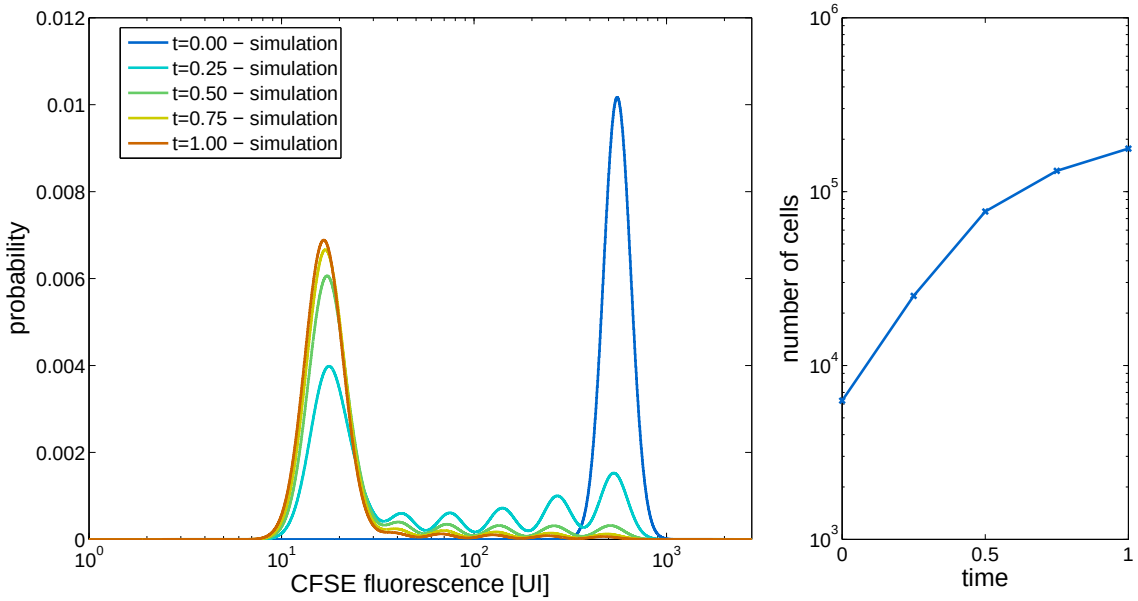

Figure 2.5: The left subfigure shows the fluorescence histograms obtained by simulating the ALDC model with the parameters listed in Table 2.1. The different colors code for the different time points. The right subfigure depicts the development of the total number of cells in the population.

slowing down, since there is no cell division in generation six only cell death and since towards $t = 1$ the majority of cells is already in the last considered generation.
In this example a small enough step size renders the numerical implementation of the ALDC model accurate enough for parameter estimation. However, the computation time is very high and the appropriate step size is not clearly evident from the initial setup of the problem.

2.5 Summary

In this chapter a new model has been developed to describe the proliferation of cells that can occur in different cell types and change those types in the progress of proliferation while possibly having age-, division number-, cell type- and time-dependent rates. Therefor a system of coupled PDEs was used. The solution has been derived by decomposition and the method of characteristics. Furthermore, the relation to existing models has been discussed. We extended the existing implementation for the DALSP model and found that it is not robust enough, slow and costly. This leads to the conclusion that the present implementation is too expensive. To avoid solving the PDE (2.2.9) the age structure can be discretized yielding a less complex ODE system.

Chapter 3

Cell state-, label-, division number- and cell type-structured population model for CFSE-data

In Chapter 2 we found that the numerical simulation of the ALDC model can require very small step sizes to attain a good approximation of the true solution. A key problem for the accuracy of the numerical scheme was the approximation of integrals in the solution of the PDEs modeling the population with continuous age structure. To circumvent this problem, we will establish in this chapter a new model class that, inspired by the cell cycle, accounts for discrete cell age states rather than continuous cell age. The model is named cell state-, label-, division number- and cell type-structured population (SLDC) model. Similarly to Section 2.2 the SLDC model will be decomposed. We then relate the SLDC model to existing models by comparing the division rates and the inter-division time.

3.1 Formulation of SLDC model

In this section we formulate the SLDC model. Therefor we consider a discrete age structure and introduce the discrete cell states $c \in \{1, \dots, C\}$. C denotes the cell state in which cell division takes place. A cell with division number i first has to go through all the states 1 to C to divide and arrive at division number $i+1$. Theoretically, a cell can change its cell type in each stadium of the cycle without cell division or change of state. This corresponds to the transitions with rate $\delta_i^{j\tilde{j}}$ in the age-dependent model. Analogously to the age-dependent model the cell type can also change during cell division. For simplicity, rates in the SLDC model that describe the same processes as in the ALDC model will be denoted with the same symbol as their age-dependent counter parts. Again we introduce the joint number density for generation i, $n_i(c, x, j|t)$. The number of cells of cell type j in cell state c that at time t have undergone i divisions and have a label concentration in a certain interval $x \in [x_1, x_2]$ can be computed by integrating the number density over the label x, $\int_{x_1}^{x_2} n_i(c, x, j|t)dx$.
The model accounts for the following processes,

- the change of the number density with respect to time, $\frac{\partial n_i(c,x,j|t)}{\partial t}$,

- the change of the number density due to label degradation with dilution rate $\nu(t,x)$, $\frac{\partial(\nu(t,x)n_i(c,x,j|t))}{\partial x}$,

- the change of the number of cells in the subpopulation with cell type j, division number i and state c due to cell death with rate $\beta_i^j(t,c)$ and change of cell state with rate $\rho_i^j(t,c)$, $-(\rho_i^j(t,c)+\beta_i^j(t,c))n_i(c,x,j|t)$,

- the change of number of cells in the subpopulation with division number i and cell type j and state $c \in \{2,\dots,C\}$ through advancement of cell state of cells in generation i with cell type j and cell state $c-1$ with rate $\rho_i^j(t,c-1)$, $\rho_i^j(t,c-1)n_i(c-1,x,j|t)$,

- the change of the number of cells in the subpopulation with division number i, cell type j and state $c=0$ through cell division of cells with division number $i-1$, cell type $\tilde{j}$ and state $c=C$ with division rate $\rho_{i-1}^{\tilde{j}}(t,C)$ with simultaneous change of cell type from type $\tilde{j}$ to j with probability $\omega_{i-1}^{j\tilde{j}}(t,C)$,
$2\gamma\rho_{i-1}^{\tilde{j}}(t,C)\omega_{i-1}^{j\tilde{j}}(t,C)n_{i-1}\left(C,\gamma x,\tilde{j}|t\right)$, and

- the change of the number of cells in the subpopulation with division number i, cell state c and cell type j due to spontaneous change of cell type from cell type $\tilde{j}$ to j cells with transition rate $\delta_i^{j\tilde{j}}(t,c)$,
$\delta_i^{j\tilde{j}}(t,c)n_i\left(c,x,\tilde{j}|t\right)$.

The dilution rate $\nu(t,x)$ is the same as in Chapter 2. The rates that describe cell aging, death, division and cell type change do now depend on the cell state. Their domain is $\{1,\dots,C\}\times\mathbb{R}_+$. The rates related to population dynamics are

- the death rate at time t for cells in generation i, cell state c and with cell type j, $\beta_i^j(t,c)$,

- the rate at time t at which cells in in generation i, cell state c and with cell type j change their cell state, $\rho_i^j(t,c)$,

- the probability at time t that a cell in generation i, cell state C and with cell type $\tilde{j}$ changes the cell type from $\tilde{j}$ to j during cell division, $\omega_i^{j\tilde{j}}(t)$,

- the rate at time t at which cells with division number i and cell state c and cell type $\tilde{j}$ spontaneously change their cell type from $\tilde{j}$ to j, $\delta_i^{j\tilde{j}}(t,c)$.

For $c=C$ the rate $\rho_i^j(t,C)$ additionally describes the rate at which cells in state C divide. The rate $\omega_i^{j\tilde{j}}(t)$ is factually independent of the cell state since cell division only occurs in state C. It holds that $\sum_{\tilde{j}=1}^{J}\omega_i^{j\tilde{j}}(t)=1$. The spontaneous change of cell type described by rate $\delta_i^{j\tilde{j}}(t,c)$ can occur in every cell state and does not change it.

The combination of the described processes yields the evolution equations

$$\begin{aligned}\frac{\partial n_i(c,x,j|t)}{\partial t} + \frac{\partial(\nu(t,x)n_i(c,x,j|t))}{\partial x} &= -(\rho_i^j(t,c)+\beta_i^j(t,c))n_i(c,x,j|t) \\ &\quad - \sum_{\tilde{j}=1}^{J}\delta_i^{\tilde{j}j}(t,c)n_i(c,x,j|t) + \sum_{\tilde{j}=1}^{J}\delta_i^{j\tilde{j}}(t,c)n_i\left(c,x,\tilde{j}|t\right) \\ &\quad + \begin{cases} 0 & i=0, c=1 \\ \rho_i^j(t,c-1)n_i(c-1,x,j|t) & i\geq 0, c>1 \\ 2\gamma\sum_{\tilde{j}=1}^{J}\rho_{i-1}^{\tilde{j}}(t,C)\omega_{i-1}^{j\tilde{j}}(t,C)n_{i-1}\left(C,\gamma x,\tilde{j}|t\right) & i>0, c=1 \end{cases}\end{aligned} \quad . \tag{3.1.1}$$

The initial conditions are

$$\begin{aligned} i=0: &\qquad n_0(c,x,j|0) = n_0(c,j)p_0(x) \\ i\geq 1: &\qquad n_i(c,x,j|0) \equiv 0. \end{aligned}$$

Analogously to the ALDC model, we denote the initial cell state distribution of cells of type j that did not undergo any cell division by $n_0(c,j)$. The number of cells with division numbers $i \geq 1$ is assumed to be zero at the time of staining, since cell division can only be observed after staining. The initial label distribution is denoted by the probability density $p_0(x)$ which describes the probability of a cell to have label concentration x after staining. We assume a product form of the initial conditions $n_0(c,x,j|0) = n_0(c,j)p_0(x)$, as the initial label distribution is independent of the initial cell state distribution.
Again, formulation in matrix-vector-form can improve the legibility by reducing summation to matrix-vector-multiplication. Therefor a vector $\mathbf{n_i}(c,x|t)$ is considered which, as before, is structured by cell type,

$$\mathbf{n_i}(c,x|t) = \begin{pmatrix} n_i(c,x,1|t) \\ n_i(c,x,2|t) \\ \vdots \\ n_i(c,x,J|t) \end{pmatrix}. \tag{3.1.2}$$

With the Definition 3.1.2 the model becomes

$$\begin{aligned}\frac{\partial \mathbf{n_i}(c,x|t)}{\partial t} + \frac{\partial(\nu(t,x)\mathbf{n_i}(c,x|t))}{\partial x} &= -(\boldsymbol{\rho_i}(t,c)+\boldsymbol{\beta_i}(t,c)-\boldsymbol{\delta_i}(t,c))\mathbf{n_i}(c,x|t) \\ &\quad + \begin{cases} 0 & i=0, c=1 \\ \boldsymbol{\rho_i}(t,c-1)\mathbf{n_i}(c-1,x|t) & i\geq 0, c>1 \\ 2\gamma\boldsymbol{\omega_{i-1}}(t)\boldsymbol{\rho_{i-1}}(t,C)\mathbf{n_{i-1}}(C,\gamma x,t) & i>0, c=1 \end{cases}\end{aligned} \tag{3.1.3}$$

with $\boldsymbol{\rho_i}(t,c)$ and $\boldsymbol{\beta_i}(t,c)$ being diagonal matrices with entries $\rho_i^j(t,c)$ and $\beta_i^j(t,c)$ respectively

$$\boldsymbol{\rho_i}(t,c) = \begin{pmatrix} \rho_i^1(t,c) & 0 & \dots & 0 \\ 0 & \rho_i^2(t,c) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_i^J(t,c) \end{pmatrix} \tag{3.1.4}$$

$$\boldsymbol{\beta_i}(t,c) = \begin{pmatrix} \beta_i^1(t,c) & 0 & \dots & 0 \\ 0 & \beta_i^2(t,c) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_i^J(t,c) \end{pmatrix}. \tag{3.1.5}$$

The matrix $\boldsymbol{\delta_i}(t,c)$ is defined analogously to (2.1.8)

$$\boldsymbol{\delta_i}(t,c) = \begin{pmatrix} -\sum_{\tilde{j}=1}^{J} \delta_i^{\tilde{j}1}(t,c) & \delta_i^{12}(t,c) & \dots & \delta_i^{1J}(t,c) \\ \delta_i^{21}(t,c) & -\sum_{\tilde{j}=1}^{J} \delta_i^{\tilde{j}2}(t,c) & \dots & \delta_i^{2J}(t,c) \\ \vdots & \vdots & \ddots & \vdots \\ \delta_i^{J1}(t,c) & \delta_i^{J2}(t,c) & \dots & -\sum_{\tilde{j}=1}^{J} \delta_i^{\tilde{j}J}(t,c) \end{pmatrix} \tag{3.1.6}$$

and the matrix $\boldsymbol{\omega_i}(t)$ is similar to (2.1.12) defined as

$$\boldsymbol{\omega_i}(t) = \begin{pmatrix} 1-\sum_{\tilde{j}=1}^{J} \omega_i^{\tilde{j}1}(t) & \omega_i^{12}(t) & \dots & \omega_i^{1J}(t) \\ \omega_i^{21}(t) & 1-\sum_{\tilde{j}=1}^{J} \omega_i^{\tilde{j}2}(t) & \dots & \omega_i^{2J}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \omega_i^{J1}(t) & \omega_i^{J2}(t) & \dots & 1-\sum_{\tilde{j}=1}^{J} \omega_i^{\tilde{j}1}(t) \end{pmatrix}.$$

The initial conditions are transformed to

$$\begin{aligned} i = 0: &\qquad \mathbf{n_0}(x,c|0) = \mathbf{n_0}(c)p_0(x) \\ i \geq 1: &\qquad \mathbf{n_i}(x,c|0) \equiv \mathbf{0}. \end{aligned} \tag{3.1.7}$$

Having formulated the SLDC model, we can now analyze it, analogously to the ALDC model.

3.2 Analysis of SLDC model

The SLDC model still consists of coupled systems of PDEs and is fairly complex. In the following, we will show that it can be simplified by decomposition, similarly to the ALDC model.

Theorem 3.2.1. *For dilution rates $\nu(x,t) = -k(t)x$ the number density $n_i(c,x,j|t)$ can be decomposed into two independent parts*

$$n_i(c,x,j|t) = n_i(c,j|t)p(x|i,t). \tag{3.2.1}$$

The function $p(x|i,t)$ describes the label distribution for a certain generation i, at a certain time t. It fulfills the same evolution equation (2.2.5) as in the last chapter

$$\frac{\partial p(x|i,t)}{\partial t} + \frac{\partial(\nu(x,t)p(x|i,t))}{\partial x} = 0 \tag{3.2.2}$$

with initial conditions

$$p(x|i,0) = \gamma^i p_0(\gamma^i x). \tag{3.2.3}$$

The number $n_i(c,j|t)$ of cells in generation i, cell state c and with cell type j has to satisfy

$$\begin{aligned}\frac{\partial n(c,i,j|t)}{\partial t} =& -(\rho_i^j(t,c)+\beta_i^j(t,c))n(c,i,j|t)\\ &-\sum_{\tilde{j}=1}^{J}\delta_i^{\tilde{j}j}(t,c)n(c,i,j|t)+\sum_{\tilde{j}=1}^{J}\delta_i^{j\tilde{j}}(t,c)n\left(c,i,\tilde{j}|t\right)\\ &+\begin{cases}0 & i=0,c=1\\ \rho_i^j(t,c-1)n(c-1,i,j|t) & i\geq 0,c>1\\ 2\sum_{\tilde{j}=1}^{J}\rho_{i-1}^{\tilde{j}}(t,C)\omega_{i-1}^{j\tilde{j}}(t,C)n\left(C,i-1,\tilde{j}|t\right) & i>0,c=1\end{cases}\end{aligned} \tag{3.2.4}$$

with initial conditions

$$\begin{aligned} i=0: &\qquad n(c,0,j|0)=n_0(c,j)\\ i\geq 1: &\qquad n(c,i,j|0)\equiv 0.\end{aligned}$$

In matrix-vector-form (3.2.1) becomes

$$\mathbf{n_i}(c,x|t) = \mathbf{n_i}(c|t)p(x|i,t) \tag{3.2.5}$$

with $\mathbf{n_i}(c|t)$ being the vector-function

$$\mathbf{n_i}(c|t) = \begin{pmatrix} n_i(c,1|t)\\ n_i(c,2|t)\\ \vdots\\ n_i(c,J|t)\end{pmatrix}. \tag{3.2.6}$$

The matrix-vector-formulation of (3.2.4) is

$$\begin{aligned}\frac{\partial \mathbf{n_i}(c|t)}{\partial t} =& -(\boldsymbol{\rho_i}(t,c)+\boldsymbol{\beta_i}(t,c)-\boldsymbol{\delta_i}(t,c))\mathbf{n_i}(c|t)\\ &+\begin{cases}\mathbf{0} & i=0,c=1\\ \boldsymbol{\rho_i}(c-1,t)\mathbf{n_i}(c-1|t) & i\geq 0,c>1\\ 2\boldsymbol{\omega_{i-1}}(t,C)\boldsymbol{\rho_{i-1}}(t,C)\mathbf{n_{i-1}}(C|t) & i>0,c=1\end{cases}\end{aligned} \tag{3.2.7}$$

with initial conditions

$$\begin{aligned} i=0: &\qquad \mathbf{n_0}(c|0)=\mathbf{n_0}(c)\\ i\geq 1: &\qquad \mathbf{n_i}(c|0)\equiv \mathbf{0}.\end{aligned} \tag{3.2.8}$$

Proof: We prove Theorem 3.2.1 by showing that plugging the ansatz (3.2.5) in both sides of (3.1.3) yields the evolution equation of (3.1.3) and that the initial conditions are consistent, analogously to the proof of Theorem 2.2.1 and [5]. The term obtained by

inserting the ansatz (3.2.5) on the left hand side can be differentiated in the next step using the product rule,

$$\begin{aligned}
&\frac{\partial \mathbf{n_i}(c,x|t)}{\partial t} + \frac{\partial(\nu(t,x)\mathbf{n_i}(c,x|t))}{\partial x} \\
&\quad = \frac{\partial(\mathbf{n_i}(c|t)p(x|i,t))}{\partial t} + \frac{\partial(\nu(t,x)(\mathbf{n_i}(c|t)p(x|i,t)))}{\partial x} \\
&\quad = \frac{\partial \mathbf{n_i}(c|t)}{\partial t}p(x|i,t) + \frac{\partial p(x|i,t)}{\partial t}\mathbf{n_i}(c|t) + \frac{\partial(\nu(t,x)p(x|i,t))}{\partial x}\mathbf{n_i}(c|t) \\
&\quad = \frac{\partial \mathbf{n_i}(c|t)}{\partial t}p(x|i,t) + \left(\frac{\partial p(x|i,t)}{\partial t} + \frac{\partial(\nu(t,x)p(x|i,t))}{\partial x}\right)\mathbf{n_i}(c|t).
\end{aligned}$$

From line three to four the terms were factorized using that $p(x|i,t)$ and its derivatives are scalar functions. We can now insert the evolution equations of the decomposed model (3.2.2) and get

$$\frac{\partial \mathbf{n_i}(c|t)}{\partial t}p(x|i,t) + \left(\frac{\partial p(x|i,t)}{\partial t} + \frac{\partial(\nu(t,x)p(x|i,t))}{\partial x}\right)\mathbf{n_i}(c|t) = \frac{\partial \mathbf{n_i}(c|t)}{\partial t}p(x|i,t). \quad (3.2.9)$$

When inserting (3.2.7) three cases have to be distinguished,

$$\begin{aligned}
i=0, c=1: \quad & \frac{\partial \mathbf{n_i}(c|t)}{\partial t}p(x|i,t) = -\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c|t)p(x|i,t) \\
i\geq 0, c>1: \quad & \frac{\partial \mathbf{n_i}(c|t)}{\partial t}p(x|i,t) = \left(-\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c|t)\right. \\
& \qquad \left. + \boldsymbol{\rho_i}(c-1,t)\mathbf{n_i}(c-1|t)\right)p(x|i,t) \\
i>0, c=1: \quad & \frac{\partial \mathbf{n_i}(c|t)}{\partial t}p(x|i,t) = \left(-\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c|t)\right. \\
& \qquad \left. + 2\boldsymbol{\omega_{i-1}}(t,C)\boldsymbol{\rho_{i-1}}(t,C)\mathbf{n_{i-1}}(C|t)\right)p(x|i,t).
\end{aligned}$$

The same three cases have to be distinguished, if we inserte the ansatz (3.2.5) on the right hand side of (3.1.3). For $i=0$ and $c=1$ we obtain

$$-\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c,x|t) = -\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c|t)p(x|i,t),$$

for $i\geq 0$ and $c>1$ the right hand side of (3.1.3) becomes

$$\begin{aligned}
&-\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c,x|t) + \boldsymbol{\rho_i}(c-1,t)\mathbf{n_i}(c-1x|t) \\
&\quad = -\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c|t)p(x|i,t) + \boldsymbol{\rho_i}(c-1,t)\mathbf{n_i}(c-1|t)p(x|i,t) \\
&\quad = \left(-\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c|t) + \boldsymbol{\rho_i}(c-1,t)\mathbf{n_i}(c-1|t)\right)p(x|i,t)
\end{aligned}$$

by inserting (3.2.5) and factorizing and in the case $i\geq 1$ and $c=1$ the right hand side of (3.1.3) is equal to

$$\begin{aligned}
&-\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c,x|t) + 2\gamma\boldsymbol{\omega_{i-1}}(t)\boldsymbol{\rho_{i-1}}(t,C)\mathbf{n_{i-1}}(C,\gamma x|t) \\
&\quad = -\left(\boldsymbol{\rho_i}(t,c) + \boldsymbol{\beta_i}(t,c) - \boldsymbol{\delta_i}(t,c)\right)\mathbf{n_i}(c|t)p(x|i,t) \\
&\qquad + 2\gamma\boldsymbol{\omega_{i-1}}(t)\boldsymbol{\rho_{i-1}}(t,C)\mathbf{n_{i-1}}(C|t)p(\gamma x|i-1,t).
\end{aligned}$$

For the first two cases inserting (3.2.5) gives the same result on both sides. In the third case we can use that $\gamma p(\gamma x|i-1,t) = p(x|i,t)$ for $\nu(x,t) = -k(t)x$ as we already showed in Section 2.2 to conclude that inserting (3.2.1) on the right and left hand side of (3.1.3) yields the same equations. Hence, the ansatz fulfills (3.1.3). For $i = 0$ the initial conditions are consistent as it holds that

$$\mathbf{n_i}(x|0) = \mathbf{n_0}(0)p(x|0,0) = \mathbf{n_0}p_0(x).$$

For $i \geq 1$ inserting the ansatz in both sides of (3.1.7) yields

$$\begin{aligned}&\mathbf{n_i}(x|0) \equiv \mathbf{0}\\ &\mathbf{n_i}(0)p(x|i,0) = \mathbf{0}\gamma^i p_0(\gamma^i x) \equiv \mathbf{0}.\end{aligned}$$

This concludes the proof. □

With this decomposition the label dynamics stay the same as in Chapter 2 and age, generation and cell type dynamics are simplified to an ODE system. This ODE system can be solved numerically by time efficient and accurate algorithms. In contrast to Chapter 2, we do not need an analytical solution for the subpopulation dynamics.

3.3 Relation of the SLDC model to existing models

Since the concept of discrete age structure is new in the context of the analysis of proliferation data, the question arises how the SLDC model relates to the existing model classes. Can it produce the same results as existing models? Is it more or less general than the existing models considered in this thesis? Especially interesting is the relation of discrete to continuous age structure.
The relation to the DCLSP model is clearly evident. The DCLSP model can be obtained by setting the number of cell states C in the SLDC model equal to one. As the DCLSP model is a generalization of the DLSP model, the DSP model and the LSP model, these are, as well, special cases of the SLDC model.
For the comparison of the SLDC model and the ALDC model we want to focus on the differences that arise due to the different age structure. The models can not be directly related as they are fundamentally different in their structure. We can, however, compare them in terms of inter-division time, i.e. the time a cell stays in a certain generation between to divisions.

Theorem 3.3.1. *If there is no flux between cell types and the rates $\rho_i(t,c)$ are continuously differentiable in time for all c and i, the rate $\alpha_i(t,a)$ of the ALDC model can be chosen for all i to yield an identical distribution of the inter-division time as the SLDC model with rates $\rho_i(t,c)$.*

Proof: We will prove Theorem 3.3.1 by constructing $\alpha_i(t,a)$ for a given function $\rho_i(t,c)$. When comparing the ALDC and the SLDC model in the decomposed form we find that the label dynamics are the same. To compare the influence of age structure for the SLDC model and the ALDC model, we therefore merely have to study the population dynamics of $\mathbf{n_i}^{ALDC}(a|t)$ and $\mathbf{n_i}^{SLDC}(c|t)$, respectively. The distribution of inter-division time of

cells corresponds to the flux of one generation into the next generation, if all cells have the same age $a = 0$ at time $t = 0$, there is no influx at $a = 0$, no cell death and there is no transition between cell types. This corresponds to the situation of the ALDC model and SLDC model at $i = 0$ with $\boldsymbol{\beta_i} = \mathbf{0}$, $\boldsymbol{\delta_i} = \mathbf{0}$ and $\boldsymbol{\omega_i} = \mathrm{I}$ being the J-dimensional unit matrix. As we do not consider transitions between cell types, we can without loss of generality assume that there is just one cell type $J = 1$. In this case the ALDC model is identical to the DALSP model as pointed out in Section 2.3. If $J = 1$ the function $\mathbf{n_i}^{SLDC}(c|t) = n_i^{SLDC}(c|t)$ in the SLDC, $\mathbf{n_i}^{ALDC}(a|t) = n_i^{ALDC}(a|t)$ in the ALDC and the rates become scalar functions. Furthermore, the initial age distribution lies completely at $a = 0$ in the ALDC model and $c = 0$ in the SLDC model, respectively, because all cells start with age zero. Thus, the initial age distribution of the ALDC model has to fulfill $n_0^{ALDC}(a|0) = 0$ for $a > 0$ and $\int_{\mathbb{R}_+} n_0^{ALDC}(a|0)da = 1$. This can be modeled using the delta distribution $\delta_0(a)$. The initial cell distribution is then $n_0^{ALDC}(a) = \delta_0(a)$. For the SLDC model we model the initial distribution by setting $n_0^{SLDC}(c|t) = (1, 0, \ldots, 0)^T$. The flux into the next generation is a function of t and since we start with a total number of one cell at $t = 0$, the flux into the next generation corresponds to the distribution of inter-division times.

To prove Theorem 3.3.1 by considering the flux into a certain generation, we can confine us to prove the statement for the inter-division time between generation $i = 0$ and generation $i = 1$. All other cases can be proved analogously by setting $\rho_0(t, c) = \rho_i(t, c)$ and $\alpha_0(t, a) = \alpha_i(t, a)$ because the inter-division time only depends on the rate $\rho_i(t, c)$ and $\alpha_i(t, a)$, respectively.

In the situation defined above the flux into $i = 1$ for the SLDC model is $f(t) := 2\rho_0(t, C)n_0^{SLDC}(C|t)$, where $n_0^{SLDC}(C|t)$ is the solution of the ODE

$$\begin{aligned}
\dot{n}_0^{SLDC}(1|t) &= -\rho_0(t, 1)n_0^{SLDC}(1|t) \\
\dot{n}_0^{SLDC}(2|t) &= -\rho_0(t, 2)n_0^{SLDC}(2|t) + \rho_0(t, 1)n_0^{SLDC}(1|t) \\
&\vdots \\
\dot{n}_0^{SLDC}(C|t) &= -\rho_0(t, C)n_0^{SLDC}(C|t) + \rho_0(t, C - 1)n_0^{SLDC}(C - 1|t).
\end{aligned} \tag{3.3.1}$$

The ODE is of the form $\dot{n} = F(t, n)$, where $F(t, n)$ is locally Lipschitz continuous in n and continuous in t as $\rho_i(t, c)$ is continuous on $\mathbb{R}_+$. With the theorem of Picard-Lindelöf there exists a unique continuously differentiable local solution. Because $\rho_i(t, c)$ is itself continuously differntiable, the distribution of inter-division times, $f(t)$, is a compositum of C^1 functions and thus a continuously differentiable function. For the ALDC model the flux into $i = 1$ is $2\boldsymbol{\psi_0(t)} = 2\int_{\mathbb{R}_+} \boldsymbol{\alpha_0}(t, a)\, \mathbf{n_0}^{ALDC}(a|t)da$ and can be calculated using the solution for $\mathbf{n_0}^{ALDC}(a|t)$ from (2.2.13)

$$\begin{aligned}
2\int_{\mathbb{R}_+} \alpha_0(t, a)\, n_0^{ALDC}(a|t)\, da &= 2\int_{\mathbb{R}_+} \alpha_0(t, a)\, \delta_0(a - t) \exp\left(-\int_0^t \alpha_0\left(\tilde{t}, \tilde{t} + a - t\right) d\tilde{t}\right) da \\
&= 2\alpha_0(t, t) \exp\left(-\int_0^t \alpha_0\left(\tilde{t}, \tilde{t}\right) d\tilde{t}\right).
\end{aligned}$$

Hence, the flux into generation $i = 1$ depends only on t and not on a. We can therefore consider $\alpha_0(t) = \alpha_0(t, t)$. To choose $\alpha_0(t)$ so that these functions are the same for given

$\rho_0(t,c)$ we have to solve the problem

$$\begin{aligned} 2f(t) &= 2\alpha_0(t)\exp\left(-\int_0^t \alpha_0\left(\tilde{t}\right)d\tilde{t}\right) \\ \Leftrightarrow f(t) &= \alpha_0(t)\exp\left(-\int_0^t \alpha_0\left(\tilde{t}\right)d\tilde{t}\right) \end{aligned}$$

Hence, we can deduct that if $f(t) > 0$ it follows that $\alpha_0(t) > 0$ and for $f(t) = 0$ it follows that $\alpha_0(t) = 0$. Therefor, if $f(t) \neq 0$, it holds that $\alpha_0(t) \neq 0$ as well and we can divide by $\alpha_0(t)$

$$\begin{aligned} \frac{f(t)}{\alpha_0(t)} &= \exp\left(-\int_0^t \alpha_0\left(\tilde{t}\right)d\tilde{t}\right) \\ \Leftrightarrow \log\left(f(t)\right) - \log\left(\alpha_0(t)\right) &= -\int_0^t \alpha_0\left(\tilde{t}\right)d\tilde{t}. \end{aligned}$$

The natural logarithm can be applied to $f(t)$ and $\alpha_0(t)$ since both functions are nonnegative and we excluded the case $f(t) = 0$ and $\alpha_0(t) = 0$. Derivation in time direction yields

$$\frac{\dot{f}(t)}{f(t)} - \frac{\dot{\alpha_0}(t)}{\alpha_0(t)} = -\alpha_0(t).$$

Hence, $\alpha(t)$ is either equal to zero or has to fulfill the following ODE

$$\dot{\alpha}(t) = \frac{\dot{f}(t)}{f(t)}\alpha_0(t) + \alpha_0^2(t). \tag{3.3.2}$$

The initial conditions can be derived from equation $\alpha_0(0) = f(0)$. With the existence theorem of Peano the ODE (3.3.2) has a solution, if $\frac{\dot{f}(t)}{f(t)}\alpha_0 + \alpha_0^2$ is continuous in t and α. As $f(t) \neq 0$ and continuously differentiable this is the case. Hence, we can construct a piecewise continuous $\alpha_0(t,t)$ that generates the same inter-division time between $i = 0$ and $i = 1$ in the ALDC model as the rates $\rho_i(t,c)$ in the SLDC model. □

With this finding the question arises, if there is an analogous result for the death rate $\beta_i(t,c)$. This question is much harder to answer because the observation of cell death in different cell states requires a flux between cell states. Therefor the analysis with respect to $\beta_i(t,c)$ cannot be done independently of $\rho_i(t,c)$ which complicates the calculations. Furthermore, it is not clear which quantity to use for the comparison of the impact of the death rates. Still, it would be desirable to show that, given the death rates of the SLDC model, the death rates of the ALDC model can be determined to yield the same result as the SLDC model.
For general rate $\rho_i(t)$ the calculation of $f(t)$ is complex and mostly not analytically possible. In the case $\rho_i(t,c) = \rho = \text{const}$ and $n_0(1|t) = 1$ for $c = 1$ and $n_0(c|t) = 0$ for $c \in \{2,\ldots,J\}$ a calculation is possible. The solution is the Erlang distribution with parameters ρ and C,

$$f(t) = \rho^C \frac{t^{C-1}}{(C-1)!}\exp(-\rho t)$$

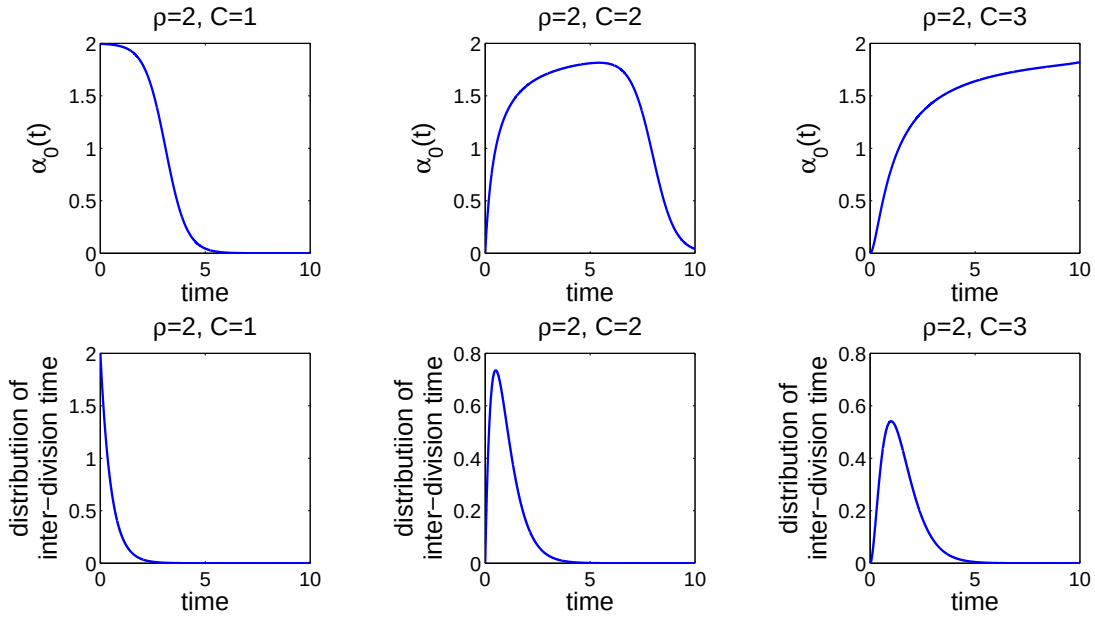

Figure 3.1: The figures in the top row depict the numerical computation of rate $\alpha_0(t)$ that has to be used in the ALDC model to create the same inter-division time as the SLDC model for rate $\rho = 2$ for $t \in [0, 10]$ distinguishing between three different cases regarding the number of cell states C. The left picture shows the computation for $C = 1$, the one in the middle the computation for $C = 2$ and the left figure shows the computation for $C = 3$. The figures in the bottom row depict the distribution of inter-division time derived by using the rate pictured in the figure above.

and

$$
\begin{aligned}
C = 1 : \dot{f}(t) &= -\rho^2 \exp(-\rho t) \\
C > 2 : \dot{f}(t) &= \frac{\rho^C}{(C-1)!} \left(-\rho t^{C-1} \exp\left(-\rho t\right) + (C-1)\, t^{C-2} \exp\left(-\rho t\right) \right) \\
&= \frac{\rho^C}{(C-1)!} \exp\left(-\rho t\right) t^{C-2} \left(-\rho t + C - 1\right).
\end{aligned}
$$

If we plug this in the ODE for α_0 we obtain

$$
\begin{aligned}
t = 0 : \alpha_0(0) &= f(0) = 0 \\
t > 0 : \dot{\alpha}_0 &= \left(\frac{C-1}{t} - \rho \right) \alpha_0 + \alpha_0^2.
\end{aligned}
$$

A numerical solution for this is can be calculated. In Figure 3.1 the computed $\alpha_0(t)$ is shown exemplary for $\rho = 2$ and three different C in the figures in the top row with the resulting inter-division time in the figures at the bottom. For $C = 1$ the numerically calculated rate is not constant, although the obvious solution would be to set $\alpha_0(t) = 2$ for all $t \in [0, 10]$. This is due to the numerical implementation and the special numerical properties of this problem. Furthermore, for large t it gets nearly impossible to determine $\alpha_0(t)$ numerically as $\exp\left(-\int_0^t \alpha_0\left(\tilde{t}, \tilde{t}\right) d\tilde{t} \right)$ becomes much smaller than one. This is a

possible reason for the behavior of the rate $\alpha_0(t)$ in the case $C = 3$. The rate is still increasing at $t = 10$ even though the inter-division time distribution is going to zero.
In this section we found that a lot models mentioned in the introduction can be seen as special cases of the SLDC. However, the ALDC model and the DALSP model are more general with respect to the inter-division time distribution. Which leads to the question if the SLDC still can produce a large variety of populations. This will be answered in the next section.

3.4 Numerical example

We now want to illustrate the properties of the SLDC with a small numerical example. Firstly, we discuss how to choose the number of cell states and with this result we will perform an exemplary simulation.

3.4.1 Choice of the number of cell states

The choice of the number of cell states C that shall be considered in the SLDC model can be discussed by reference to the case with $\rho_i(t) = \rho = \text{const}$. As mentioned before, in this case, the inter-division time distribution is an Erlang distribution. The choice of time- or generation-dependent $\rho_i(t)$ will increase the variability of the population. The objective is, hence, to find the number of cell states that can generate populations with homogeneous distributions of inter-division time. This enables us to model homogeneous as well as, with respect to inter-division time, more heterogeneous cell populations. In Figure 3.2 the influence of the number of cell states on the distribution of inter-division times is illustrated. The left subfigure shows Erlang distributions for different sets of parameters ρ and C where ρ is chosen in such a way that the mean of the distribution is one. The variance of the distribution of inter-division time decreases as the number of cell states increases. The right subfigure depicts the coefficient of variation depending on the number of cell states C. As the situation with constant ρ is the most homogeneous, the value of C determines the lower bound of the coefficient of variation. With the choice of $C = 20$ the coefficient of variation is already at around 0.22 and decreases further to approximately 0.16 for the choice of forty cell states $C = 40$. Heuristically, the choice of around twenty to forty cell states should enable the modeling of homogeneous enough cell populations as well as more variable ones.

3.4.2 Setup of the numerical example

For the numerical example we set the number of cell states to $C = 30$ as discussed above. We will consider a population consisting of $J = 2$ cell types, where one cell type ($j = 1$) divides faster than the other ($j = 2$) and there is change of cell type during cell division but no spontaneous change of cell type, i.e. $\boldsymbol{\delta_i}(t, c) = 0$ for all i, t, c. We will observe the behavior of this population for a time $t \in [0, 10]$ at the time points $t = \{0, 2, 4, 6, 8, 10\}$ and for $i \in \{0, \dots, 14\}$. The rates $\beta_i^j(t, c) = \beta^j$, $\rho_i^j(t, c) = \rho^j$ and $\omega_i^{j\tilde{j}} = \omega^{j\tilde{j}}$ only depend on the cell type j and if applicable $\tilde{j}$. The dilution rate $\nu(x, t)$ is modeled by a Gompretz decay process with parameters k_0 and c, i.e. $\nu(x, t) = -k_0 \exp(-ct)x$ [5] and

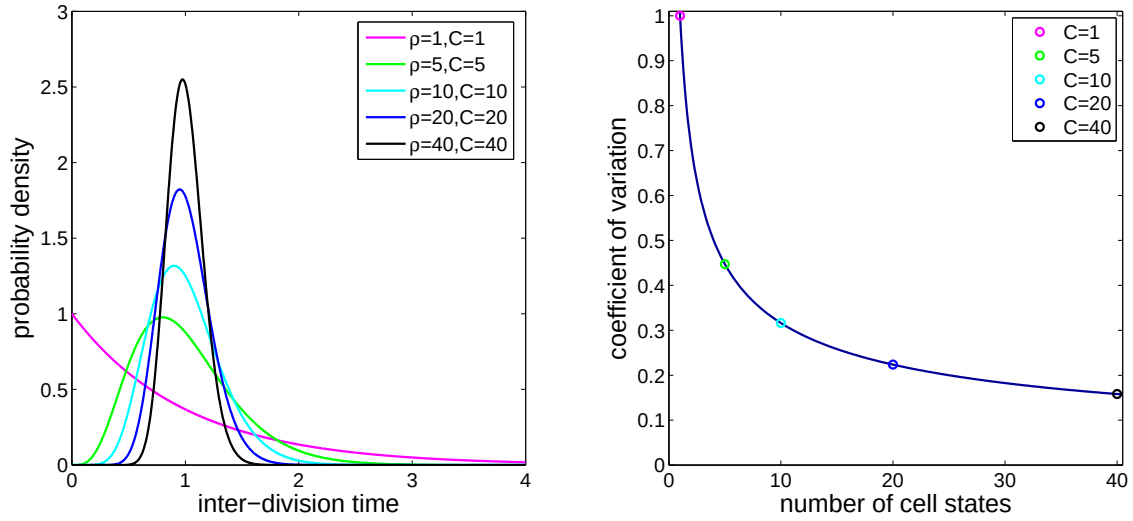

Figure 3.2: On the left the Erlang distribution for different values of ρ and C are shown. The rate ρ is chosen equal to the number of cell states C. Hence, the mean of the distribution of inter-division time is one. The variance of the distribution becomes less for increasing numbers of cell states C. The right figure shows the coefficient of variation for different numbers of cell states C. The coefficient of variation decreases with increasing numbers of cell states C.

Table 3.1: Parameters used for the numerical example

parameter	$\log_{10}(\rho^1)$	$\log_{10}(\rho^2)$	$\log_{10}(\beta^1)$	$\log_{10}(\beta^2)$	$\log_{10}(k)$	$\log_{10}(c)$
value	1.5	0.8	-5	-8	-3.5	-6.5
parameter	$\log_{10}(n_0(1))$	$\log_{10}(n_0(2))$	μ_x	$\log_{10}(\sigma_x)$	ω^{21}	ω^{12}
value	2	3	10	-0.2	0.1	0.4
parameter	γ	μ_n	σ_n			
value	2	35	25			

we set $\gamma = 2$. To model the background fluorescence we use a normal distribution with mean μ_n and standard deviation σ_n, $\mathcal{N}(\mu_n, \sigma_n^2)$. The initial age distribution in generation $i = 0$ is uniform, i.e. there are equally many cells in each cell state c. Hence, as initial distribution for cells with cell type j we get for all $c \in \{1, \dots C\}$, $j \in \{1, 2\}$ and $i = 0$ that $n_0(c, j) = n_0(j)$. We use a log-normal distribution with parameters μ_x and σ_x to model the initial label distribution, $p_0(x) = \log\mathcal{N}(\mu_x, \sigma_x^2)$. The used parameters are given in Table 3.1.

3.4.3 Simulation

The left subfigure of Figure 3.3 depicts the probability distribution of the binned fluorescence of the population with added normal distributed background fluorescence simulated by the SLDC model. Since one cell type divides faster than the other, the fluorescence histogram of day four exhibits two modes and the histograms of day six to ten are skewed to the right. The growth of the total population is depicted in the right subfigure. From day zero to day two the population grows slower than from day two to the next days,

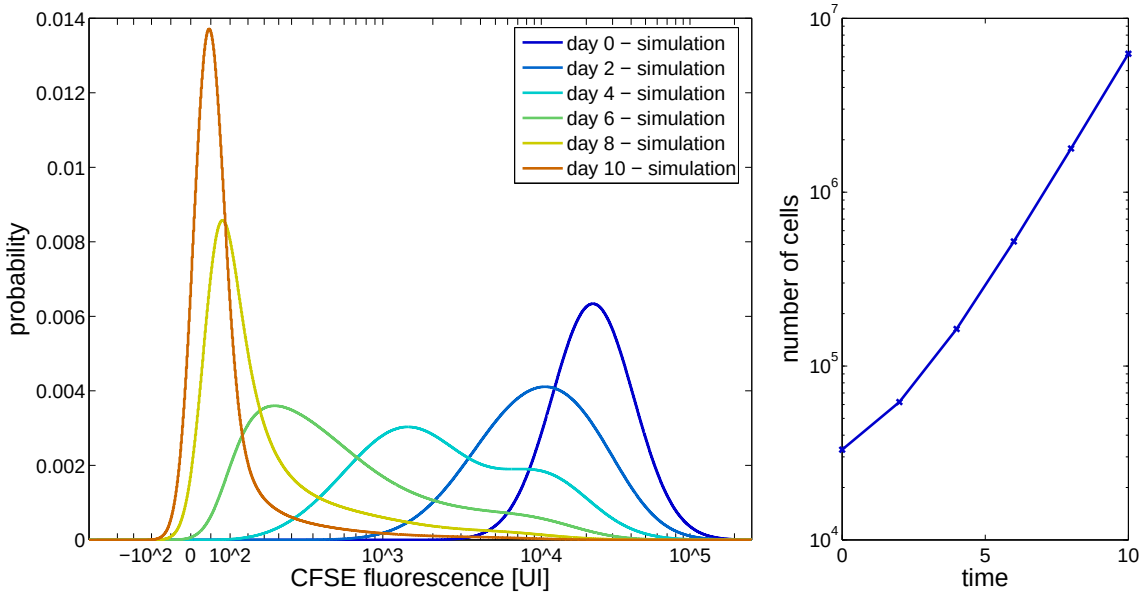

Figure 3.3: The left subfigure shows the fluorescence histograms obtained by simulating the SLDC model with the parameters listed in Table 3.1. The different colors code for the different time points. The right subfigure depicts the development of the total number of cells in the population.

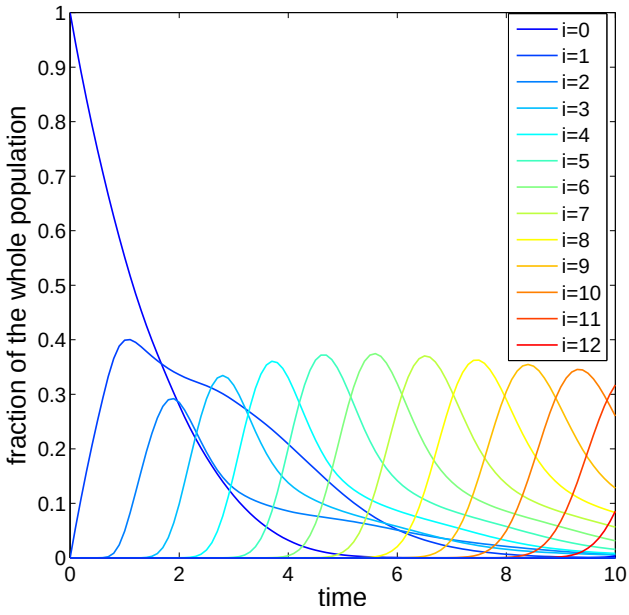

Figure 3.4: This figure depicts the percentage cells in generation i contribute to the total amount of cells at a time t. The different generations are represented by different colors.

as the second, more slowly dividing cell type starts with ten times more cells. After day two the majority of the growth originates from the faster growing first cell type. Figure 3.4.3 illustrates the composition of the total population in terms of generations which are

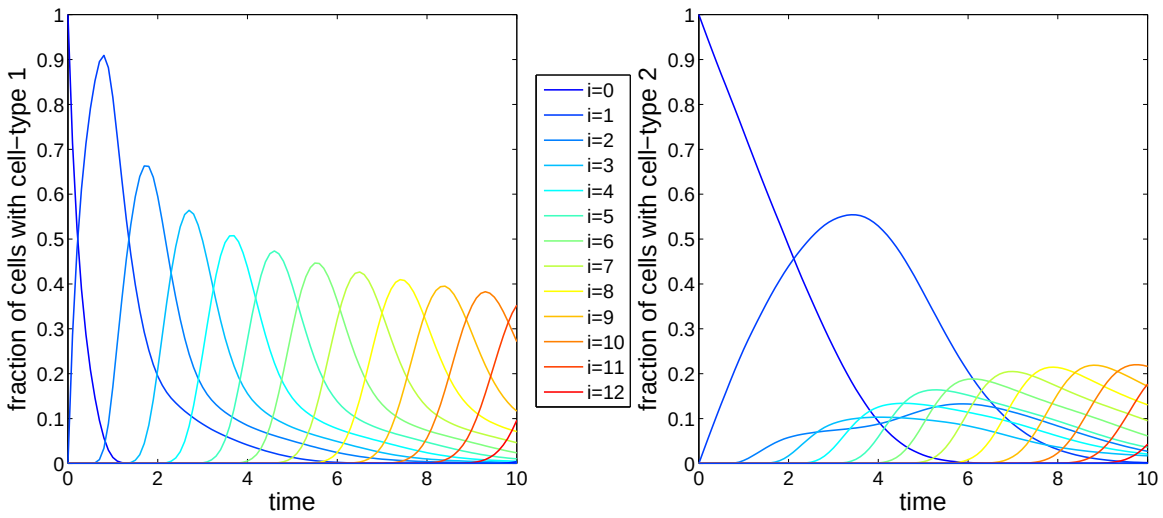

Figure 3.5: The figures show the percentage cells with cell type j and in generation i contribute to the total amount of cells with cell type j at time t.

Table 3.2: Generation- and cell type-dependent parameters used for the comparison runtime between ALDC model and SLDC model

parameter	$\log_{10}(\rho_i^1)$	$\log_{10}(\rho_i^2)$	$\log_{10}(\beta_i^1)$	$\log_{10}(\beta_i^2)$
value	$0.8 + 0.5i$	$0.8 + 0.2i$	$-1.5 - i$	$-5 - i$

represent by the different colors. It can be seen, that for the considered time span and parameters the cells divide at most twelve times. The slowly dividing cell type has a strong influence on the first generations and cells that divided only once are present until day eight. In the later generations the influence of the slowly dividing cell type decreases, as the fast dividing cell type outnumbers the former. In Figure 3.5 the generation structure for each cell type is shown. The subfigures depict the fraction each generation comprises of the total number of cells with the same cell type at time t. Here the influence of flux between cell types due to change of cell type during cell division can be observed. The faster dividing cell type is skewed to the right indicating an influx of cells that divide more slowly. The effect the change of cell type has on the subpopulation of the slow dividing cell type can best be seen in generation $i = 2$. There are obviously two peaks, the first one originating from cells of cell type $j = 1$ that changed their cell type during division, the second originating from regular cell division. An second aspect observable in these figures is that age structure with respect to cell division is inherent even without explicitly incorporating cell state-dependent rates ρ. There is an distinct delay between the emersion of one generation and the next.

3.4.4 Comparison of run time of the SLDC model and the ALDC model

We can modify this example to compare the run time of the SLDC model and the ALDC model. In the example in Section 2.4 we considered seven generations $S = 6$ and generation- and cell type-dependent rates during the time interval $t \in [0, 3]$. The accurate simulation of just the subpopulation sizes $\mathbf{n_i}(a|t)$ using the ALDC model with step size $\Delta = 0.001$ took approximately 175 seconds. To allow for comparability we consider the same number of generations and the same time interval for our numerical implementation of the SLDC model. As in the example in Section 2.4, the rates $\rho_i^j(t, c) = \rho_i^j$ and $\beta_i^j(t, c) = \beta_i^j$ depend generation and cell type. They are depicted in Table 3.2. The remaining parameters were left unchanged and can be found in Table 3.1. The simulation of the subpopulation sizes, $\mathbf{n_i}(c|t)$, takes roughly 0.5 seconds. Hence, the calculation of the population and subpopulation sizes with the SLDC model is more than 300 times faster than the calculation with our implementation of the ALDC model.

3.4.5 Results

Even though the ALDC model is in some sense more general than the SLDC model, a large enough choice of C renders the SLDC model capable to model a broad variety of cell populations. We could nevertheless produce an example of a complex population structure, even without exploting the entire possibilities generated by the potential structure of the rates $\rho_i^j(t, c)$ and $\beta_i^j(t, c)$. And furthermore, the reduced complexity of the system and especially the ODE form for $n_i(c, t)$ and a therefore much shorter run time compensate for the loss of generality.

3.5 Summary

In this chapter we introduced the SLDC model. The concept of cell states constitutes a novel and functional approach to the modeling of age structure in the context of parameter estimation for CFSE-data without losing significant capability compared to the modeling with continuous age structure. The SLDC model is capable of modeling age,- label-, division number- and cell type-structured populations using systems of PDEs. After decomposition, the system describing the population dynamics is simplified to an ODE system. This enables an efficient numerical implementation that is much faster and more robust than our numerical scheme developed for the ALDC model. Therefore, the SLDC model is a very suitable option for the simulation of proliferation data especially with regard to parameter estimation.

Chapter 4

Application of the SLDC model to CFSE-data for the proliferation of acute lymphatic leukemia

In this chapter we will present a preliminary analysis and parameter estimation of *in vivo* proliferation data of acute lymphatic leukemia using the SLDC model. We will start by presenting the data and then discuss the model-variants and the algorithm used for the parameter estimation. We will conclude the chapter by discussing the challenges. As parameter estimation is a complex problem, the main focus of this thesis was developing appropriate models and numerical algorithms. Aditionally, finding fitting parameters is a time consuming task and the used methods have to be adapted to the problem repeatedly. Therefore, the time frame of this thesis just allowed a preliminary analysis of the data.

4.1 Data and experimental setup

Acute lymphatic leukemia (ALL) is a cancer of bone marrow and blood, where white blood cells start growing uncontrolled. When the cancer is diagnosed it is immediately treated. At detection the cancer is normally already in an advanced state. Due to these circumstances there is little knowledge of the cancer proliferation *in vivo* and the early development of ALL. Yet, this knowledge could reveal potential drug targets and possibly enable an earlier detection and thus improve the prospects of patients.
To study the proliferation dynamics of ALL, the research team of Professor Jeremias (AGV Research Unit Gene Vectors, Helmholtz Zentrum München) extracted acute lymphatic leukemia cells from three patients. The cells were then preprocessed to express certain surface markers and red fluorescence to simplify the sorting of cancer and mouse tissue. At day zero the ALL cells got stained with CFSE and injected into immune suppressed mice. At different days the bone marrow got extracted. After the extraction the bone marrow cells were separated from the mouse cells using magnetic activated cell sorting (MACS) and fluorescence activated cell sorting (FACS) utilizing the transgenic properties of the human cancer cells. After these two purification steps most mouse cells are sorted out. The cancer cells are then used to create a CFSE fluorescence spectrogram. The Figure 4.1 (middle) shows the probability distribution of the binned fluorescence that

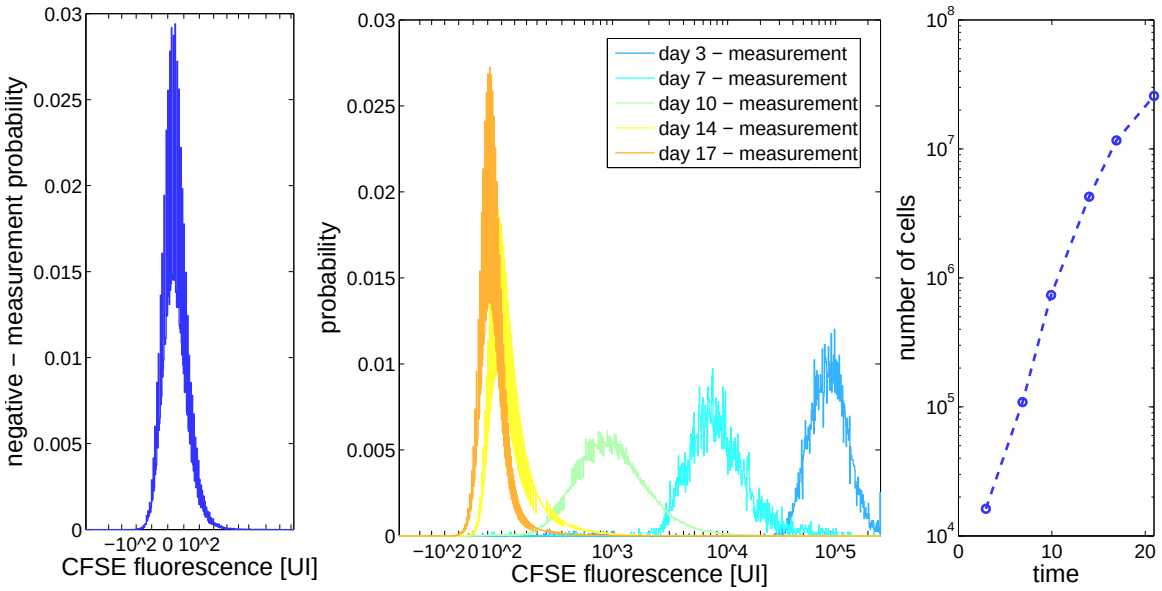

Figure 4.1: This figure depicts the data gathered by the team of Professor Jeremias. In the right figure the probability distribution of a cell to have a certain background fluorescence is depicted. The middle subfigure shows the binned fluorescence distribution of the cell population at different days where each day is represented by a different color. The right subfigure illustrates the total number of cancer cells measured at these days.

was measured in the cell population where the different days of extraction are represented by different colors. The right subfigure shows the total number of cancer cells measured. The measurement at day 0 is not depicted as it represents the fluorescence and total cell number at injection. Since just a fraction of the injected cells actually home, i.e. grow on onto the mouse tissue, the values at day zero do not represent the initial number and label distribution of the cells that actually proliferate. To account for the background fluorescence a fluorescence spectrogram of unlabeled cells, a negative control, was produced. It is shown in Figure 4.1 (left).

4.2 Model alternatives

As we do not know the actual processes that determine the behavior of the population, we will test various model variants and choose the one that fit the data best. These model variants differ in the assumed structure of parameters and population. We start by testing four model alternatives. The first three models consider a population consisting only of one cell type. Here the first variant, $\mathcal{M}_1$, assumes that the rates $\rho_i(t,c) = \rho$ and $\beta_i(t,c) = \beta$ are constant. This is the simplest case. The second model variant, $\mathcal{M}_2$, models a division and state advancement rate ρ that depends linear on the generation number i, $\rho_i(t,c) = \rho_0 + ik_\rho$. The rate $\beta_i(t,c) = \beta$ is constant. In the third variant, $\mathcal{M}_3$, we consider the flipped case where $\rho_i(t,c) = \rho$ is constant and $\beta_i(t,c) = \beta_0 + ik_\beta$ depends linear on the generation i. The last model variant we want to study, $\mathcal{M}_4$, considers a

population consisting of two cell types that are independent of each other, i.e. there are no fluxes between cell types. The rates $\rho_i^j(t,c) = \rho^j$ and $\beta_i^j(t,c) = \beta^j$ just depend on the cell type and are otherwise constant. In summary, we have the model variants

- $\mathcal{M}_1$: $J = 1$, $\rho_i(t,c) = \rho$ and $\beta_i(t,c) = \beta$ are constant,
- $\mathcal{M}_2$: $J = 1$, $\rho_i(t,c) = \rho_0 + ik_\rho$ is linearly dependent on the generation i and $\beta_i(t,c) = \beta$ is constant,
- $\mathcal{M}_3$: $J = 1$, $\rho_i(t,c) = \rho$ is constant and $\beta_i(t,c) = \beta_0 + ik_\beta$ is linearly dependent on the generation i,
- $\mathcal{M}_4$: $J = 2$, $\rho_i^j(t,c) = \rho^j$ and $\beta_i^j(t,c) = \beta^j$ just depend on the cell type j, $\boldsymbol{\delta_i}(t,c) = 0$ and $\boldsymbol{\omega_i}(t,c) = I$.

In all four cases, the dilution rate $\nu(x,t) = -k_0 \exp(-ct)x$ is modeled by a Gompertz decay process with parameters k_0 and c [5]. Like in Section 3.4 we assume an uniform initial age distribution $n_0(c,j) = n_0(j)$ for all $c \in \{1,\dots C\}$, $j \in \{1,2\}$ and $i = 0$. This results in a initial population size of Cn_0 cells. The initial label distribution is assumed to be a log-normal distribution with parameters μ_x and σ_x, $p_0(x) = \log\mathcal{N}(\mu_x, \sigma_x^2)$, and we set $\gamma = 2$. The background fluorescence will be modeled by a Johnson SU distribution with parameters γ_{JSU}, $\lambda_{\text{JSU}} > 0$, $\sigma_{\text{JSU}} > 0$ and ξ_{JSU} that will be determined before the parameter estimation from the control data. The parameters of a model that have to be estimated can be presented as a parameter vector θ. In our case the parameter vector for the four model variants is

$$\theta = [\theta_\rho, \theta_\beta, k_0, c, n_0, \mu_x, \sigma_x], \tag{4.2.1}$$

where θ_ρ and θ_β are the set of parameters used for the parameterization of $\rho_i(t,c)$ and $\beta_i(t,c)$, respectively, depending on the model variant.

4.3 Estimation algorithm

In the following, we will use maximum likelihood (ML) methods for the estimation of the model parameters. ML methods search for the set of parameters θ^* that maximizes the conditional probability, in other words the likelihood, that the measured data $\mathcal{D}$ was generated by our model with this set, i.e. we use the parameter set that maximizes the likelihood function $\mathcal{L}(\theta) = \mathbb{P}(\mathcal{D}|\theta)$ as estimator. The vector θ denotes the parameters of the model and the set of parameters θ^* is called the maximum likelihood estimator (MLE) [2, 5]. The measured data $\mathcal{D}$ is given as fluorescence histograms $\{\bar{H}_{t_k}^l\}_{l=1}^{d_l}$ and the total number of cells $\bar{N}_{t_k} = \sum_{l=1}^{d_l} H_{t_k}^l$ for each measured time point $t_k = 1,\dots,d_k$, where $l = 1,\dots,d_l$ denotes the fluorescence bin $[y_l, y_{l+1})$ and $\bar{H}_{t_k}^l$ is the number of cells that exhibit a total fluorescence $\bar{y}$, i.e. CSFE-induced fluorescence plus backgroundfluorescence, in this intensity range [2, 5]. We now want to deduct a likelihood function for this situation. The derivation of the likelihood function is in large parts not model specific and has been done in [2, 5]. We will present a short outline of the reasoning given in [2]. As stated in [2] the probability that the data was created by the SLDC model with a certain set of parameters θ is the product of the probabilities that the measurements at each time point t_k were generated by θ. The probability to observe a certain histogram $\{\bar{H}_{t_k}^l\}_{l=1}^{d_l}$

at time t_k can be transformed into the conditional probability to observe the histogram $\{\bar{H}^l_{t_k}\}_{l=1}^{d_l}$ given the total number of cells $\bar{N}_{t_k}$ and parameter vector θ times the probability to generate $\bar{N}_{t_k}$ cells with parameter vector θ [2]. Hence, the likelihood function for this case is

$$\mathcal{L}(\theta) = \mathbb{P}(\mathcal{D}|\theta) = \prod_{t_k=1}^{d_j} \mathbb{P}(\{\bar{H}^l_{t_k}\}_{l=1}^{d_l}|\bar{N}_{t_k}, \theta)\mathbb{P}(\bar{N}_{t_k}|\theta). \tag{4.3.1}$$

The two factors can be considered separatedly. Measuring the fluorescence intensity of $\bar{N}_{t_k}$ cells and sorting the cell into the respective bin is approximately a series of independent Bernoulli experiments with d_l possible outcomes. Hence, the probability to obtain a certain histogram $\{\bar{H}^l_{t_k}\}_{l=1}^{d_l}$ from $\bar{N}_{t_k}$ cells is multinomial distributed. The probability $p^l_{t_k}(\bar{y}|\theta)$ of a cell to exhibit a fluorescence intensity $\bar{y} \in [y_l, y_{l+1})$ and thus being sorted into bin l at time t_k depends on the chosen parameter θ and is computed from the SLDC model. The measured total fluorescence $\bar{y}$ consists of the CFSE-induced fluorescence, which is proportional to the label concentration in the cell with proportionality constant r, rx, plus background fluorescence y_a, $y = rx + y_a$. We can neglect the difference between label concentration and label-induced fluorescence, since the evolution equations stay the same and we do not know the initial label distribution [2, 5]. The sum of background fluorescence distribution and the calculated distribution of CFSE-induced fluorescence $p(x|i,t)$ is the convolution of both densities, denoted by $p(y|i,t)$ [2, 5]. We can now calculate the total fluorescence distribution of the cell population at time t_k,

$$n(y|t_k) = \sum_{i=0}^{S}\sum_{j=1}^{J}\sum_{c=1}^{C} n_i(c,j|t_k)p(y|i,t_k).$$

This distribution is then binned according to the bins of the data to obtain simulated histograms $\{H^l_{t_k}\}_{l=1}^{d_l}$. To calculate the probability that a cell is in a particular bin, $p^l_{t_k}(\bar{y}|\theta)$, we need to divide the number of cells in bin l by the total number of cells $N_{t_k} = \sum_{l=1}^{d_l} H^l_{t_k}$. As a correction for outliers we chose to add the constant value p_0/d_l to each histogram bin and normalize the the resulting distribution to yield integral one [2]. Hence, the probability that a cell randomly chosen out of a population of $\bar{N}_{t_k}$ cells exhibits a fluorescence intensity in the intensity range $\bar{y} \in [y_l, y_{l+1})$ is

$$p^l_{t_k}(\bar{y}|\theta) = \left(\frac{H^l_{t_k}}{N_{t_k}} + \frac{p_0}{d_l}\right)\frac{1}{1+p_0}.$$

With this definition of $p^l_{t_k}(\bar{y}|\theta)$, the likelihood to measure the histograms $\{\bar{H}^l_{t_k}\}_{l=1}^{d_l}$ given a measured total number of cells $\bar{N}_{t_k}$ is

$$\mathbb{P}(\{\bar{H}^l_{t_k}\}_{l=1}^{d_l}|\bar{N}_{t_k}, \theta) = \frac{\bar{N}_{t_k}!}{\prod_{l=1}^{d_l}\bar{H}^l_{t_k}!}\prod_{l=1}^{d_l} p^l_{t_k}\left(\bar{y}|\theta\right)^{\bar{H}^l_{t_k}}.$$

For the second factor, the likelihood of measuring $\bar{N}_{t_k}$ cells, $\mathbb{P}(\bar{N}_{t_k}|\theta)$, we assume, that the measurment error is log-normally distributed with parameters $\log\left(N_{t_k}\right)$ and σ^2_M

$$\mathbb{P}(\bar{N}_{t_k}|\theta) = \frac{1}{\bar{N}_{t_k}\sqrt{2\pi\sigma^2_M}}\exp\left(-\frac{\left(\log\left(N_{t_k}\right) - \log\left(\bar{N}_{t_k}\right)\right)^2}{2\sigma^2_M}\right).$$

Table 4.1: Lower and upper bounds for the optimization

parameter	k_0	c	μ_x	σ_x	n_0	$n_0(j)$	β, β_0, k_β, β^j	ρ, ρ^j	ρ_0, k_ρ
lower bound	10^{-4}	10^{-8}	6	10^{-1}	10^2	10^1	10^{-6}	10^{-3}	10^{-2}
upper bound	10^0	10^0	12	10^0	10^6	10^6	10^1	10^3	10^2

For numerical reasons we consider the log-likelihood,

$$\log \mathcal{L}(\theta) = \sum_{k=1}^{d_k} \sum_{n=1}^{\bar{N}_{t_k}} \log(n) - \sum_{l=1}^{d_l} \sum_{n=1}^{\bar{H}^l_{t_k}} \log(n) + \sum_{l=1}^{d_l} \bar{H}^l_{t_k} p^l_{t_k} - \frac{1}{2} \left(\log\left(2\pi\sigma_M^2 \bar{N}_{t_k}\right) - \left(\frac{\log(\bar{N}_{t_k}) - \log(N_{t_k})}{\sigma_M} \right)^2 \right).$$

To compute the MLE θ^*, we will perform multi-start local optimizations. The starting points for the local optimizations are sampled using latin hypercube sampling. The local optima found by the individual optimizations are compared and the parameter set corresponding to the largest one is used as MLE.
For the optimizations we will use the lower and upper bounds for the parameters provided in Table 4.1. These bounds are chosen on the basis of a previous analysis of the data performed Jan Hasenauer, PhD. The bounds for the initial population size, Cn_0, and the parameter μ_x of the initial label distribution can also be deducted directly from the data. By using the estimation algorithm outlined in this section and the optimizer for constrained optimization provided by MATLAB we can now find the MLEs for the model variants.

4.4 Preliminary results

In this section we will present some preliminary results and we will, furthermore, discuss how to improve the estimation algorithms.
As stated in Section 4.2, we want to approximate the background fluorescence by a Johnson SU distribution with parameters γ_{JSU}, $\lambda_{\text{JSU}} > 0$, $\sigma_{\text{JSU}} > 0$ and ξ_{JSU}. The probability density function of the Johnson SU distribution is given by

$$\frac{\sigma_{\text{JSU}}}{\lambda_{\text{JSU}}\sqrt{2\pi}\sqrt{\left(\frac{x-\xi_{\text{JSU}}}{\lambda_{\text{JSU}}}\right)^2 + 1}} \exp\left(-0.5 \left(\gamma_{\text{JSU}} + \sigma_{\text{JSU}} \sinh^{-1}\left(\frac{x - \xi_{\text{JSU}}}{\lambda_{\text{JSU}}} \right) \right)^2 \right). \tag{4.4.1}$$

By minimizing the square error of our simulation we find the set of parameters that produces the best approximation. The set of parameters is shown in Table 4.2 and Figure 4.2 depicts the best fit of the Johnson SU distribution in comparison to the data.
For the parameter estimation of the ALL data we performed multi-start local optimization with 100 starts. The computation time for this amount of starts is very high hence limiting the number of parameter estimations done in this thesis.
In the following, we want to present preliminary results for the parameter estimation by plotting the best fit together with the data. The figure for each model variant consist of

Table 4.2: Parameters of the Johnson SU distribution used for the modeling of the background fluorescence

parameter	γ_{JSU}	σ_{JSU}	λ_{JSU}	ξ_{JSU}
value	-0.9823	$10^{0.3988}$	$10^{1.9067}$	-12.1699

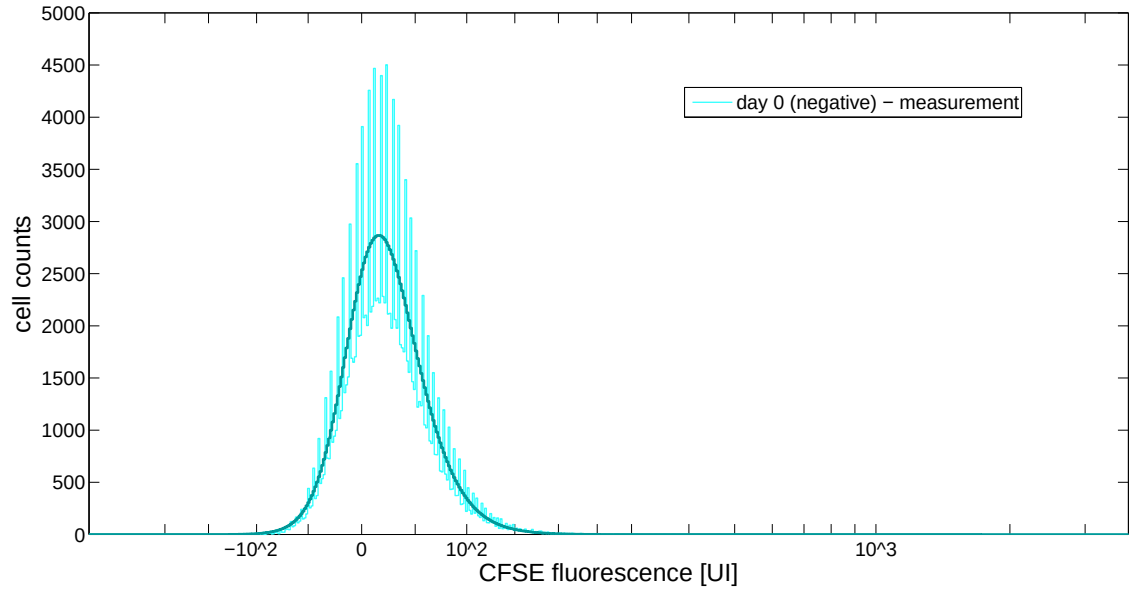

Figure 4.2: Measured background fluorescence and the fit with the Johnson SU distribution.

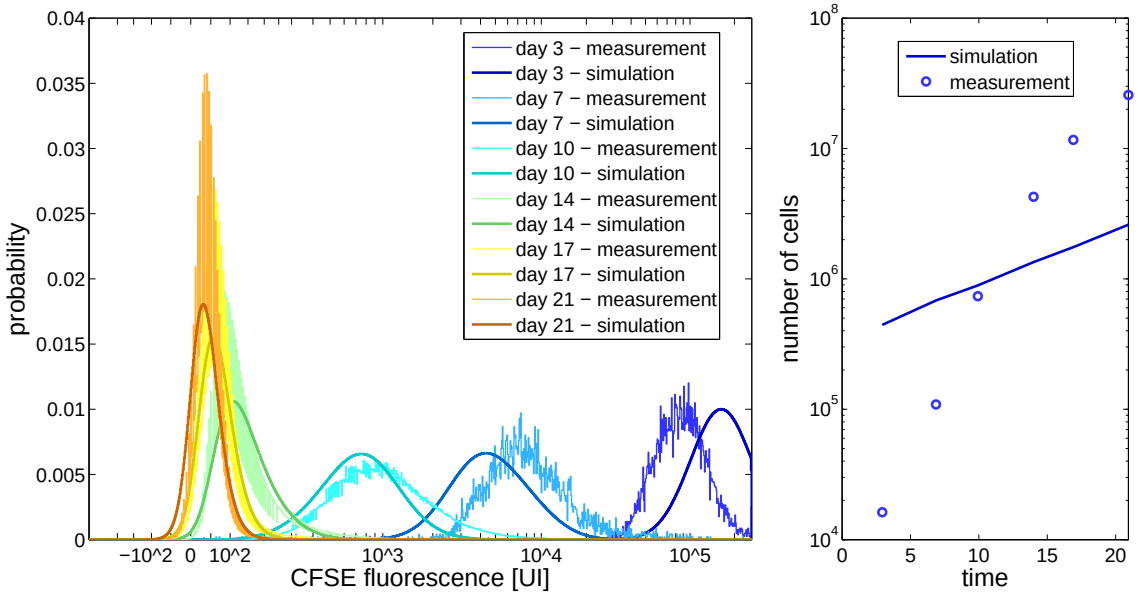

Figure 4.3: The left subfigure depicts the measured label histograms for different time points which are represented by different colors together with the corresponding prediction created by using model variant $\mathcal{M}_1$ of the SLDC model. The right subfigure illustrates the measured population size as dots and the predicted population size as a straight line.

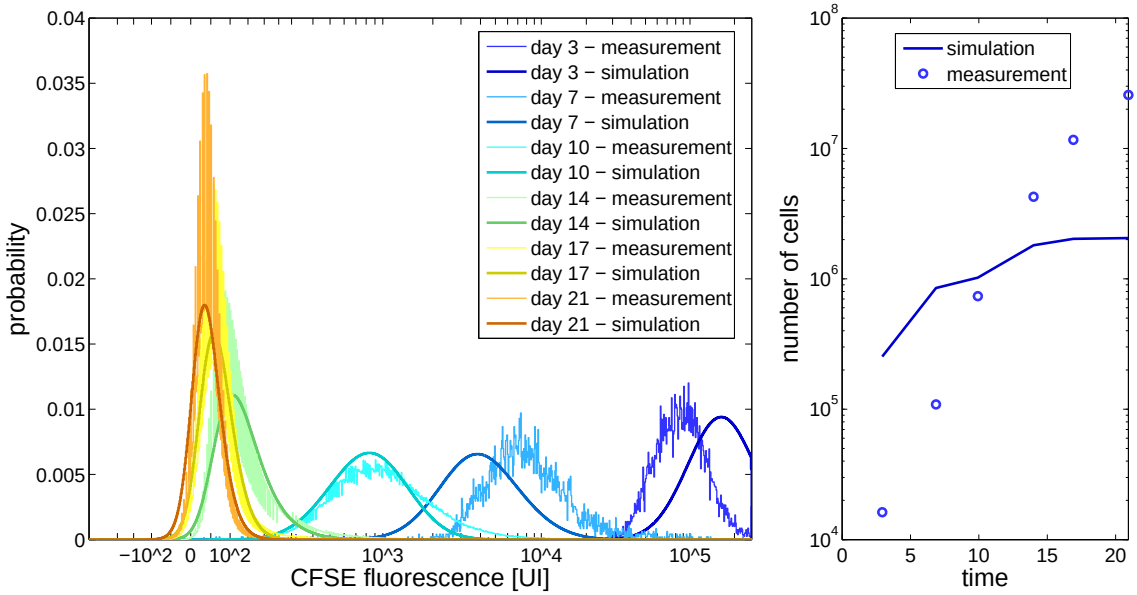

Figure 4.4: The left subfigure depicts the measured label histograms for different time points which are represented by different colors together with the corresponding prediction created by using model variant $\mathcal{M}_2$ of the SLDC model. The right subfigure illustrates the measured population size as dots and the predicted population size as a straight line.

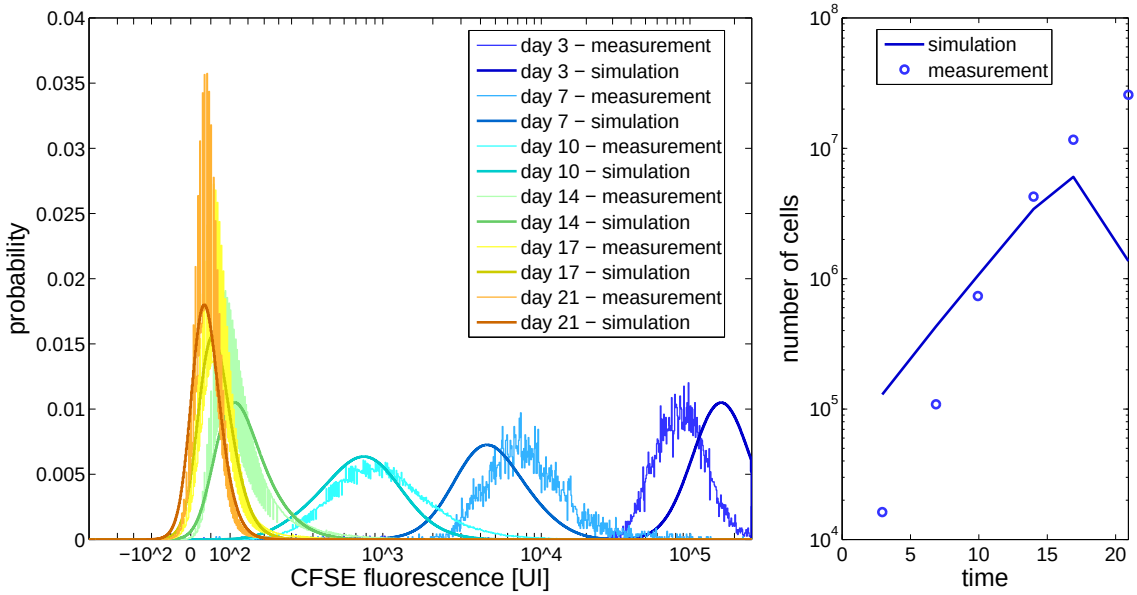

Figure 4.5: The left subfigure depicts the measured label histograms for different time points which are represented by different colors together with the corresponding prediction created by using model variant $\mathcal{M}_3$ of the SLDC model. The right subfigure illustrates the measured population size as dots and the predicted population size as a straight line.

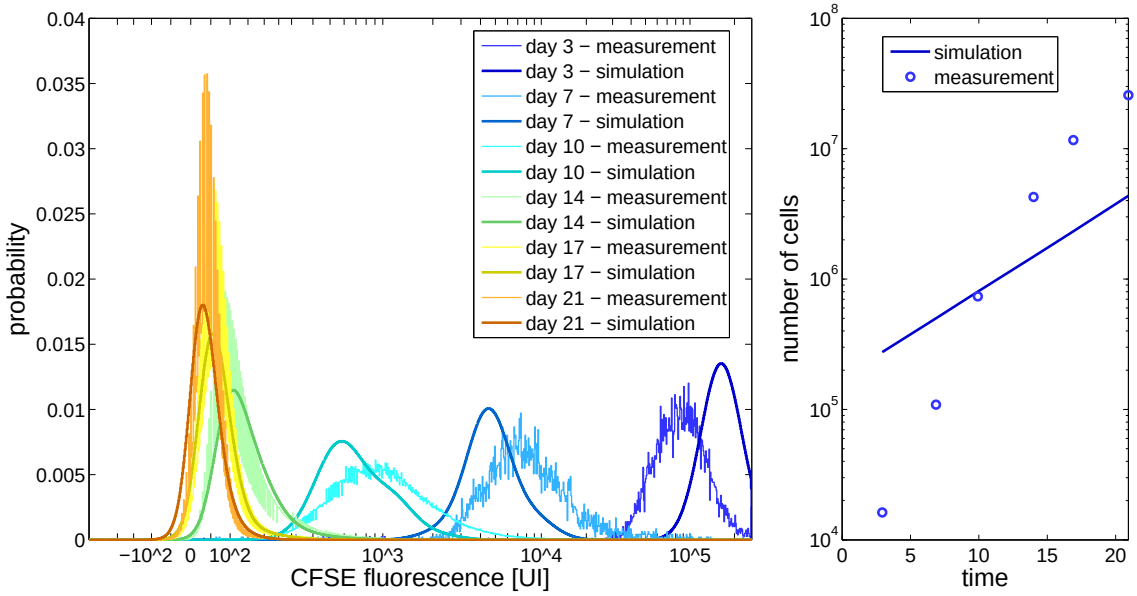

Figure 4.6: The left subfigure depicts the measured label histograms for different time points which are represented by different colors together with the corresponding prediction created by model using variant $\mathcal{M}_4$ of the SLDC model. The right subfigure illustrates the measured population size as dots and the predicted population size as a straight line.

two subfigures. The left one depicts the fluorescence histograms at the different extraction days represented by the lighter color together with the histograms predict by the model variant represented by the darker colors. The right subfigure displays the evolution of the total number of cells predicted by the model variant as a straight line and the measured total numbers as circles. Figure 4.3, Figure 4.4, Figure 4.5 and Figure 4.6 depict the best fit for model variant $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$, respectively. The predicted fluorescence intensity for the first three time points is apparently different from the measured one for all four model variants. However, the agreement of the prediction and the measurements seems to get better for later time points. The initial population size is overestimated by all model variants and it seems that none of the models can explain the population growth well. In conclusion, none of the fits is convincing.

There are several possible reasons for the unconvincing fits. An obvious problem is that the optimization reached the upper bound for μ_x in all cases. This is surprising, as the predicted label distribution for day 3 is already overestimating the measured distribution for $\mu_x = 12$. In addition, 100 starts of local optimizations seem to be insufficient. The best point is only found once and we think that there might be better estimates. Hence, the significance of these results is limited to begin with. Yet, the bad fit can have further possible reasons. For example, the model variants are too simple and we need to consider more complex dependencies. Further, the assumption of log-normal distributed initial label concentration and uniform initial age distribution could be too restricting. Extending these approaches by taking into account a mixture of log-normal distributions to model the initial label distribution as well as more flexible initial age distributions could improve the fit for the first day. The optimization could be improved by more sophisticated and reliable gradient computations using the sensitivity equation of the SLDC model. The

fact that the fit of the predicted fluorescence intensity distribution is better for later time points than the earlier ones might suggest that the amount of measured cells has an influence on the value of the likelihood function. A good fit of the fluorescence intensity at time points at which more cells are measured seems to be favored in the optimization. Hence, one could argue for using a modified objective function that balances out these effects. The influence of the suggested approaches remains to be investigated.

4.5 Summary

In this chapter we presented the proliferation data for acute lymphatic leukemia gathered by the team of Professor Jeremias. We proceeded by outlining the estimation algorithm to use for parameter estimation in the case of CFSE proliferation data. In the last section we presented preliminary estimation results for four model-variants. These results were not convincing. We examined possible reasons and discussed options for the improvement of the estimation results. It is clear, that the parameters estimated so far do not explain the ALL data. However, these were just preliminary results which already provided significant insights into the problems which have to be addressed. Furthermore, we note that using classical age-structured models would have been computationally infeasible. Further research with using the suggested improvements will hopefully yield more fitting outcomes.

Chapter 5

Conclusion and outlook

5.1 Conclusion

In this thesis we introduced two models for the analysis of CFSE-data that extend the existing models by considering division number-, age-, label- and cell type- structured populations.
The first model, the ALDC model, incorporated a continuous age structure and can be seen as a generalization of the existing models. In particular, it is a combination of the DCLSP model introduced in [8] and the DALSP model presented in [5, 6]. We showed, that the model can be decomposed into two subsystems governing label and population dynamics respectively. By studying a numerical example we found, however, that our numerical implementation of the ALDC model is not robust enough and, especially for complex population structures, too time consuming to be used for parameter estimation.
To circumvent the problems of the ALDC model, we introduced the SLDC model that uses cell states to create a discrete age structure and created a novel approach to modeling age structure in the context of CFSE-data. The SLDC model can be decomposed into a set of PDEs governing the label dynamics and a system of ODEs governing the population dynamics. By using an ODE-solver the population dynamics can be simulated very efficiently. We further found that the SLDC model is a generalization of many existing models, but less general than the ALDC model. However, the number of cell states can be chosen in such a way, that the SLDC model can generate a broad enough variety of cell populations. In some sense, the SLDC model can be interpreted as a discretization of the ALDC model.
The estimation algorithms developed for the parameter estimation using existing models could, in large parts, be adopted for the parameter estimation of the presented ALL proliferation data using the SLDC model. We presented preliminary results that are not yet convincing. We encountered some challenges regarding the parameter estimation of the ALL data. As the population was studied up to 21 days we needed a model capable of dealing with many generations. Moreover, the multi-start optimizations were computationally very costly yielding a long run time. Furthermore, we were dealing with *in vivo* data which makes outliers in the data or unobserved factors more probable. It was further not possible to identify a subpopulation- or generation-structure directly from the data. Hence, we had to start by testing some basic model variants. It also seems that the varying sample sizes complicate the optimization. The value of the objective function is

likely dominated by the likelihood at time points for which a large amount of cells was measured.

5.2 Outlook

One can think of several ways to advance and refine the SLDC model for further applications. For example, the SLDC model could be adapted to model asymmetric cell division. Another idea could be to include signaling pathways into the model creating a tool for the analysis of data composed of CFSE histograms and measurements of a pathway component. An exemplary application for this extended model could be the study of the effects of therapeutic agents on the cell proliferation.
As seen in Chapter 4, the preliminary results do not yield a good approximation of the data. Hence, further work on the parameter estimation is necessary. The next steps should include the improvement of the estimation algorithm by including sensitivities to improve the optimization as well as incorporating uncertainty analysis of the estimates by using profile likelihoods to determine the significance of the estimates. If the improved algorithm for parameter estimation still does not yield good results for the proposed model variants, more complex model variants have to be tested.
After improving and extending the estimation algorithm, the parameter estimation for the remaining two patient samples gathered by the team of Professor Jeremias and later a combined analysis of all three data sets should be done. By this it could be possible to identify differences between the samples and to draw inferences about the differences in the progression of ALL in different patients.
A major future aim would be to do model selection using for example the Akaike information criterion (AIC) and Bayes information criterion (BIC) to find the model variant that most likely explains the biological situation, in particular, answering whether the cell population consists of one or more cell types.
Furthermore, it would be interesting to test the SLDC model for other types of CFSE proliferation data as the full potential of this new approach has still to be determined.
Clearly, there are many open questions, however, the SLDC model we introduced in this thesis is the first model that allows for addressing these issues in a computationally manageable manner.

Bibliography

[1] H.T. Banks, K.L. Suttona, W.C. Thompson, G. Bocharov, D. Roose, T. Schenkel, and A. Meyerhans. Estimation of cell proliferation dynamics using CFSE data. *Bulletin of Mathematical Biology*, 73:116–150, 2010.

[2] J. Hasenauer. *Modeling and parameter estimation for heterogeneous cell populations.* PhD thesis, University of Stuttgart, 2013.

[3] J. Hasenauer, D. Schittler, and F. Allgöwer. Analysis and simulation of division- and label-structured population models: A new tool to analyze proliferation assays. *Bulletin of Mathematical Biology*, 74:2692–2732, 2012.

[4] T. Luzyanina, D. Roose, T. Schenkel, M. Sester, S. Ehl, A. Meyerhans, and G. Bocharov. Numerical modelling of label-structured cell population growth using CFSE distribution data. *Theoretical Biology and Medical Modelling*, 4:26, 2007.

[5] P. Metzger. A unified growth model for division-, age- and label-structured cell populations. Diploma thesis, Univerisity of Stuttgart, 2012.

[6] P. Metzger, J. Hasenauer, and F. Allgöwer. Modelling and analysis of division-, age- and label-structured cell populations. In *Proceedings of the Workshop on Computational Systems Biology (WCSB)*, Ulm, Germany, 2012.

[7] D. Schittler, J. Hasenauer, and F. Allgöwer. A generalized population model for cell proliferation: Integrating division numbers and label dynamics. In *Proceedings of the Workshop on Computational Systems Biology (WCSB)*, pages 165–168, Zürich, Switzerland, 2011.

[8] D. Schittler, J. Hasenauer, and F. Allgöwer. A model for proliferating cell populations that accounts for cell types. In *Proceedings of the Workshop on Computational Systems Biology (WCSB)*, Ulm, Germany, 2012.

[9] D. V. Widder. *Advanced Calculus.* Prentice-Hall, Inc., Englewood Cliffs, N.J., second edition, 1961.