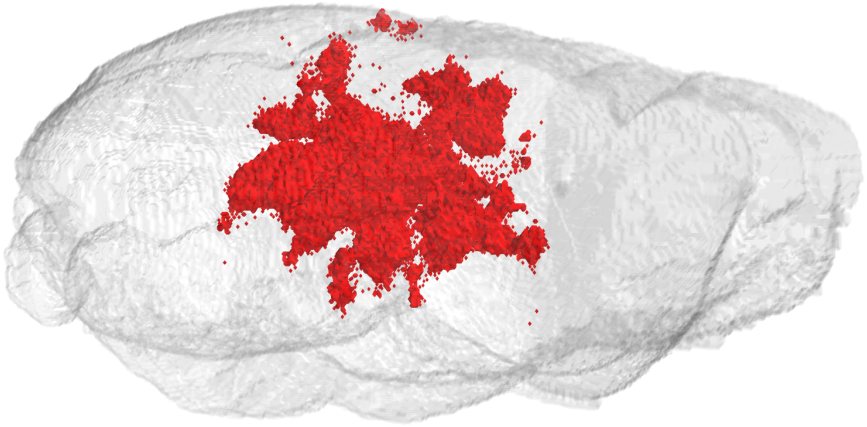
CHALMERS
UNIVERSITY OF TECHNOLOGY



Modelling Glioblastoma Growth in Anisotropic Tissue

Identification of Cell Line-Specific Parameters Governing Distinct Growth Patterns

Master's thesis in Engineering Mathematics and Computational Science

HENRIK BERGGREN

DEPARTMENT OF MATHEMATICAL SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

Master's thesis 2021

Modelling Glioblastoma Growth in Anisotropic Tissue

Identification of Cell Line-Specific Parameters Governing Distinct Growth Patterns

Henrik Berggren

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
Gothenburg, Sweden 2021

Modelling Glioblastoma Growth in Anisotropic Tissue
Identification of Cell Line-Specific Parameters Governing Distinct Growth Patterns
Henrik Berggren

Supervisor: Adam Malik, Department of Department of Immunology, Genetics and Pathology, Uppsala University
Supervisor: Philip Gerlee, Department of Mathematical Sciences, Chalmers University of Technology
Examiner: Philip Gerlee, Department of Mathematical Sciences, Chalmers University of Technology

Master's Thesis 2021
Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone: +46 31 772 1000

Cover: The brain cancer glioblastoma simulated in anisotropic tissue, here proliferation and migration of cancer cells are influenced by brain vasculature and white matter structures.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Modelling Glioblastoma Growth in Anisotropic Tissue
Identification of Cell Line-Specific Parameters Governing Distinct Growth Patterns
Henrik Berggren
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

Glioblastoma, or glioma grade IV is one of the most common types of brain cancer, unfortunately, the cancer is highly aggressive with short survival times related to the diffuse and invasive behaviour of the disease. In many cases, the cancer cells spread to large areas of the brain resulting in difficulties for therapeutic methods used to treat the disease. The mechanisms of glioblastoma spread and growth are still poorly understood, hence this study aims to contribute to the subject by an attempt to characterize tumour growth patterns observed in experimental studies.

With support in previous studies, we propose a novel modelling approach consisting of a stochastic, agent-based mathematical model for simulations of glioblastoma development. Anisotropic tissue properties of the brain, in terms of white matter tracts and the brain vasculature, are taken into consideration and influence the proliferation and migration of individual cells. The model is included in a framework for parameter inference, Approximate Bayesian Computation, to identify tumour characteristics from experimental data, originating from cell line-derived xenograft mouse models.

By studying parameters of synthetic data, tumours generated by the model itself, we show that model parameters are identifiable to a large extent. Specifically, the method is successful in the task of estimating known tumour properties such as the rate of migration and proliferation as well as the preference of migration in relation to the brain tissue. Furthermore, we investigate three experimental data sets in terms of coronal brain slices with respect to rates governing growth and spread as well as a quantification of the influence from the microenvironment. We present results showing a partial ability of the model to estimate parameters from the experimental data. To improve the estimation of parameters there is a need for an increased number of simulations to enhance the sampling density. Moreover, the growth patterns possible to simulate are affected by the available medical data which originates from different individuals compared to the experimental results. This introduces sources of error in the task to characterize observed tumour development.

Keywords: Glioblastoma, Glioma grade IV, Agent-based mathematical modelling, Approximate Bayesian Computation (ABC).

Acknowledgements

First and foremost, I would like to thank my supervisors Philip Gerlee and Adam Malik for their thorough preparation of this thesis, paving the way for my involvement and a successful project. I would also like to express my gratitude for their continuous support, making the thesis a genuinely enjoyable experience.

Further, I would like to thank all researchers in Sven Nelander's research group at Uppsala University for their generous contribution in form of derivation and processing of medical data as well as the execution of experimental studies, without which this study had not been possible.

Thanks also to Martin Raum at Mathematical Sciences for providing the essential high-performance computing resources for the study.

Henrik Berggren, Gothenburg, September 2021

Table of Contents

1 Introduction **1**
1.1 Aim . . . 2
1.2 Limitations . . . 3

2 Theory **5**
2.1 Former mathematical modelling of glioblastoma . . . 5
2.2 Medical imaging . . . 6
2.2.1 Diffusion tensor imaging . . . 7
2.2.2 Brain vasculature . . . 8
2.2.3 Xenotransplantation . . . 8
2.3 Bayesian inference . . . 9
2.3.1 Approximate Bayesian Computation . . . 9
2.3.1.1 Summary statistics . . . 11

3 Material and Methods **15**
3.1 Experimental tumour growth . . . 15
3.2 DTI data and pre-processing . . . 17
3.3 Brain vasculature and pre-processing . . . 17
3.4 Combination of medical data sets . . . 18
3.5 Cell-based model for glioblastoma growth . . . 18
3.5.1 Proliferation . . . 19
3.5.2 Migration . . . 19
3.6 Generation of synthetic data . . . 21
3.7 Threshold derivation, synthetic data . . . 23
3.7.1 Direct comparison . . . 24
3.7.2 Three-levelled comparison . . . 25
3.8 Parameter identification, synthetic data . . . 26
3.8.1 Prior selection . . . 27
3.8.2 Simulations and estimated parameters . . . 27
3.9 Parameter identification, experimental data . . . 27
3.9.1 Threshold derivation . . . 28
3.9.2 Prior selection, fixation of parameters . . . 28
3.9.3 Simulations and acceptance . . . 29

4 Results **31**
4.1 Model consistency . . . 31
4.2 Cell line-specific parameter identification . . . 35
4.2.1 Brain 1 . . . 35
4.2.2 Brain 2 . . . 37

4.2.3 Brain 3 39

5 Discussion **43**
5.1 Model consistency 43
5.2 Cell line-specific parameter estimation 45
5.2.1 Brain 1 45
5.2.2 Brain 2 46
5.2.3 Brain 3 47
5.3 Social and Ethical Aspects 47

6 Conclusion **49**

References **51**

A Complementary results of parameter estimations **A-1**

B Preliminary examination **B-1**

1. Introduction

Glioblastoma is the most common type of primary brain cancer (tumours with onset in the brain) among adults, moreover, it is one of the most aggressive classes of cancer resulting in short survival times for patients [1]–[5]. The severity of the disease is acknowledged by the World Health Organisation's classification of tumours where glioblastoma is categorized at the maximum grade, grade IV [6]. Hence the cancer is also termed glioma grade IV, the nomenclature originates from the name of glial cells where cancer have been believed to develop from. Recent studies, however, show that astrocytes, neural stem cells and oligodendrocyte progenitor cells could be precursor cells as well [4].

Despite persistent research regarding glioblastoma the progress concerning patient outcome have been limited. Glioma grade IV is currently incurable and exhibits a short expected survival time, around one year for affected patients treated with standard therapy [1]. For glioblastoma, the usual treatment consists of maximal possible tumour resection followed by radio- and chemotherapy [1], [2], [5]. Unfortunately, this treatment finds difficulty in removing the complete tumour and there is a high risk of tumour recurrence [2]. One of the main reasons for the inefficiency of the therapeutic methods is the invasive behaviour and the broad heterogeneity contained in the concept of glioblastoma [7]. Previously, glioma grade IV was denominated glioblastoma multiforme due to its vastly differing characteristics. The tumour irregularity is present both in tumour location and the topographic structure of the tumours [1]. The cancer cells often show a high activity of migration through brain tissue resulting in an invasion of cancer cells into distant parts of the brain compared to the primary tumour location [8]. For clinicians, this complicates the surgical resection and constantly leaves residual tumour cells, leading to tumour recurrence and the high resistance of glioblastoma to therapy [3].

The difficulty in treating glioblastoma is in part due to our lack of knowledge concerning the mechanisms governing migration and proliferation dynamics of glioblastoma cells. Furthermore, a large disparity has been observed in comparisons of *in vivo* and *in vitro* studies, where the former methods show a significantly increased invasive behaviour [7], suggesting the impact of brain tissue. This has been known for a substantial time, already in 1940 the pioneering work by Scherer showed an active migration of glioma cells away from the primary tumour, through the brain tissue [9].

Brain tissue is distinguished into two categories, white- or grey brain matter. White brain matter is a category of brain tissue constituted by myelin-coated nerve fibres and a supporting structure of glial cells. The second major tissue category is the

grey brain matter, containing a network of neural cell bodies and related glial cells with a high density [10].

The research by Scherer has been followed by multiple studies, specifically establishing the impact of white matter in relation to glioblastoma cell migration. In particular, cancer cells have been shown to migrate along white matter fibres [11], [12]. Similar effects have been observed for brain vasculature, where glioblastoma cells migrate along the surface of blood vessels [13], [14]. Additionally, studies have suggested an effect of increased cell proliferation with a sufficient supply of oxygen or excretion of certain signalling proteins [14], leading to an even stronger relationship between tumour development and brain vasculature. However, while brain structures, such as white matter and blood vessels, have been observed to have effects on cancer growth, the actual quantitative impact of mechanical and chemical interactions with the tumour microenvironment is to a large extent unknown [11], [13].

1.1 Aim

The present state of knowledge regarding glioblastoma growth and invasion supports further investigations to understand the heterogeneous behaviour of tumours resulting in the broad spectrum of different tumour characteristics and, ultimately, to improve the therapeutic methods. For example, by the means of pharmaceutical development impeding tumour cell migration to force compact tumours, which is easier to approach with surgery.

One of the ways to investigate tumour development and the different cancer characteristics is through *in silico* modelling of the cell behaviour. Here it is possible to simulate the effects of novel therapeutic methods or to investigate the impact of different anatomical structures on tumour morphology. In this study, we aim to approach the second issue by suggesting a stochastic mathematical model developed specifically to describe the spatio-temporal growth of glioblastoma. As mentioned, the heterogeneity of the brain has a significant impact on the behaviour of individual cancer cells and in extension the entire tumour. Hence, the suggested model will take advantage of information from medical data to describe the microenvironment of the tumour cells. With support of previous research, we will focus on the impact of white matter tracts (derived from diffusion tensor imaging) and brain vasculature to account for the complex structure of the brain both concerning cell migration and proliferation. The stochasticity of the model is included to capture the irregularity in cell proliferation and migration. The mathematical model will be implemented and used for simulations with the ultimate aim to estimate unknown tumour properties governing the distinct growth patterns observed in experiments.

It is possible to concretize the aim of the study into a set of sub-tasks that this study proposes to address.

1. Investigate different means to compare simulation outcome with experimental

data.

2. Examination of model consistency through parameter estimations of synthetic data.

3. Estimate model parameters governing growth patterns observed in experiments.

The first issue to investigate is different methods to compare the similarity of simulations with experimental data, this will be used to determine the correspondence between simulated tumours to growth patterns observed in experiments. Continuing, the second task is intended to establish the consistency of simulations with the mathematical model, i.e. is it possible to estimate known model parameters from synthetic data, generated by the model itself? If the model consistency is verified or possible weaknesses are identified, the study proceeds by estimating model parameters from experimental data and explore the third main issue of the study. Does the mathematical model have the possibility to simulate distinct and reliable results, describing tumour development and growth patterns, guided by the available medical data?

1.2 Limitations

Although the ultimate goal of this study is to contribute to the understanding of the human disease, this study will exclusively treat glioblastoma where the domain of simulations will be the brain of mice. This is the area where the experimental review has been performed and corresponding medical data of diffusion tensor imaging and vasculature are available. The diffusion tensor imaging originates from two mice derived by Jiang & Johnsson in [15]. The brain vasculature emerges from a study by Todorov *et al.* [16] where different mice are the subject of the data.

More in detail, the experimental results are characterised by cross-sectional scans of the mouse brain 40 weeks after xenotransplantation with human-specific antibodies highlighted to observe the tumour spread after the specified time, hence simulations will be carried out over the same time to find comparable results. Note that the experimental results originate from different individuals compared to the diffusion tensor imaging and blood vessel data, introducing sources of error in the study. In conclusion, the available experimental and medical data suggest the scope of the study and limits the investigation to simulations of tumour development of grade IV glioblastoma with possible applications in the brain of mice.

2. Theory

In the following chapter, the theoretical background for the remaining parts of the report is presented. The chapter includes previous modelling approaches to glioblastoma development, background regarding the medical images and statistical inference techniques used in this study.

2.1 Former mathematical modelling of glioblastoma

Mathematical modelling of glioblastoma is often used to investigate the mechanisms governing tumour formation, growth and dissemination. Mathematical modelling is an asset with the potential to improve the general outcome of the disease. For instance with the possibility to predict tumour development [17]. In the following section, we will present some previously used methods for describing glioblastoma growth.

In the 1990s, the modelling of glioblastoma was approached with deterministic mathematical models used to describe tumour growth, based on exponential or logistic growth laws. These temporal models were extended to spatial models with the same concept of growth [18]. However, it was found that straightforward exponential models, which had been successful in other types of cancer modelling, were not applicable for glioblastoma modelling due to the peculiar, diffusive, behaviour of the disease [18].

Starting from simple mathematical models, more advanced models have been designed. For example, models based on partial differential equations to describe the spatio-temporal variation of cell density. An innovative study was published in the year 2000 by Swanson *et al.* [19] where the brain structure was included in a reaction-diffusion equation to simulate tumours in a more anatomically accurate brain. Here the concept of variable cell migration was incorporated by different diffusion coefficients for white and grey matter. The spatial dependence of diffusion was determined from the, at the time, novel BrainWeb database [20] which provides a three-dimensional representation of location and proportions of white and grey matter in the brain [19].

More recently, Painter & Hillen [21] as well as Swan *et al.* [22] used diffusion tensor imaging, an imaging technique that allows to accurately derive white matter tracts, to account for anisotropic diffusion. The information was incorporated in the mathematical model by allowing for different rates of migration in different directions

and in this way model the phenomenon of glioblastoma migration along white matter fibres [22]. Models by Swan *et al.* as well as Painter & Hillen are examples of continuum models where the details of individual cells are disregarded in favour of a tumour perspective where measures such as local cell densities are simulated according to governing differential equations. One disadvantage of the aforementioned macroscopic methods, by construction, is the difficulty to describe phenomena on a cell-level spatial scale.

To resolve the disadvantages with the continuum models it is possible to reduce the spatial scale and model the tumour development on a microscopic scale. The microscopic modelling approach treats concepts related to cell-level scales, e.g. the dynamics of and between cells. For example, cell proliferation and migration describing the tumour development [17]. One such class of computational models are agent-based models. This type of model has diverse opportunities for application including tumour modelling where the agent can represent a cell. One set of cell-based models are cellular automaton models, where one or more cells inhabit one of the discrete lattice sites and move from site to site as the simulation proceeds, i.e. each agent has a time-dependent position in the simulation domain [23]–[25]. This approach provides a straightforward way to describe cell interaction with its microenvironment. This is used by Hatzikirou & Deutsch in [23], where the authors include tensor diffusion imaging information to simulate cell invasion in the brain. A similar study has been done by Alacrón *et al.* in [24]. Here tumour growth and tumour-induced formation of blood vessels are simulated in a heterogeneous environment with a cellular automaton model, with the inclusion of blood flow effects for transportation of cancer cells. The importance of blood vessels was also considered by Baker *et al.* in [26], where an agent-based model was used to investigate the interaction between brain vasculature and glioblastoma growth as well as the effect of therapy reducing the growth of new blood vessels.

With respect to the previous mathematical modelling techniques used for simulations of glioblastoma growth described in this section, we propose a stochastic, cell-based model aided by diffusion tensor imaging and brain vasculature to investigate cell line-specific growth patterns. The details of the model will be described in section 3.5.

2.2 Medical imaging

In this section, we will introduce the imaging techniques used in the remainder of the report. The medical images are included in this study to aid the mathematical model and to include essential tissue information to capture concepts on a cell level spatial scale. The medical information is derived from diffusion tensor imaging and the brain vasculature. Furthermore, the experimental results consist of image scans of a developed tumour, allowing for a comparison between simulations and experiments.

2.2.1 Diffusion tensor imaging

One way to study the tissue properties of the brain is the *in vivo* technique diffusion tensor imaging (DTI). DTI uses magnetic resonance images (MRI) as a source to estimate the rate of diffusion of water molecules in different directions [27]. Recent advances in magnetic resonance techniques in combination of staining of important regions provides a high isotropic resolution of $20-50\mu m$. Furthermore, MRI provides good contrast in the recreation of soft tissue and it is three-dimensional while being non-invasive [28]. Thus it is a suitable technique for brain imaging where the main use of the data is to get information on tissue composition.

The primary DTI data consists of a diffusion tensor (DT) for every voxel in the spatial domain. The tensor elements are formed by an application of magnetic field gradients to create an MR image representing diffusion in a specific direction. By repeated application for different directions, the complete diffusion tensor is estimated and in combination, all tensors create a tensor field of the spatial domain [29]. The diffusion tensor, $\mathrm{DT} \in \mathbb{R}^{3\times3}$, is a symmetric and positive definite matrix. As a result of the matrix properties there exists three orthogonal eigenvectors (v_1, v_2, v_3) with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The principal eigenvector, v_1 (corresponding to the largest eigenvalue) is aligned with the main diffusion direction [29].

To illustrate the diffusion, it is common to condense the information from the diffusion tensor to a scalar representation of the diffusion. This is possible to do through different techniques, for example through the average of the eigenvalues of the tensor as an estimation of the diffusion magnitude. This measure is often called mean diffusivity (MD) or apparent diffusion coefficient (ADC) [29]. Another of the most frequently used measures is fractional anisotropy (FA). This measure represents the part of diffusion deviating from isotropic diffusion (i.e. anisotropic diffusion) and is estimated as the normalized variance of the eigenvalues [29], according to:

$$\mathrm{FA} = \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}. \tag{2.1}$$

From the condensed measures describing diffusion, it is feasible to estimate tissue anatomy. Most importantly for this study, white matter tracts. As previously mentioned DTI describes the diffusion of water molecules that diffuse unhindered along the white matter tracts. However, in the perpendicular direction to white matter tracts the diffusion is minimal due to the lipid-rich myelin covering the axons [27]. Hence it is possible to derive the information of fibre location from the DTI data resulting in a spatial map of fibre tracts of white matter given in three dimensions [30].

2.2.2 Brain vasculature

Detailed information of the brain vasculature structure and composition is of high value in the research of many diseases, including glioblastoma. Common techniques to capture blood vessel information includes MR imaging and micro-computed tomography (micro CT). Unfortunately, both techniques fail to capture capillaries in brain tissue due to limitations in resolution [16]. Light-sheet microscopy or fluorescence microscopy is another common imaging technique for cell biology since it gives a possibility to observe physiology on a sub-cellular level [31], this also includes investigation of blood vessels. However, until recently it was only possible to use fluorescence microscopy for tissue with the maximum thickness of 200 µm, too shallow for a complete brain examination. But, with novel tissue-clearing techniques this obstacle has been defeated [16]. Recently Todorov *et al.* suggested an approach based on innovative tissue clearing techniques, two types of staining of blood vessels to allow for light-sheet microscopy of the complete brain domain. The three-dimensional scan from light-sheet microscopy was followed by image segmentation of blood vessels. The described approach resulted in an improvement in brain vasculature mapping with a resolution to the micrometre scale while maintaining the accuracy corresponding to manual segmentation by clinicians [16].

2.2.3 Xenotransplantation

There exist multiple experimental techniques to study tumour growth, of which xenograft models is one category. These methods consist of transplantation of cancer tissue from a diagnosed patient into immunodeficient mice. In the new environment the tumour is allowed to develop for a certain time, determined by research interest and ethical regulations [32]. One often distinguishes between specific types of xenograft models, in detail, we have the specific cell line-derived xenograft (CDX) and patient-derived xenograft (PDX) models. In CDX models, tumour cells from diagnosed patients are surgically extracted, cultivated externally before they are used for injection in immunocompromised mice. The difference in PDX is characterised by the resection of intact tissue to a larger extent and the transplantation is performed directly [33]. Both methods are frequently used with different advantages and disadvantages. CDX is popular for the efficiency in time and cost of the experiments. One downside of the method is that the cell lines often consists of late-stage cancer cells, which is not ideal for modelling early tumour development [33]. PDX models have the possibility to exhibit the heterogeneity in a population of cancer cells but it fails to recreate the specific tumour microenvironment and the interactions between tumour cells and the host immune system [32].

Moreover, to observe tumour propagation it is possible to perform labelling of cells before transplantation. Here one approach is to tag human-specific antibodies for staining to display the cancer cells and enhance the tissue contrast after the tumour growth [34]. This provides a way to study the migration of tumour cells in a natural environment where the final state of the tumour is possible to capture with a scan of fluorescent antibodies. In general, it is possible to decide the initial tumour location as well as the number of transplanted cells or the tumour size in xenograft models.

For this study it is possible to take advantage of this information for a corresponding initial condition for tumour simulations.

2.3 Bayesian inference

Bayesian inference is a field within statistics and a specific type of statistical inference based upon Bayesian statistical principles. In Bayesian inference, as well as its frequentist counterpart, the aim is to derive information about a population, based on data from the same population [35]. The main distinction between the frequentist and Bayesian approach is the inclusion of prior knowledge, with respect to the data at hand in the case of Bayesian statistics. The incorporation of prior knowledge is showed in Bayes' theorem, (2.2), the foundation of Bayesian statistics. Bayes' theorem presents a relation between the conditional probability of a specific parameter value, θ, given data, Y, and the probability of Y given θ:

$$\pi(\theta|Y) = \frac{L(Y|\theta) \cdot \pi(\theta)}{\pi(Y)}. \tag{2.2}$$

With this notation, θ represents the parameter of the underlying distribution. Y is the data such that $L(Y|\theta)$ is known as the likelihood of the data given the parameter θ. $\pi(\theta)$ is the prior distribution of the parameter. Here our previous knowledge of θ is contained, before any data is taken into consideration. The denominator, $\pi(Y)$, consists of the marginal likelihood and acts merely as a constant of normalisation in many cases [36]. It is also important to note the specific distinction that parameters are regarded as random variables in Bayesian inference as opposed to the frequentist approach.

2.3.1 Approximate Bayesian Computation

The central part of this study is to infer parameters governing the different patterns of tumour growth possible to observe. Parameters related to the migration and proliferation of cells are, however, not straightforward to determine since the likelihood function is unknown. We approach this problem by using the likelihood-free method Approximate Bayesian Computation (ABC). This technique has gained a reputation in the field of biology in recent years with the increase in computational capacity and is regularly used on problems where the likelihood is impossible to compute [37]. ABC is also widely used as an approximation in situations where the likelihood function is too computationally costly to evaluate [38].

The technique follows the scheme illustrated in figure 2.1. First, a set of parameter candidates, or a vector of parameters, are sampled from the prior distribution in question, $\theta_i' \sim \pi(\theta)$. For each parameter candidate, a simulation is performed to produce a result according to the specific problem, raising the requirement for a generative model to simulate data: $X_i \sim L(Y|\theta_i')$. The second step is followed by an evaluation of the result in relation to the observed data, this could, for instance, be characterised by an experiment. The evaluation is performed to investigate the agreement between simulation and observation. If $X_i = Y$ the parameter candidate

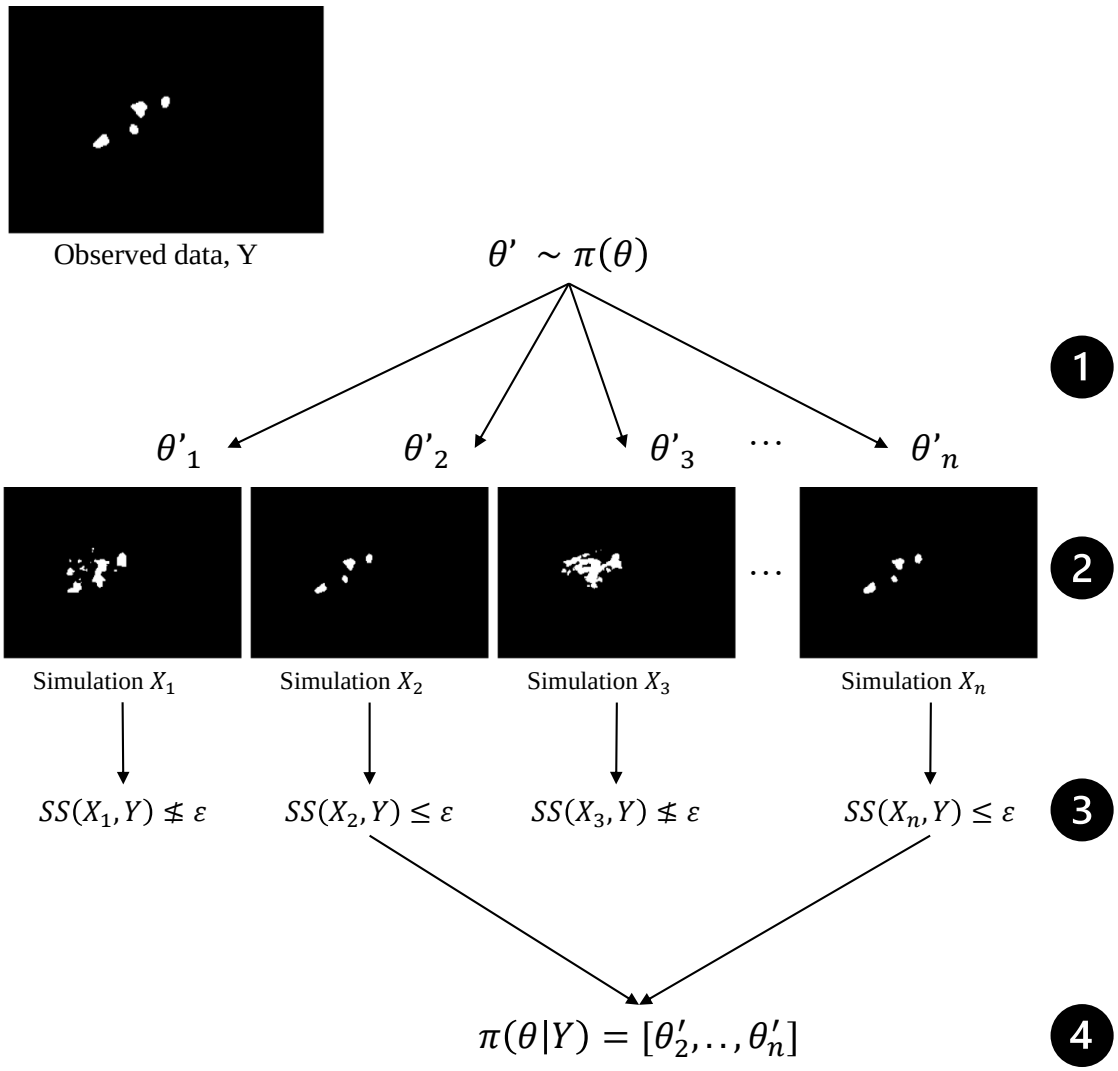


Figure 2.1: Example of the ABC-procedure, recreation of illustration from [38]. Step 1: Draw parameter candidates, θ'_i, from the prior distribution $\pi(\theta)$. Step 2: Perform a simulation with each parameter candidate. Step 3: Compute the summary statistics between simulated and observed data. Compare to the threshold ε. Step 4: The approximate posterior distribution, $\pi(\theta|Y)$, is composed of the distribution of accepted parameter candidates. Parameters above the threshold are rejected.

is regarded to belong to the posterior distribution, $\pi(\theta|Y)$, hence it is saved. However, for a continuous variable, the probability to estimate the true parameter is zero. Furthermore, an exact concurrence is not necessarily interesting and it is common to calculate a set of summary statistics that describes the observed and simulated data and compare the summary statistics instead of the complete data [39]. It is important that the summary statistics describe the data in a good way to avoid unnecessary approximations. But the choice of summary statistics is not straightforward to select and differs from task to task. The difference between summary statistics is evaluated with respect to a selected threshold, ε, where the threshold represents the tolerable approximation of the posterior distribution. The final step of the procedure is to save the parameter candidates underlying an accepted summary statistic to the approximate posterior and reject the remaining parameters, represented by stage 4 in figure 2.1.

2.3.1.1 Summary statistics

To assess the similarity between simulation results and experimental results a selection of summary statistics were determined. To capture different characteristics of the observed and simulated results five measures have been identified and was used in combination as a summary statistic in the ABC-procedure. In this study the observed and simulated results are expressed by images, hence a selection of image similarities have been chosen for the task.

The Jaccard index, also known as Intersection over Union (IoU), is often used for object detection in the area of computer vision [40], [41].

Definition 2.3.1 (Jaccard index)**.** Let I_X and I_Y denote the sets of non-zero pixels for the images X and Y, then the Jaccard index is defined as

$$J(X,Y) = 1 - \frac{|I_X \cap I_Y|}{|I_X \cup I_Y|}. \tag{2.3}$$

Note that $J(X, Y) \in [0, 1]$, where the lower boundary indicates two identical images and two disjoint images results in the maximum value of the measure. The motivation of using the Jaccard index is to evaluate the simulated tumour shape, it is however important to remember that this measure does not take the amount of displacement into consideration and will therefore not describe a tumour difference completely on its own.

Definition 2.3.2 (Mean squared error)**.** The mean squared error for two images, X and Y, with identical dimension, $[I \times J]$, is defined as

$$\mathrm{MSE}(X,Y) = \frac{1}{I \cdot J} \sum_{i}^{I} \sum_{j}^{J} (X_{i,j} - Y_{i,j})^2. \tag{2.4}$$

For binary images X and Y, $\mathrm{MSE}(X, Y) \in [0, 1]$, where the minimum value is obtained for two identical images. The mean squared error aims to describe the difference in tumour size and overlap, suggesting additional aspects of tumour representation.

Directed Hausdorff distance is a measure of dissimilarity for point sets [42]. In this study, we apply the measure for the image indices representing tumour presence to capture the displacement between different results.

Definition 2.3.3 (Directed Hausdorff distance)**.** For two, non-empty, discrete sets, I_X and I_Y, represented by non-zero image coordinates,

$$d_H(I_X, I_Y) = \max_{\mathbf{x} \in I_X} \left\{ \min_{\mathbf{y} \in I_Y} \{d(\mathbf{x}, \mathbf{y})\} \right\}. \tag{2.5}$$

Here $d(\mathbf{x}, \mathbf{y})$ is the Euclidean distance for images, defined by (2.6) for two coordinates $\mathbf{x}$ and $\mathbf{y}$.

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{2.6}$$

In the context of image comparison, the sets I_X and I_Y is the non-zero pixel positions in two images, X and Y, respectively. Furthermore, we expand the definition in the case of empty sets, a possible event for an image without tumour presence. This extension is tailor-made for an image comparison task. Comparing a non-empty image with an empty image results in the worst possible measure, the diagonal distance of the image, given by (2.7), for images with the dimension $[I \times J]$.

$$d_H(I_X, \emptyset) = d_H(\emptyset, I_Y) = \sqrt{I^2 + J^2}. \tag{2.7}$$

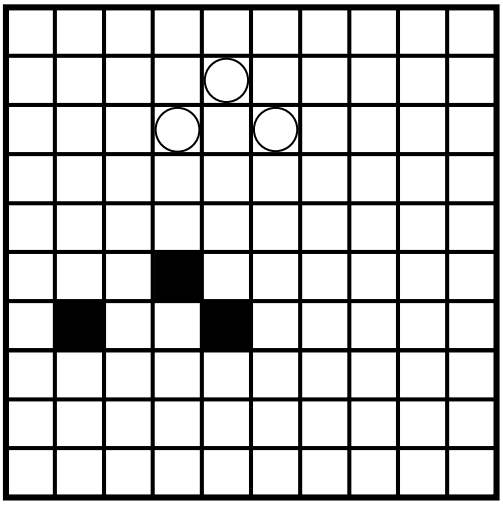
However, for two empty sets we have a perfect image agreement and we define $d_H(\emptyset, \emptyset) = 0$. Note that the directed Hausdorff distance is not symmetric, where $d_H(I_X, I_Y) \neq d_H(I_Y, I_X)$, in general. This fact and the rationale of the directed Hausdorff distance is shown in figure 2.2. The asymmetric characteristic of the Hausdorff distance gives a more nuanced description of image difference and it is often used in medical applications, for example in the evaluation of image segmentation of medical images such as MRI or CT-scans [42].

The Wasserstein distance, also known as Earth mover's distance, is originally a measure of similarity between two probability distributions with the rationale of the amount of mass needed to be moved for two distributions to become equal.

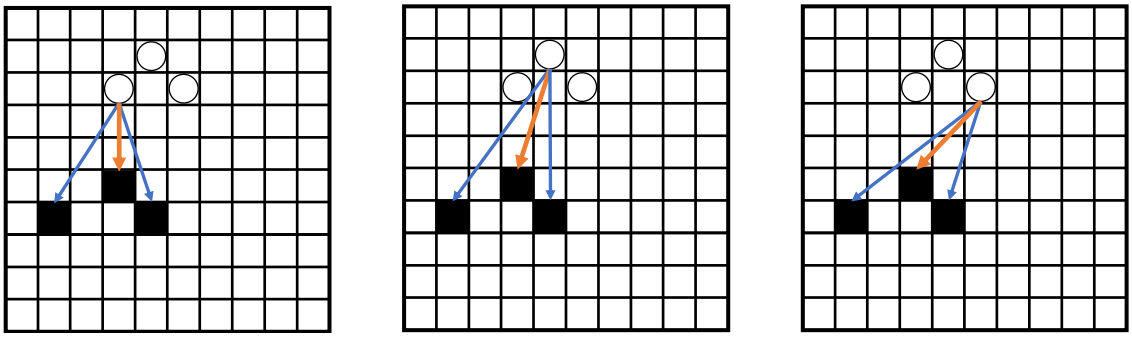
Definition 2.3.4 (Wasserstein distance)**.** For two, one-dimensional probability distributions μ and ν the Wasserstein distance is defined as:

$$W(\mu, \nu) = \inf_{P \sim \mu,\, Q \sim \nu} \left\{ E\left[\|P - Q\| \right] \right\}. \tag{2.8}$$

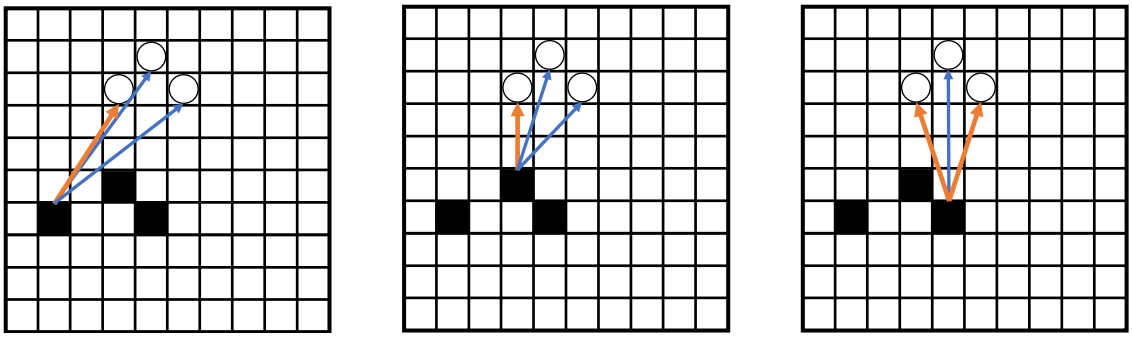
For general distributions, the measure is finite, $W(\mu, \nu) < \infty$, if $E[\|P\|] + E[\|Q\|] < \infty$, where $E[\|P\|]$ denotes the expected value of the random vector P, [43]. For the task of image comparison, we propose a windowed Wasserstein distance, denoted $WW(X, Y)$. In the windowed case, we divide the complete image into a number of sub-images and apply the Wasserstein distance to the one-dimensional distribution of pixel values in the sub-image. The total measurement is achieved by a summation of the Wasserstein distances over all sub-images.



(a) Two hypothetical sets, I_X represented by circles and I_Y represented by filled squares, placed in a fictional figure of 10×10 pixels.



(b) $d_H(I_X, I_Y) = \max_{\mathbf{x} \in I_X} \{\min_{\mathbf{y} \in I_Y} \{d(\mathbf{x}, \mathbf{y})\}\} \approx \max\{2.0, 4.12, 3.61\} = 4.12.$



(c) $d_H(I_Y, I_X) = \max_{\mathbf{y} \in I_Y} \{\min_{\mathbf{x} \in I_X} \{d(\mathbf{y}, \mathbf{x})\}\} \approx \max\{4.47, 2.0, 4.12\} = 4.47.$

Figure 2.2: Rationalization of the directed Hausdorff distance as well as an illustration of its non-symmetric properties.

3. Material and Methods

This chapter will present the material and data used in this project. This is followed by the implementation of the theory. Including a presentation of the mathematical model and the way it has been applied to produce synthetic data from which parameters have been estimated. Lastly, the methodology to estimate parameters governing growth patterns observed in experiments will be described.

3.1 Experimental tumour growth

Researchers in Sven Nelander's research group at Uppsala University has performed experiments using cell line-derived xenografts (CDX) where cells from multiple patients with diagnosed glioblastoma have been labelled with GFP-luciferase to identify human-specific antibodies. The cells were transplanted into the brain of immunodeficient mice according to a stereotactic procedure, a well-established method for precise and reproducible brain surgery [44]. The glioblastoma cells were allowed to develop during the time of 40 weeks in agreement with research interest and ethical regulations, after which the brains were sliced into five elements in the coronal plane. Each element was stained and scanned for highlighted antibodies representing tumour cell presence.

In this study, three experiments are considered and these were the foundation of the identification of cell line-specific parameters. We call the experiments "brain 1", "brain 2" and "brain 3" throughout this report, table 3.1 shows information regarding each brain. Here the initial tumour location (in Cartesian coordinates of the simulation domain), the number of transplanted cells and the coronal slices of acceptable quality are reported.

Table 3.1: Information of performed experiments and the available data consisting of scans of coronal slices (CS-scans).

Brain name	Transplantation loc.	# Cells	CS-scans
Brain 1	[93, 133, 101]	10^5	70
Brain 2	[93, 133, 101]	10^5	150, 95, 70
Brain 3	[93, 110, 101]	10^5	140, 90, 70

Following the scan of individual brain sections, a segmentation of the images was performed by the Nelander group. Here, a semantic segmentation deep convolutional neural network (CNN) architecture based on the U-Net was used to distinguish between tumour cells, normal mouse brain cells and background tissue. The resulting,

segmented, images related to the three brains are shown in figure 3.1, figure 3.2 and figure 3.3. Here the cancer cells are coloured red, the brain structure is blue and white, while the background is black. As seen in the table above, brain 1 contains one scan of a coronal slice as opposed to three coronal slices for brain 2 and -3, this is due to the quality of the experimental results where the other coronal slices have been damaged during the slicing procedure or the absence of tumour cells in distant parts of the brain.

As preparation for comparison between simulation results and experiments, the high-resolution segmented scans were rescaled to the dimensions of the simulations, from $[1908 \times 1304]$ to $[250 \times 182]$. Essentially, resulting in a percentage of cancer cells in each pixel in the reduced image. Furthermore, this image was mapped to the intensity range of the simulation, corresponding to the carrying capacity of the model.

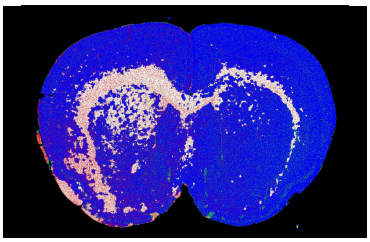


Figure 3.1: Cancer cells segmented from the xenograft scan of brain 1, coronal slice no. 90, glioblastoma cells in red.

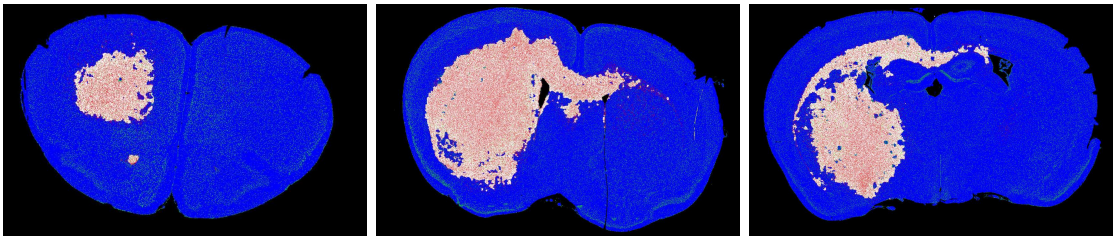


Figure 3.2: Cancer cells segmented from the xenograft scan of brain 2, coronal slice no. 70, 95 and 150 from left to right, glioblastoma cells in red.

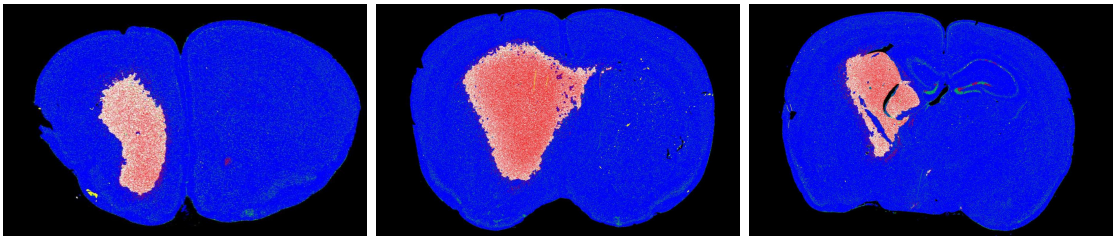
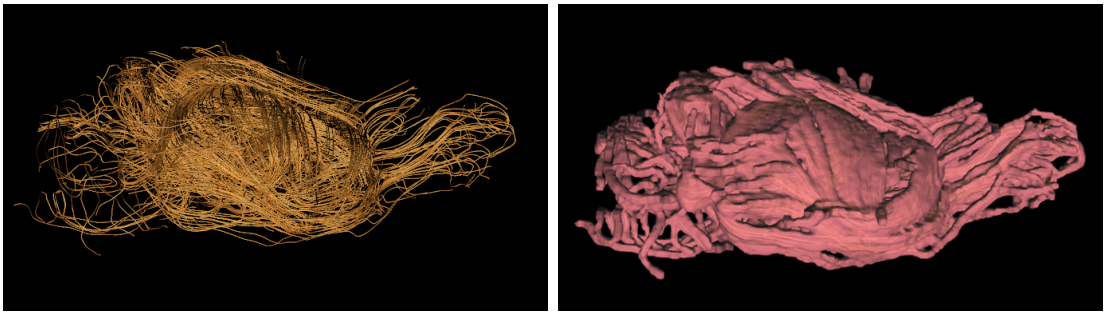


Figure 3.3: Cancer cells segmented from the xenograft scan of brain 3, coronal slice no. 70, 90 and 140 from left to right, glioblastoma cells in red.

3.2 DTI data and pre-processing

As mentioned in section 2.2.1, the DTI data reflects the diffusion of water molecules. Thus, the information of white matter tracts has to be derived from the underlying DTI data. Here researchers at Uppsala University have used an existing DTI data set from Jiang & Johnson, [15], representing the brain of a C57BL/J6 mouse as foundation. To derive the white matter fibres, DSI Studio was used, this is a tractography software tool for DTI analysis. This procedure results in the fibre position, illustrated in figure 3.4a. The thickness of the tract was determined by including a neighbourhood of one voxel to create white matter tracts seen in figure 3.4b. The result of the data processing is a three-dimensional, binary matrix where 1 indicates white matter in a specific voxel.



(a) Position of white matter tracts.

(b) White matter tracts used to guide migration in the mathematical model.

Figure 3.4: Result of the pre-processing of DTI data set which was used to direct the migration of glioblastoma cells in the mathematical model. Here viewed in the sagittal plane.

3.3 Brain vasculature and pre-processing

The open-source data, derived by Todorov *et al.* in [16], representing high-resolution brain vasculature, was attained by the Nelander group and reduced in dimension for reasonable memory consumption. The result is once more a three-dimensional, binary matrix representing blood vessel location, illustrated with a saggital view in figure 3.5.

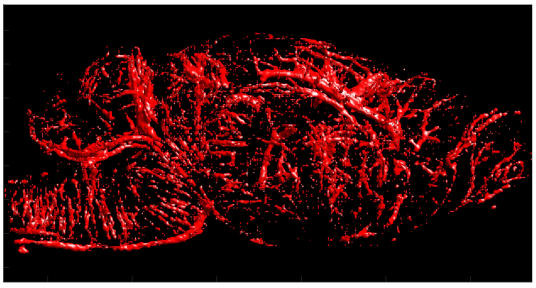


Figure 3.5: Result of the pre-processing of brain vasculature data set, viewed in the sagittal plane.

3.4 Combination of medical data sets

To achieve comparable white matter tracts and brain vasculature both representations were cropped and rotated manually, followed by a transformation to get the two brains to the same domain. The transformation was found by an iterative closest point algorithm, an algorithm suited for three-dimensional point clouds. The result is binary matrices of dimension [250, 346, 182] where voxels are equally spaced, with the volume 43^3 µm^3. The simulation domain is derived as the union of both brains, enclosing all tissue information and suggesting the space available for cancer migration in simulations.

3.5 Cell-based model for glioblastoma growth

As mentioned in the background regarding previous glioblastoma modelling, section 2.1, several models have been proposed, recently taking white matter tracts or brain vasculature into consideration. In this study, we propose a novel mathematical model for the task, a stochastic, cell-based model guided by both white matter tracts and the brain vasculature for simulations of glioblastoma propagation. We simulate the evolution of the tumour development in a three-dimensional domain representing the mouse brain.

From a selected initial condition regarding location and number of cells, cells are allowed to proliferate and migrate with probabilities dictated by decided rates. The 40 week tumour development-time is discretized in T constant time steps of duration δt. For $t \in [0, 1, \ldots, T]$ such that $t \cdot \delta t = [0, \ldots, 40\text{w}]$, we denote the total number of cells at time t by N_t . At each time step, we sample N_t cells with replacement, each sampled cell is subject to a possible proliferation and migration event. The proliferation and migration events occur with a spatially dependent probability P_p and a constant probability of migration P_m. If a proliferation event is carried through the result is one additional cell, placed in the same voxel as the parent cell. If a cell is subject to migration the result is a cell-movement in one of six neighbour voxels where the direction is guided by the surrounding microenvironment by a probability

distribution over all directions. This directional distribution governing migration is denoted D_m. The model incorporates a carrying capacity for each voxel to capture the concept of the available space in a voxel. We denote the carrying capacity K and if a voxel is inhabited by K cells any proliferation in this voxel or migration to the filled voxel is aborted. After each of the N_t cells has been the subject of proliferation and migration events the time is increased to $t+1$ and the procedure by sampling of N_{t+1} cells is repeated until the final time step. In the following two sections the incorporation of tissue properties in proliferation and migration will be described in detail.

3.5.1 Proliferation

The proliferation of a glioblastoma cell is governed by the spatially varying probability, $P_p(\mathbf{x})$, denoting the probability of a proliferation event in voxel $\mathbf{x}$. In detail, we determine the total probability of a proliferation event as the sum of a base probability, base_p, representing the natural cell cycle of glioblastoma cells and an additive term, $\text{BV}_p \cdot \text{BV}_{pres}$, representing an increased probability of proliferation if blood vessels are present in the surroundings of a cell,

$$P_p = \text{base}_p + \text{BV}_p \cdot \text{BV}_{pres}. \tag{3.1}$$

The term BV_{pres} denotes the percentage of voxels where there exist blood vessels in a sphere around the current voxel. For our discrete simulation domain, we define the vicinity as the set of coordinates in a closed ball with radius r, centred in $\mathbf{x}$ as $B_r[\mathbf{x}] = \{\mathbf{y} \in \mathbb{R}^3 : \|\mathbf{x}-\mathbf{y}\| \leq r\}$, hence we determine the percentage of blood vessels by,

$$\text{BV}_{pres} = \frac{\sum_{\mathbf{y}\in B_r[\mathbf{x}]} \text{BV}(\mathbf{y})}{|B_r[\mathbf{x}]|}. \tag{3.2}$$

Specifically for the proliferation event, we select the radius $r = 3$ voxels acting as a distance where cancer cells are affected by growth factors such as signalling proteins and oxygen resulting in an increased probability to model the stimulation of proliferation activity. The notation $|B_r[\mathbf{x}]|$ represents the total number of voxels in the discretized sphere.

3.5.2 Migration

The migration direction is governed by a probability distribution over all directions in the three-dimensional von Neumann neighbourhood of each voxel, $P_m(\mathbf{x}) \in \mathbb{R}^{1\times 6}$. Similarly to the proliferation probability, the migration direction is dependent on the brain tissue, characterised by an influence from blood vessels and white matter tracts in the vicinity of the voxel.

$$D_m = (1 - \text{BV}_w - \text{WM}_w) \cdot D_h + \text{BV}_w \cdot D_{\text{BV}} + \text{WM}_w \cdot D_{\text{WM}} \tag{3.3}$$

Here the terms represent homogeneous migration, migration in direction of blood vessels and migration in direction of white matter respectively. The parameters BV_w and WM_w represent the migration preference for blood vessels and white matter,

with values in the range $[0, 1]$, under the additional constraint, $\mathrm{BV}_w + \mathrm{WM}_w \leq 1$. Ensuring that the probability distribution, D_m, sums to one. The residual migration is governed by homogeneous migration, equal probability of migration in each direction, $D_h = \frac{1}{6} \cdot [1,\ 1,\ 1,\ 1,\ 1,\ 1]$.

Probability distributions specifically governing blood vessels and white matter, D_{BV} and D_{WM}, is determined from a weighted centre of mass, shown for white matter in (3.4). Here the presence of blood vessels or white matter further away from the voxel gives a smaller contribution due to the decreasing concentration of nutrients further away from the source. The weighted centre of mass is illustrated by a two-dimensional example in figure 3.6, here fictional presence of white matter is highlighted by blue pixels. Vectors from the centre of the sphere to white matter presence are yellow with a width to indicate the weight which it contributes to the resultant. The centre of mass is the sum of all individual contributions, displayed by a red vector in the illustration. The weighted centre of mass is normalized by the component sum such that $\mathrm{WM}_{dir}(\mathbf{x})$ represent the proportion of mass in each direction, with a sign that indicate the positive or negative direction. Finally the probability distribution over neighbouring directions, $D_{\mathrm{WM}(\mathbf{x})}$ is formed by assigning a probability of movement, or zero depending on the sign of the corresponding component in $\mathrm{WM}_{dir}(\mathbf{x})$.

$$
\begin{aligned}
\mathrm{COM}_{\mathrm{WM}}(\mathbf{x}) &= \sum_{\mathbf{y} \in B_r[\mathbf{x}]} \mathrm{WM}(\mathbf{y}) \cdot \frac{(\mathbf{y} - \mathbf{x})}{\|\mathbf{y} - \mathbf{x}\|} \cdot (1 + r - \|\mathbf{y} - \mathbf{x}\|) \\
\mathrm{WM}_{dir}(\mathbf{x}) &= \frac{\mathrm{COM}_{\mathrm{WM}}(\mathbf{x})}{\sum_i |\mathrm{COM}_{\mathrm{WM}}(\mathbf{x})_i|} \\
D_{\mathrm{WM}}(\mathbf{x}) &= [p_x^+,\ p_x^-,\ p_y^+,\ p_y^-,\ p_z^+,\ p_z^-] \\
p_i^+ &= \max\{0, \mathrm{WM}_{dir}(\mathbf{x})_i\} \\
p_i^- &= \max\{0, -\mathrm{WM}_{dir}(\mathbf{x})_i\}
\end{aligned}
\tag{3.4}
$$

Again, we have $B_r[\mathbf{x}]$ as the set of coordinates in a closed ball with radius r, centred in $\mathbf{x}$, $B_r[\mathbf{x}] = \{\mathbf{y} \in \mathbb{R}^3 : \|\mathbf{x} - \mathbf{y}\| \leq r\}$. Furthermore, $\mathrm{WM}(\mathbf{x})$ represents the DTI data which is characterised as a matrix of the brain dimension with a binary indication of white matter presence or not, described in section 3.2. The probability distribution of migration influenced by blood vessels is derived analogously, with DTI data replaced by brain vasculature described in section 3.3. There is however a difference in the size of the local neighbourhood. In the case of white matter, the radius of the closed ball is set to 3 voxels, in contrast to the case of brain vasculature where the radius is 10 voxels based on an increased effect from the secretion of nutrients from blood vessels. The difference in radius is a result of an assumption with support in the type of interaction with the microenvironment. For blood vessels, cells are affected by chemical substances with an additional reach, compared to the attraction from white matter which is believed to be characterised primarily by mechanical concepts.

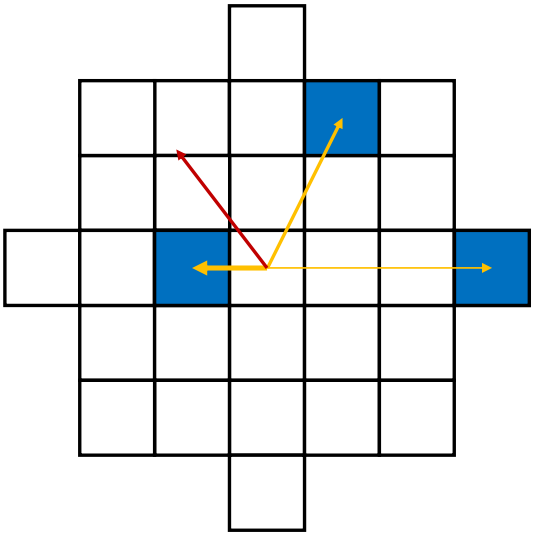


Figure 3.6: Two-dimensional illustration of weighted centre of mass, fictional white matter presence denoted by blue pixels. The individual contributions to the centre of mass is expressed by yellow vectors where the thickness indicates the corresponding weight. The resultant of all individual contributions, the weighted centre of mass is displayed by a red vector.

3.6 Generation of synthetic data

To investigate the consistency of the model, i.e. to investigate the possibility to estimate known parameters of synthetic data, the data needs to be generated. Hence, the mathematical model was used with parameters fixed according to values in table 3.2. We will use two cases of migration to investigate different tumour characteristics. Namely, heterogeneous migration where cells are set to have a preference in migration along blood vessels and white matter tracts. Furthermore, we will investigate the model consistency in the case of homogeneous migration, corresponding to an environment of isotropic tissue.

Table 3.2: Parameters used for generation of synthetic data with heterogeneous migration (Case 1) and homogeneous migration (Case 2).

Parameter	Case 1	Case 2
Time step, δ_t	6 [h]	6 [h]
Simulation time, T	40 [w]	40 [w]
Carrying capacity, K	3 [cells]	3 [cells]
Initial number of cells, N_0	1536 [cells]	1536 [cells]
Base proliferation, base_p	0.00325	0.00325
Increased proliferation, BV_p	$3 \cdot \text{base}_p$	$3 \cdot \text{base}_p$
Migration probability, P_m	0.4	0.4
Migration preference for BV, BV_w	0.51	0.0
Migration preference for WM, WM_w	0.228	0.0

Here the values were chosen with respect to computational accuracy and previous findings of glioblastoma growth. Starting from the top of the table, the time step, δt was chosen for a sufficient migration velocity. The total simulation time, T, was straightforward to determine as 40 weeks, although the synthetic data have no connection to the experimental data it is clear that we want to test the model consistency on the same time frame as it will be used with later of, to generate simulations over the 40 week period. The carrying capacity, K, was set to 3 cells per voxel. In relation to the size of a cell in relation to the voxel size, it would admittedly be appropriate with a higher value. However, the accessible space for cancer cells are a complex question varying throughout the brain dependent of the extent of the extracellular matrix. It is also possible to regard one cell in the simulation as corresponding to multiple cells in reality. The initial condition was selected as a cube with a width of eight voxels, in each voxel three cells were placed resulting in 1536 cells in total.

Continuing, both *in vivo* and *in vitro* studies of glioblastoma cells have been performed with respect to the proliferation and migration rate of the cells. Proliferation rates have been derived from tumour doubling time, for example by an *in vitro* study by Stensjøen *et al.* where the median doubling time in 106 patients was 49.6 days [45], resulting in a proliferation rate for individual cells of $r_p = 0.014\,\mathrm{d}^{-1}$. Derived from exponential cell growth,

$$\begin{aligned} N_t &= N_0 \cdot \exp(r \cdot t) \\ t = 49.6 &\Rightarrow N_t = 2 \cdot N_0 \\ \Rightarrow 2N_0 &= N_0 \cdot \exp(r \cdot 49.6) \\ \Rightarrow \ln(2) &= r \cdot 49.6 \Rightarrow r \approx 0.014. \end{aligned} \tag{3.5}$$

Similarly, Swanson *et al.* used a doubling time of 2 months in their mathematical model of tumour growth [19], resulting in a proliferation rate for individual cells of $r_p = 0.012\,\mathrm{d}^{-1}$. Consequently, the baseline proliferation probability was derived from a rate of 0.013 d^{-1}, the mean of the previous results. With the general approximation of probability from the rate of an event, (3.6), we find the probability, base_p, corresponding to the selected proliferation rate, to be 0.00325.

$$p \approx r \cdot \frac{\delta t}{24} \tag{3.6}$$

Furthermore, we decide the coefficient for an increased proliferation in presence of blood vessels to be three times as large as the baseline, $\mathrm{BV}_p = 3 \cdot \mathrm{base}_p$.

Proceeding to the parameters governing the migration of glioblastoma cells we note a large disparity in previous research with studies reporting the mean cell velocity in the range of 34 µm/d [13] up to 500 µm/d [46] both found in experimental settings. For the synthetic data, we select a migration rate, $r_m = 1.6\,\mathrm{d}^{-1}$, concluding in a migration velocity of 64µm/d, as a result of the constant voxel-width, $\Delta x = 43\,\mu\mathrm{m}$,

$$\begin{aligned} v &= r_m \cdot \Delta x \\ v &= 1.6 \cdot 43 = 68.8\,\mu\mathrm{m/d}. \end{aligned} \tag{3.7}$$

Furthermore, the probability of a migration event, $P_m = 0.4$, derived by (3.6).

Lastly, we have the two weight parameters describing the preference in favour of blood vessels, BV_w, or white matter, WM_w. In Case 2 the coefficients were fixed to zero, resulting in all migration covered by homogeneous movement. In Case 1, we resort to the previous results by Esmaeili *et al.* where 38% of tumour growth direction was aligned with white matter tracts [11]. Furthermore, Cuddapah *et al.* report that 85 % of glioma cells move into contact with blood vessels [8]. To also include some homogeneous migration in the data each coefficient was multiplied with a factor, 0.6, resulting in the parameters, $\mathrm{BV}_w = 0.51$ and $\mathrm{WM}_w = 0.228$.

With the parameters fixed according to table 3.3, the mathematical model described in section 3.5 was used to generate 100 realizations in each of the two cases. For every simulation, one slice of the simulation domain was saved, corresponding to the data of a coronal slice in the experimental results. One realization from each case is shown in figure 3.7, here the impact of anisotropic tissue, characterised by directed migration, is distinctly expressed.

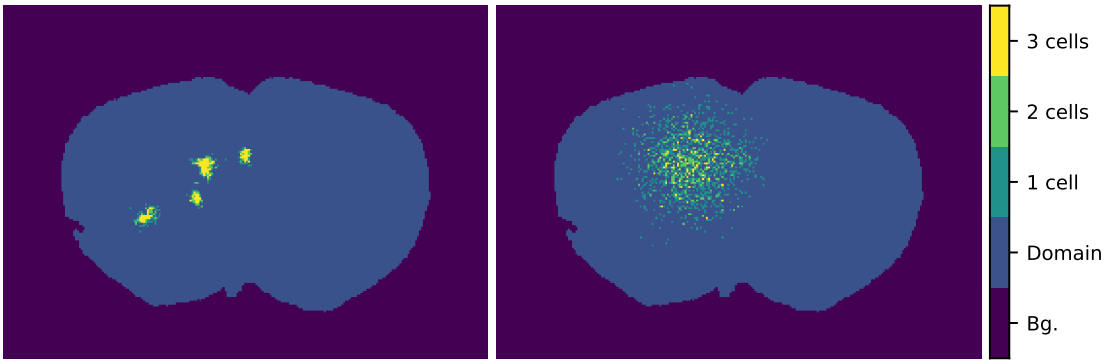


Figure 3.7: Realizations of the model with fixed parameters. Case 1 (left) and Case 2 (right). Each figure shows a coronal slice where the pixel intensity describes the cancer cell density. Furthermore, the brain domain is accentuated from be distinguished to the background (Bg.).

3.7 Threshold derivation, synthetic data

The derivation of a threshold for the acceptance of parameters in advance of the ABC procedure was based on 100 realizations with fixed parameters described in the previous section. All realizations were combined into one representation and summary statistics were evaluated between individual simulations and the combined representation. This was made with the intent to get an idea of the stochasticity in the model, i.e. express the maximum detectable difference in simulations with fixed parameters. Here two techniques were derived, first a direct comparison where each simulation was evaluated with respect to the synthetic data. This technique proved to be insufficient. Hence, a second method, a three-levelled approach was adopted.

Here the simulations and synthetic data were divided into three images based on the cell occupancy in each voxel. In the following sections, both procedures of threshold derivation will be described in additional detail.

3.7.1 Direct comparison

First, all realizations were combined into one representation by a pixel average after a Gaussian filter, with standard deviation $\sigma = 1$ and kernel size $[10 \times 10]$ pixels, had been applied. The Gaussian filter is justified since we are not interested to detect cancer development with pixel precision, but rather to capture the general characteristics of the growth pattern. The average representation was binarized with thresholding at pixel intensity 0.5, resulting in a single representation corresponding to the available data from experiments. The average representations for Case 1 and -2 of fixed parameter combinations are shown in figure 3.8.

Figure 3.8: Combined, average, representations of tumour growth with fixed parameters. Case 1 (left) and Case 2 (right). Each figure shows a binary representation of tumour cells in a coronal slice.

To find a threshold that accepts the true parameter value, all 100 realizations were evaluated in direct comparison to the average representation. If we denote individual simulations X_i and the combined representation Y, a vector of summary statistics were determined according to definition 2.3.1-definition 2.3.4, resulting in a vector with five elements, (3.8),

$$SS(X_i, Y) = \begin{bmatrix} J(X_i, Y) & \text{MSE}(X_i, Y) & d_H(I_{X_i}, I_Y) & d_H(I_Y, I_{X_i}) & WW(X_i, Y) \end{bmatrix}. \tag{3.8}$$

For each comparison we receive a vector of measures that have been achieved with the correct parameter value, these observations were used to derive two types of thresholds. First, we define $\boldsymbol{\varepsilon}_{mean}$ to be a vector of thresholds with elements determined by the mean of each observed measure, i.e. the threshold for the Jaccard index and the first element in $\boldsymbol{\varepsilon}_{mean}$ is determined by,

$$\frac{1}{100} \sum_{i=1}^{100} J(X_i, Y). \tag{3.9}$$

Furthermore, we derive a second, more relaxed threshold denoted ε_{max} as the maximum of each observed measure. For the Jaccard index the threshold is determined by,

$$\max_{i\in[1,100]}\{J(X_i, Y)\}. \tag{3.10}$$

Note that thresholds for the other measures of tumour similarity are derived analogously.

3.7.2 Three-levelled comparison

To assess the weakness of the direct comparison a new way of processing simulated results was developed, substantially consisting of a refined comparison with respect to the direct comparison. Instead of a single thresholding after the application of the Gaussian filter, the cell occupancy in each pixel was taken into consideration. Based on the number of cells in the cross-sectional image the tumour presence was divided into three images representing pixels with one, two or three cells.

To combine all model realizations into three representations a Gaussian filter with standard deviation $\sigma = 0.5$ and filter size 3×3 pixels was applied to each simulation. After which, the simulation was divided according to pixel intensity, $X_i^{(2)}$ is an image where the pixel intensity of the filtered image is greater than zero and less than or equal to one, $X_i^{(2)}$ originates from pixel intensities in the range $(1, 2]$ and $X_i^{(3)}$ from the range $(2, 3]$. Each combined image was determined with a pixel-median and a binarization to result in three images. The three image levels corresponding to heterogeneous and homogeneous migration of fixed parameters is presented in figure 3.9 and figure 3.10.

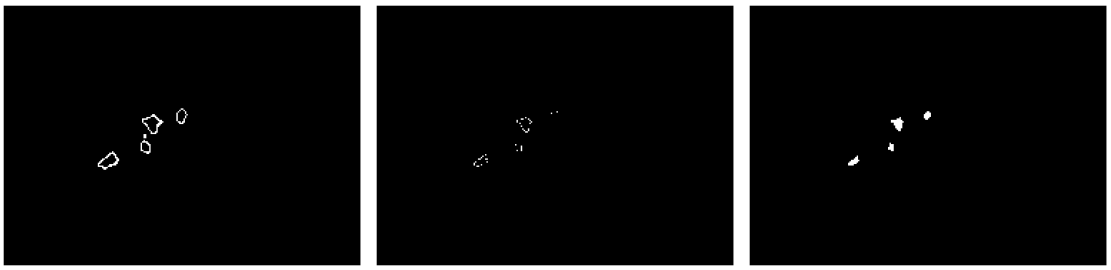


Figure 3.9: Combined, median, representations of tumour growth with fixed parameters (Case 1). Intensity level 1 (left), level 2 (centre) and level 3 (right).

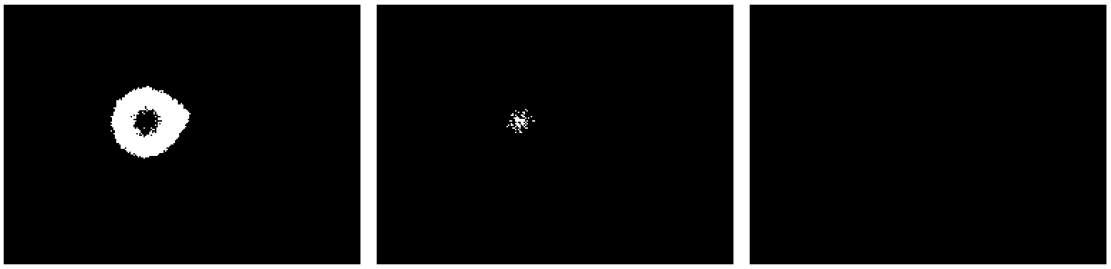


Figure 3.10: Combined, median, representations of tumour growth with fixed parameters (Case 2). Intensity level 1 (left), level 2 (centre) and level 3 (right).

Once again the full range of summary statistics, determined according to definition 2.3.1-definition 2.3.4, was used to derive thresholds for the ABC procedure. In this case the measurements are for every image level, $X_i^{(1)}$, $X_i^{(2)}$ and $X_i^{(1)}$, with the corresponding notation for median combined representation, $Y^{(1)}$, $Y^{(2)}$ and $Y^{(3)}$. Here the threshold is denoted $\boldsymbol{\varepsilon}_{mean}^{(j)}$, for a general image level j, derived according to (3.9) and $\boldsymbol{\varepsilon}_{max}^{(j)}$, derived according to (3.10).

It was found that the threshold derived by mean value, $\boldsymbol{\varepsilon}_{mean}$, was very restrictive in comparison to $\boldsymbol{\varepsilon}_{max}$. A well-established approach in Approximate Bayesian Computation is to include the best percentage of simulations in the posterior distribution, particularly when there is high uncertainty in the selection of thresholds. This approach was adopted in this study by calculating the minimal relaxation of a threshold needed for a simulation to be accepted,

$$r_i \text{ s.t. } SS(X_i^{(j)}, Y^{(j)}) \leq r_i \cdot \boldsymbol{\varepsilon}^{(j)}, \ \forall j \in [1, 2, 3]. \tag{3.11}$$

Throughout the report and the ABC procedure the vector inequality is defined as

$$\mathbf{x} \leq \mathbf{y} \text{ if } x_k \leq y_k \quad \forall k \in [1, \ldots, K], \tag{3.12}$$

for two vectors $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^{1\times K}$. In (3.11), a low value of necessary relaxation, r_i, indicate a summary statistic close to the threshold. Hence, simulations were sorted in ascending order according to the relaxation coefficient and the posterior distribution was formed by the parameter candidates with relaxation coefficient within the best percentage of all simulations. In section 4.1, a parameter estimation with $\boldsymbol{\varepsilon}_{max}^{[1,2,3]}$ is compared to an estimation conducted from the minimal relaxed mean threshold.

3.8 Parameter identification, synthetic data

In the parameter identification of synthetic data one or two parameters were treated as unknown and sampled from a uniform distribution. In the following sections, the prior selection and simulations for the ABC procedure will be presented.

3.8.1 Prior selection

To estimate known parameters the prior distributions were selected as a so-called flat prior, a uniform distribution where no specific prior knowledge is contained, except for the limits of the interval outside which it is impossible to sample a parameter candidate. The interval of the uniform distribution was chosen as $\left[\frac{1}{5} \cdot x,\ 5 \cdot x\right]$, where x represents a general model parameter. This selection was made for all parameters or the combination of parameters subject to identification. The residual parameters remained fixed according to the values given in table 3.2.

3.8.2 Simulations and estimated parameters

By following the ABC-rejection procedure parameter candidates were sampled from the described prior distribution. With the specific parameter candidate/s, a simulation was performed and the result was compared to the single representation of synthetic data. Comparisons were performed both with the direct comparison and the three-levelled approach. Parameters giving a summary statistic below the corresponding threshold was saved in the approximated posterior distribution. To get a sufficient number of accepted parameters in the posterior distribution different number of simulations was performed depending on the parameter to be estimated. The different parameters subject of estimation and the certain number of simulations are presented in presented in table 3.3. The posterior was determined both with the method of direct comparison and the three-levelled approach where max, mean, best percentage in respect to the smallest necessary relaxation of the mean threshold was used. Additionally, an investigation of the difference in using one summary statistic in comparison to all five measures in combination. All results are shown in either section 4.1 or appendix A.

Table 3.3: Parameters at subject of estimation and number of performed simulations, the remaining parameters are fixed according to the reference simulation.

Parameter/s	# Simulations	Reference simulation
Increased proliferation, BV_p	5455	Case 1
Migration preference for BV, BV_w	7347	Case 1
Migration preference for WM, WM_w	2838	Case 1
Base rates: r_p & r_m	6055	Case 1
Base rates: r_p & r_m	11 497	Case 2
Migration preference: BV_w & WM_w	11 574	Case 1

3.9 Parameter identification, experimental data

The method of parameter identification of experimental data is to a large extent similar to the estimation of known parameters presented in the previous section. The difference, however, is the non-existent threshold for accepted parameters. Here a different strategy has to be adopted.

3.9.1 Threshold derivation

Due to the shortcomings found in the direct comparison, we exclusively use the three-levelled approach in estimations of experimental parameters. With the determined threshold from synthetic data as a foundation, it was noted that regardless of heterogeneous or homogeneous migration the threshold was found to be of the same magnitude. Suggesting the random variation of the model rather than the variation corresponding to a different parameter choice. With some marginal, the threshold was chosen from the maximum values found in either of the two cases of migration. Resulting in a selection of threshold values for experimental estimation, given in table 3.4. We denote this threshold $\boldsymbol{\varepsilon}_{exp}^{[1,2,3]}$.

Table 3.4: Threshold derived from the deviation in synthetic data with fixed parameters, calculated in advance of ABC procedure.

Threshold	J	MSE	d_H	WW
$\boldsymbol{\varepsilon}_{exp}^{1}$	0.8	0.025	[50, 5]	0.5
$\boldsymbol{\varepsilon}_{exp}^{2}$	1	0.01	[60, 10]	0.6
$\boldsymbol{\varepsilon}_{exp}^{3}$	1	0.005	[60, 10]	0.2

3.9.2 Prior selection, fixation of parameters

In the estimation of cell line-specific parameters, it was decided to estimate four parameters. Namely, the rates governing proliferation and migration, r_p and r_m in combination with the migration preferences, BV_w and WM_w. The remaining parameters was fixed during the simulations. However, parameters such as the time step and the limits of the prior distribution are not straightforward to decide. This problem was assessed by a preliminary examination with homogeneous migration which was performed with different time steps to determine the necessary resolution to capture the diffuse behaviour shown by the experiments. The parameter selection and the prior limits are presented in table 3.5. Further details of the preliminary examination are presented in appendix B. Note that the increased proliferation in presence of blood vessels is fixed to the proliferation probability, hence, it follows the randomly sampled proliferation rate from simulation to simulation.

Table 3.5: Parameters and prior range used for generation of simulations in the ABC procedure to estimate cell line-specific tumour characteristics.

Parameter	Brain 1	Brain 2	Brain 3
Time step, δ_t	2 [h]	3 [h]	6 [h]
Simulation time, T	40 [w]	40 [w]	40 [w]
Carrying capacity, K	3 [cells]	3 [cells]	3 [cells]
Initial number of cells, N_0	51 231 [cells]	51 231 [cells]	51 231 [cells]
Proliferation rate, r_p	[0, 0.025] [d^{-1}]	[0, 0.02] [d^{-1}]	[0, 0.04] [d^{-1}]
Increased proliferation, BV_p	$base_p$	$base_p$	$base_p$
Migration rate, r_m	[0, 12] [d^{-1}]	[0, 8] [d^{-1}]	[0, 4] [d^{-1}]
Migration preference for BV, BV_w	[0, 1]	[0, 1]	[0, 1]
Migration preference for WM, WM_w	[0, 1]	[0, 1]	[0, 1]

3.9.3 Simulations and acceptance

The initial number of cells was placed in a sphere of radius 16 voxels, centred according to the coordinates from experimental studies given in table 3.1. In each voxel, three cells were placed resulting in 51 231 cells. The total number of cells was selected in relation to the carrying capacity, in section 3.6 it was introduced that the actual carrying capacity may be higher. With this in mind, it is possible to regard one cell in the simulation domain corresponding to multiple cells in reality. Hence, the initial condition is aimed to be comparable to the initial tumour in the experimental study.

Following the ABC procedure, with the introduced prior distributions and fixed parameters according to table 3.5, the model was used to generate simulations that were compared to experimental results with the three-levelled approach. In total 3200 simulations were performed to estimate the parameters governing the tumour characteristics of brain 1, 3400 for brain 2 and 4000 simulations for brain 3. Acceptance was determined by the simulations with the minimal necessary relaxation of $\varepsilon_{exp}^{[1,2,3]}$, again the best percentage was included in the posterior distribution. Since brain 2 and -3 contain multiple coronal slices to be compared to simulations the relaxation was derived such that all coronal slices had to be accepted, still resulting in one relaxation coefficient for every simulation.

4. Results

The results of the ABC procedure is mainly characterised by the posterior distribution of accepted parameters. In the following chapter, the accepted parameters of different cases of parameter estimations will be presented. First, results from parameter estimations of synthetic data are presented, the case when the true parameter combinations are known. This is followed by the results of parameter estimation from three experimental results. Note that complementary results of synthetic parameter estimations can be found in appendix A.

4.1 Model consistency

The investigation of the model consistency was, as mentioned previously, done to investigate the possibility to estimate known parameters with the ABC methodology. This was assessed by estimation of one and two parameters, with different techniques of image comparison in the derivation of summary statistics.

In figure 4.1, the posterior distribution of a single estimated parameter, BV_p, is shown in the form of a histogram with the true parameter value highlighted by the dashed line. Here the direct image comparison technique was used with two types of thresholds, to the left the posterior derived with ε_{max} is displayed which predictably accepts a larger number of parameters compared to the more restrictive ε_{mean}, shown to the right. The benefit of the more restrictive threshold is the increased accuracy of the estimation. In table 4.1, the mean and standard deviation of the posterior distribution is shown with outliers removed. Presenting a close estimation of the known true parameter value, $\mathrm{BV}_p = 0.00975$.

Continuing to the two-parameter estimation of base rates with homogeneous migration, we compare the two techniques of image comparison. In figure 4.2 the posterior distributions of proliferation- and migration rate is presented. The figure to the left is the result of the direct approach. Here it is possible to see the shortcomings of the direct approach, a compact, high-density, corresponding to high proliferation rate and low migration is regarded to be similar to a diffuse tumour with high migration

Table 4.1: Parameters and prior range used for generation of simulations in the ABC procedure to estimate cell line-specific parameters.

Threshold	# Accepted parameters	Mean	Standard deviation
ε_{max}	719	0.009864	0.002029
ε_{mean}	104	0.009675	0.001245

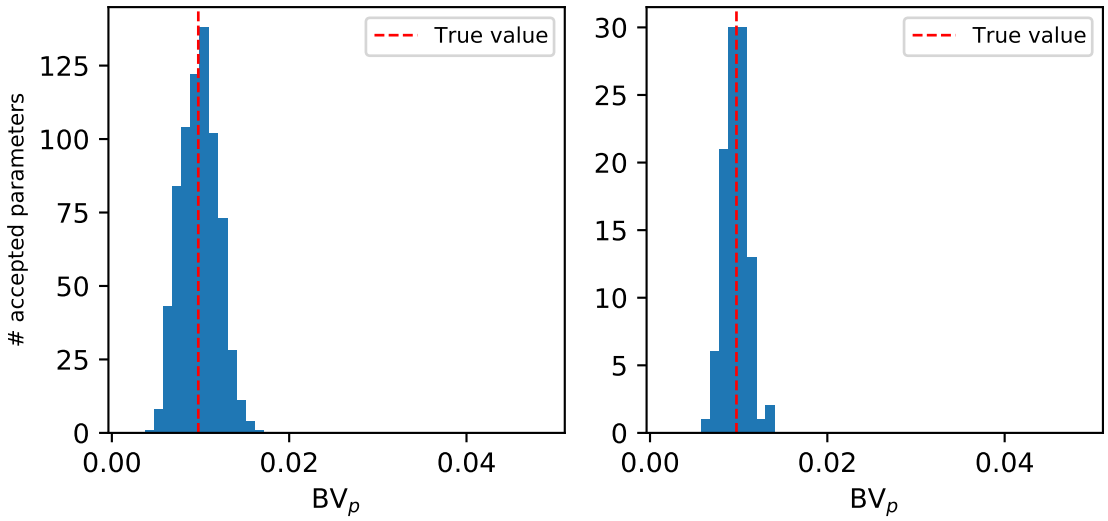


Figure 4.1: Histogram over posterior distributions with different thresholds, ε_{max} (left) and ε_{mean} (right). Direct comparison is used in both estimations.

and low proliferation. In the right of figure 4.2 the same estimation is presented, now with the three-levelled image comparison technique. The improvement in the estimation is achieved by taking the cell density into consideration. Note that the maximum threshold is used in both cases of estimation.

Remaining at the estimation of base rates with homogeneous migration, figure 4.3 illustrates the effect of multiple measures in the evaluation of image similarity. Here the posterior distribution of r_p and r_m determined with the three-levelled comparison technique is shown as a box plot. The posterior distribution of the parameters with the full range of summary statistics, $SS(X_i, Y)$ is compared to the distributions found when the measurements, $J(X_i, Y)$, MSE(X_i, Y), $d_H(X_i, Y)$, $d_H(Y, X_i)$, $WW(X_i, Y)$ are used individually. The combined use of summary statistics results in the intersection of all other posterior distributions, naturally resulting in the combined summary statistics being the most restrictive evaluation method. For this specific, case of estimation of r_p and r_m, it is also possible to see how different individual measures over- and underestimating the true parameter values, resulting in the combined procedure to produce the overall closest estimation of the true parameter values.

As mentioned in the methods section, often in the ABC procedure there is no apparent way to decide the threshold for experimental parameter estimation, here an approach with the best percentage of the simulations is used to determine the posterior distribution. This method was investigated on synthetic data and the result of estimation of migration preference is shown in figure 4.4. To the left, the posterior distribution of BV$_w$ and WM$_w$ is presented, derived from a three-levelled comparison with maximum threshold, $\varepsilon_{max}^{[1,2,3]}$. To the right, it is possible to see the posterior constituting of the simulations with the least required relaxation of the mean threshold, $\varepsilon_{mean}^{[1,2,3]}$. Here the best percentage is selected, resulting in 115 parameters in the

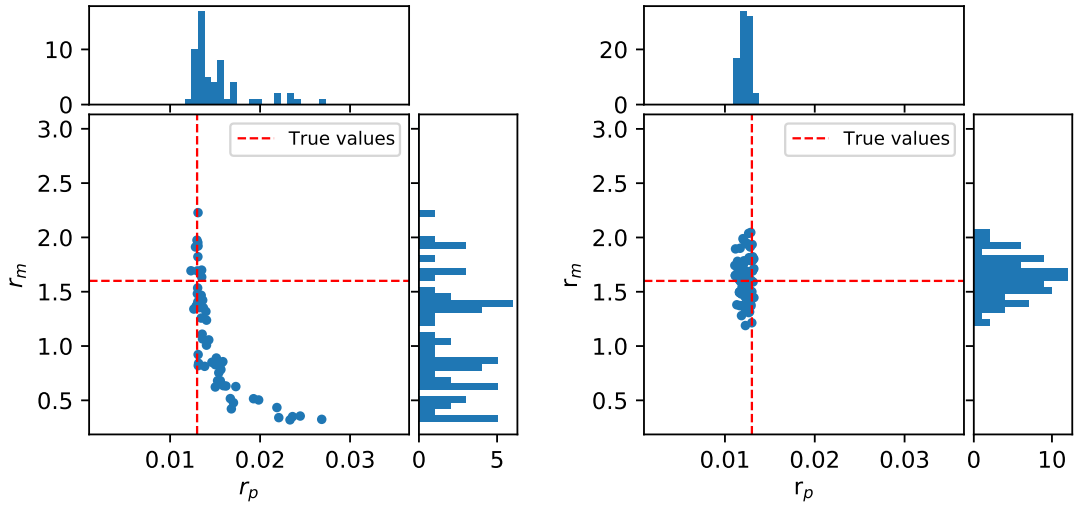


Figure 4.2: Scatter plot and histogram over posterior distributions with different image comparison techniques, direct comparison (left) and three-levelled approach (right). The parameters subject of estimation is the proliferation- and migration rates with homogeneous migration.

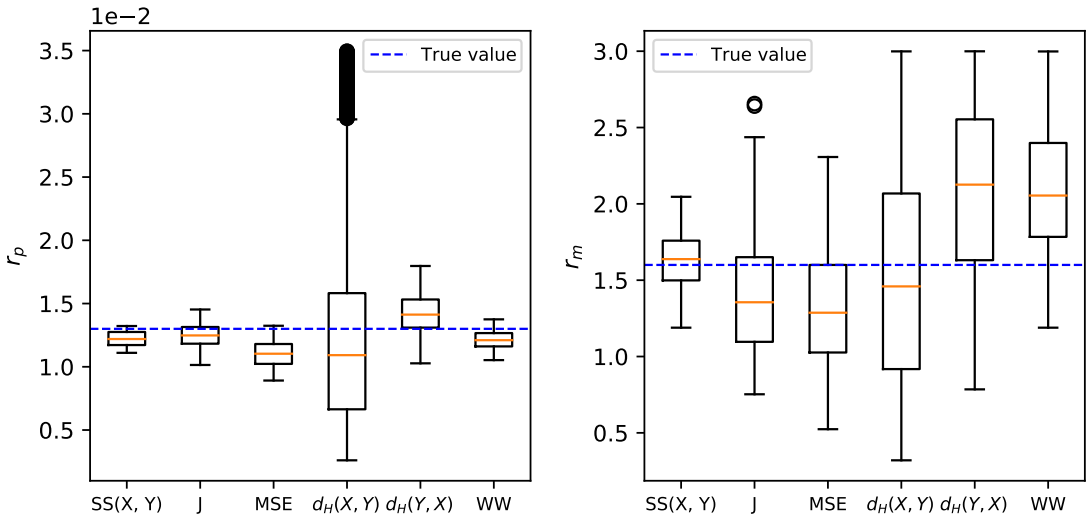


Figure 4.3: Comparison of parameter estimation with summary statistics individually and all five measures used in combination, represented as a box plot with outliers represented by circles. Here the three-levelled approach is used to estimate proliferation- and migration rates in the case of homogeneous migration.

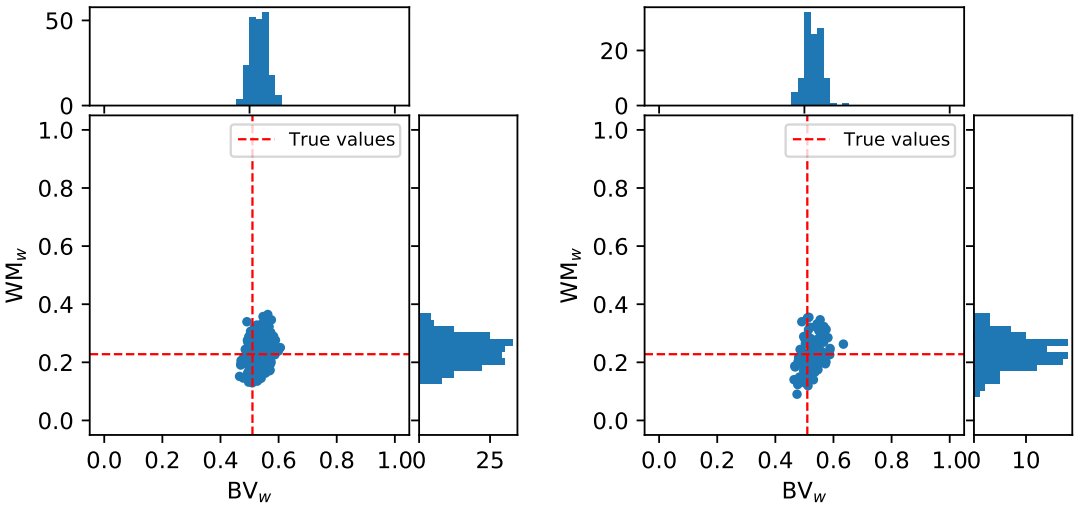


Figure 4.4: Scatter plot and histogram over posterior distributions in the investigation of migration preference, three-levelled approach. Max threshold (left) and least required relaxation of mean threshold (right).

posterior distribution.

Up to this point, the ABC procedure with the suggested generative model and the three-levelled approach of image comparison has proven successful in the task to estimate known parameters of the synthetic data. In figure 4.5 the estimation of base rates, r_p and r_m, under heterogeneous migration is shown. The results have been derived with the approach of three-levelled comparison and maximum threshold, $\boldsymbol{\varepsilon}_{max}^{[1,2,3]}$. In this illustration, we see that the method finds difficulty in the estimation of the migration rate. Furthermore, a similar difficulty is found also for a case of a more restrictive threshold. To the right in figure 4.5, 60 simulations with a reduced threshold are shown. Hence it is possible to rule out a high permittance as the reason for an imprecise estimation. Possible reasons for the behaviour of the estimation will be approached in the forthcoming discussion section.

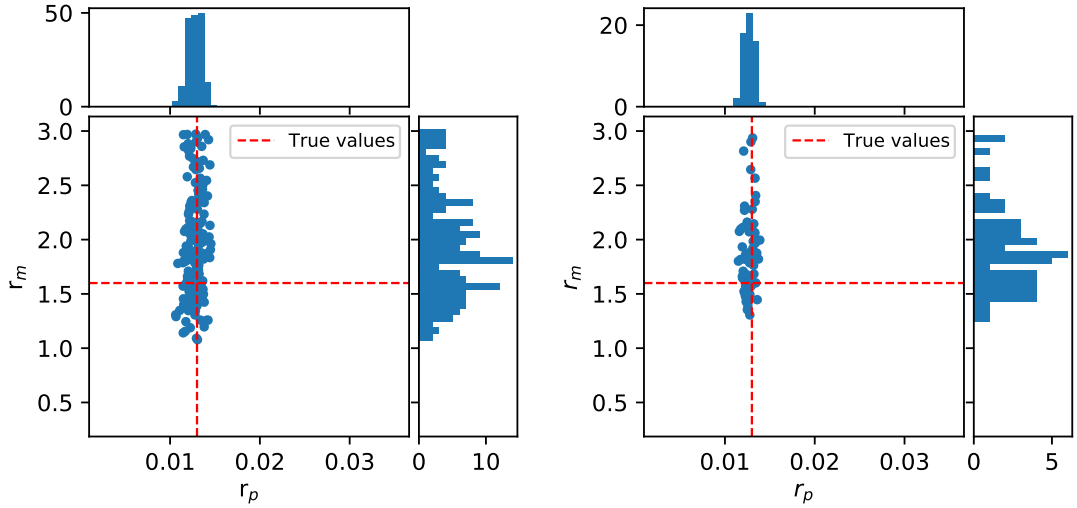


Figure 4.5: Scatter plot and histogram over posterior distributions of proliferation- and migration rates. Results derived with three-levelled comparison method. Maximum threshold (left) and reduced maximum threshold accepting 60 parameters (right).

4.2 Cell line-specific parameter identification

In the estimation of parameters from experimental results, we present the four-dimensional posterior distribution as a grid of histograms for each parameter and all combinations of marginal distributions. In the following section, the results related to brain 1, -2 and -3 is shown.

4.2.1 Brain 1

Here the best percentage of all simulations is included in the posterior distribution, resulting in 32 parameters for this estimation. In figure 4.6, histograms of each parameter are shown along the diagonal of the grid, in upper- and lower triangles of the grid scatter plots of all marginal distributions are shown. Note that the grid of the posterior distribution is symmetric in the sense that all marginal distributions are shown twice, in mirrored representations.

The most notable parameter estimation is observed in the migration preference for blood vessels, it is clear that BV_w needs to be limited for the simulations to be similar to the experimental data. The same is seen for the preference for white matter, this is reasonable in relation to the experimental results of brain 1, shown in figure 3.1, where the tumour has a low density and to a large extent, uniform spread in the brain.

In the marginal distribution of the proliferation- and migration rate it is possible to observe a low negative correlation, where a high proliferation rate in combination

with a lower migration rate is resulting in a similar tumour as a simulation with high migration and a lower proliferation. Furthermore, a moderate positive correlation is noticed in the marginal distribution of proliferation rate and white matter preference, i.e. for a simulation with a high proliferation rate an increased proliferation preference for white matter tracts were required. In the other combinations of parameters, no noteworthy correlation was discovered.

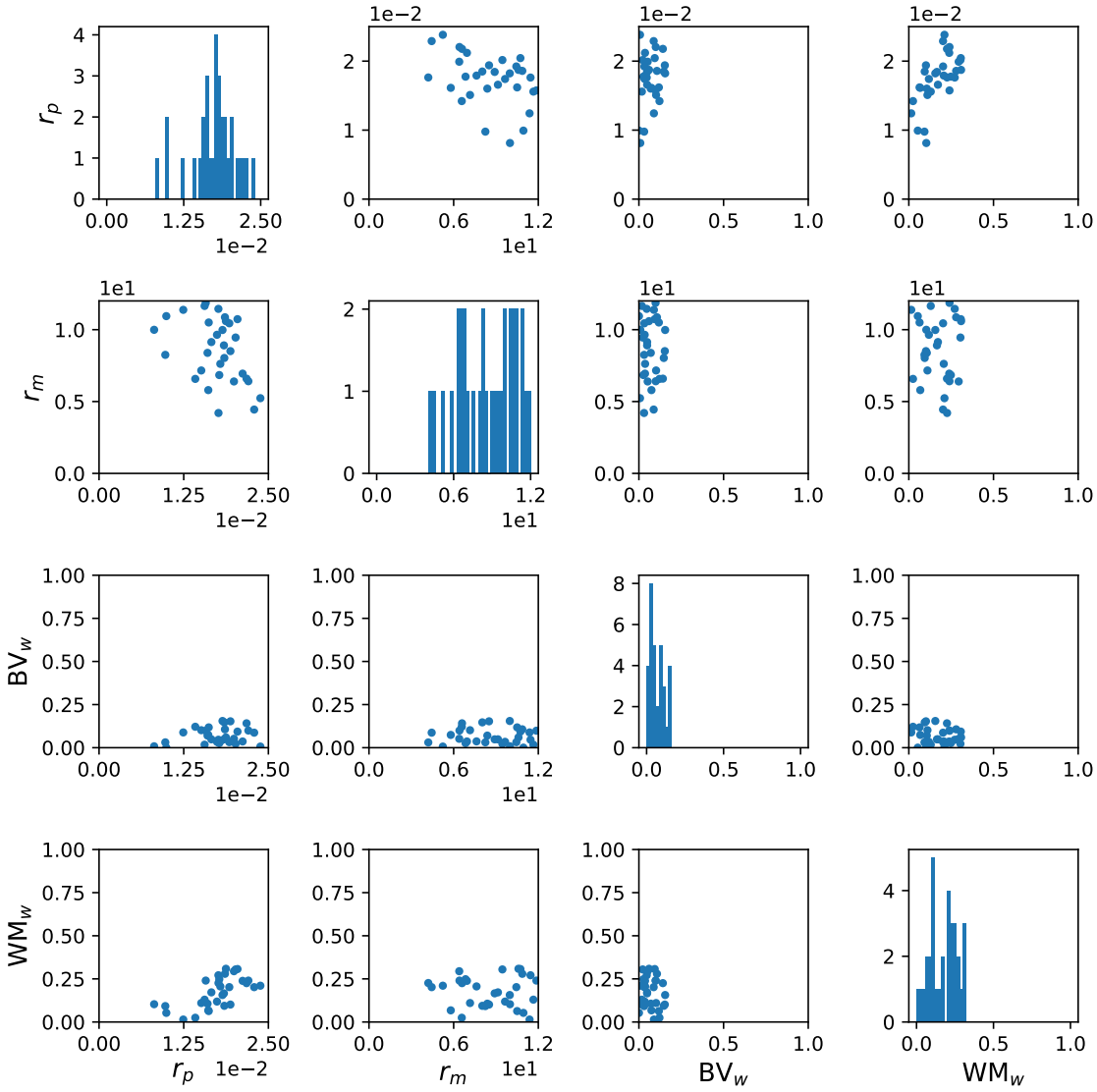


Figure 4.6: Cell line-specific parameter identification of brain 1, the four-dimensional posterior distribution is represented by histograms and marginal distributions of all parameter combinations.

In the left of figure 4.7 the experimental result of brain 1, coronal slice no. 90, is shown in its rescaled representation which has been mapped into the pixel intensity range [0, 3]. To the right, we have the corresponding best simulation derived by the least necessary relaxation of $\boldsymbol{\varepsilon}_{exp}^{[1,2,3]}$. Here the simulation is given by the parameter

combination: $r_p = 1e^{-2}$ day^{-1}, $r_m = 10.9$ day^{-1}, $BV_w = 4e^{-4}$ and $WM_w = 0.05$. The low preference for migration directed by surrounding tissue is evident in the result where cancer cells have spread close to uniformly. The migration speed in the posterior emerge from the upper region of the prior distribution resulting in a tumour that has spread largely through the brain. However, the simulated tumour fails to fully recreate the tumour reach observed in the experimental study. The proliferation rate is limited and approximately equal to the optimal proliferation rate found in the preliminary examination, shown in appendix B, producing a tumour with low density.

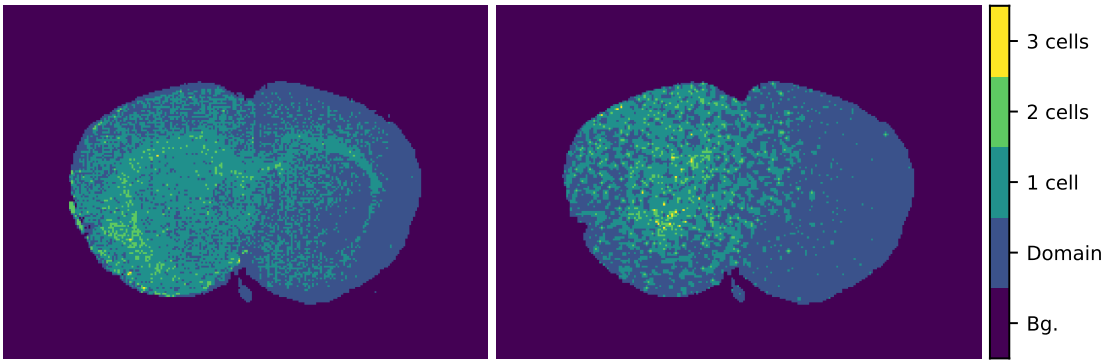


Figure 4.7: Experimental result, CS no. 90 rescaled and mapped to the intensity range of simulations (left). Simulation, CS no. 90, derived with the least required relaxation of $\boldsymbol{\varepsilon}_{exp}^{[1,2,3]}$ among all performed simulations in the estimation of brain 1 (right). Each figure shows a coronal slice where the pixel intensity describes the cancer cell density. Furthermore, the brain domain is accentuated to be distinguished from the background (Bg.).

4.2.2 Brain 2

In figure 4.8 the posterior distribution consist of the simulations with a relaxation coefficient belonging to the best percentage of all simulations. For brain 2, this results in 34 simulations, the figure contains histograms for each parameter as well as marginal distributions of all parameter combinations that are shown.

The parameter estimation shows an indication of parameter identification, most evident for the rate of migration which is required to correspond to a high migration velocity. It is also possible to determine a low preference for blood vessels as seen in the histogram regarding the parameter BV_w. The estimation of the proliferation rate is a slightly skewed distribution centred on 0.015 day^{-1}, a result similar to what was observed in the preliminary examination of brain 2, with homogeneous migration. From figure 4.8 it is evident that there is a certain demand for migration along white matter fibres. This seems promising since the well-known white matter structure called *corpus callosum* is prominent in the experimental results, given by the horizontal bundle along which cancer cells have migrated in figure 4.9. However, in figure 4.10, we observe a different structure of white matter where the cancer cells

have migrated along. The possible reasons for this are covered in section 5.2.

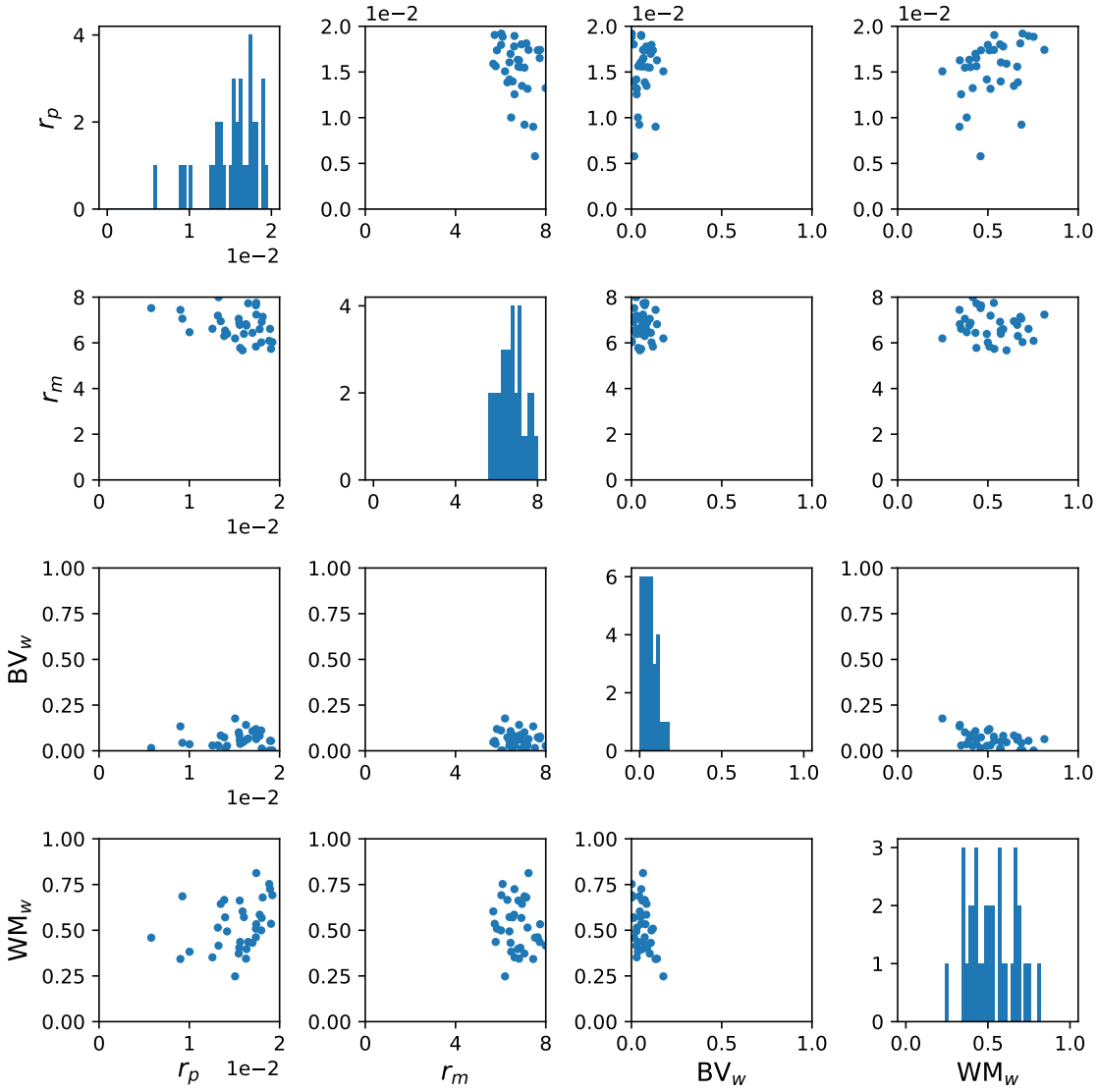


Figure 4.8: Cell line-specific parameter identification of brain 2, the four-dimensional posterior distribution is represented by histograms and marginal distributions of all parameter combinations.

Returning to the marginal distributions in figure 4.8, we detect a low negative correlation between proliferation and migration rates, similar to what was detected in figure 4.6. A low negative correlation was also observed between migration preference for blood vessels and white matter suggesting similar tumour patterns for tumour growth driven by either preference, but not both simultaneously. Lastly, there is a low positive correlation between proliferation rate and white matter preference, suggesting a relationship between the parameters where an increased proliferation rate is connected to an increased preference for white matter structures for the simulation to be included in the posterior distribution. For the remaining parameter combinations, no significant correlation was observed.

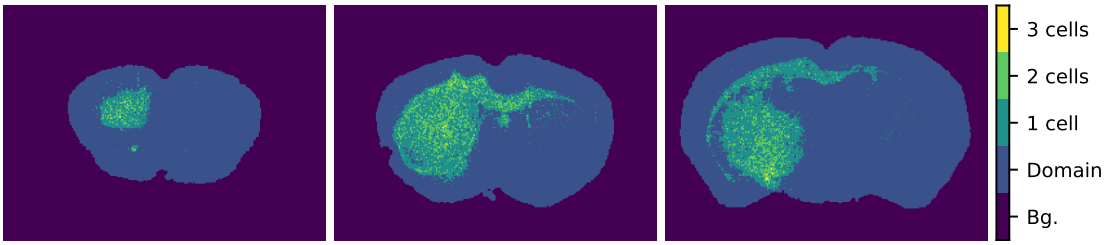


Figure 4.9: Experimental result, CS no. 70, 95 and 150 from left to right, the images have been rescaled and mapped to the intensity range of simulations. Each figure shows a coronal slice where the pixel intensity describes the cancer cell density. Furthermore, the brain domain is accentuated to be distinguished from the background (Bg.).

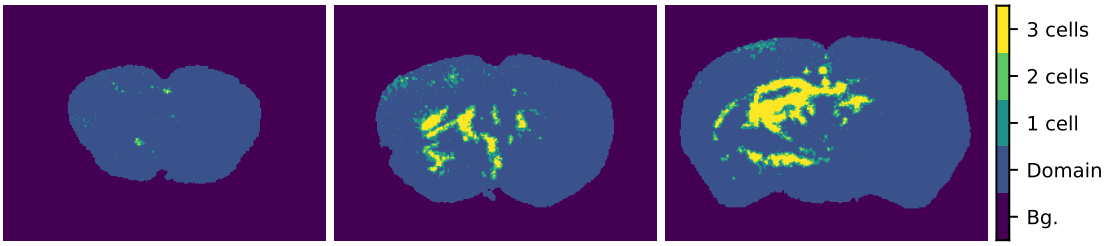


Figure 4.10: Best simulation result among all performed simulations in the estimation of brain 2, CS no. 70, 95 and 150 from left to right, derived with the least required relaxation of $\varepsilon_{exp}^{[1,2,3]}$.

4.2.3 Brain 3

To estimate the cell line-specific parameters governing the growth pattern observed in brain 3, 4000 simulations were performed. Hence, there are 40 parameter combinations in the posterior distribution illustrated by individual parameter histograms and marginal distributions for all parameter combinations in figure 4.11. Again, it is possible to observe evidence for successful parameter estimation. The experimental result of brain 3 exhibits the most compact tumour, this is captured by the estimation where the rate of migration is significantly lower compared to parameter estimations of brain 1 and -2. Furthermore, similar results are observed in the estimation of migration preference for blood vessels. This parameter has been observed to be of limited magnitude in all parameter identifications. In brain 3 the migration preference for white matter varies widely over the posterior distribution, with a high positive correlation to the proliferation rate. There is also a low observed positive correlation between WM_w and r_m suggesting an increased need for proliferation along white matter tracts to keep the tumour compact at increased rates of proliferation and migration.

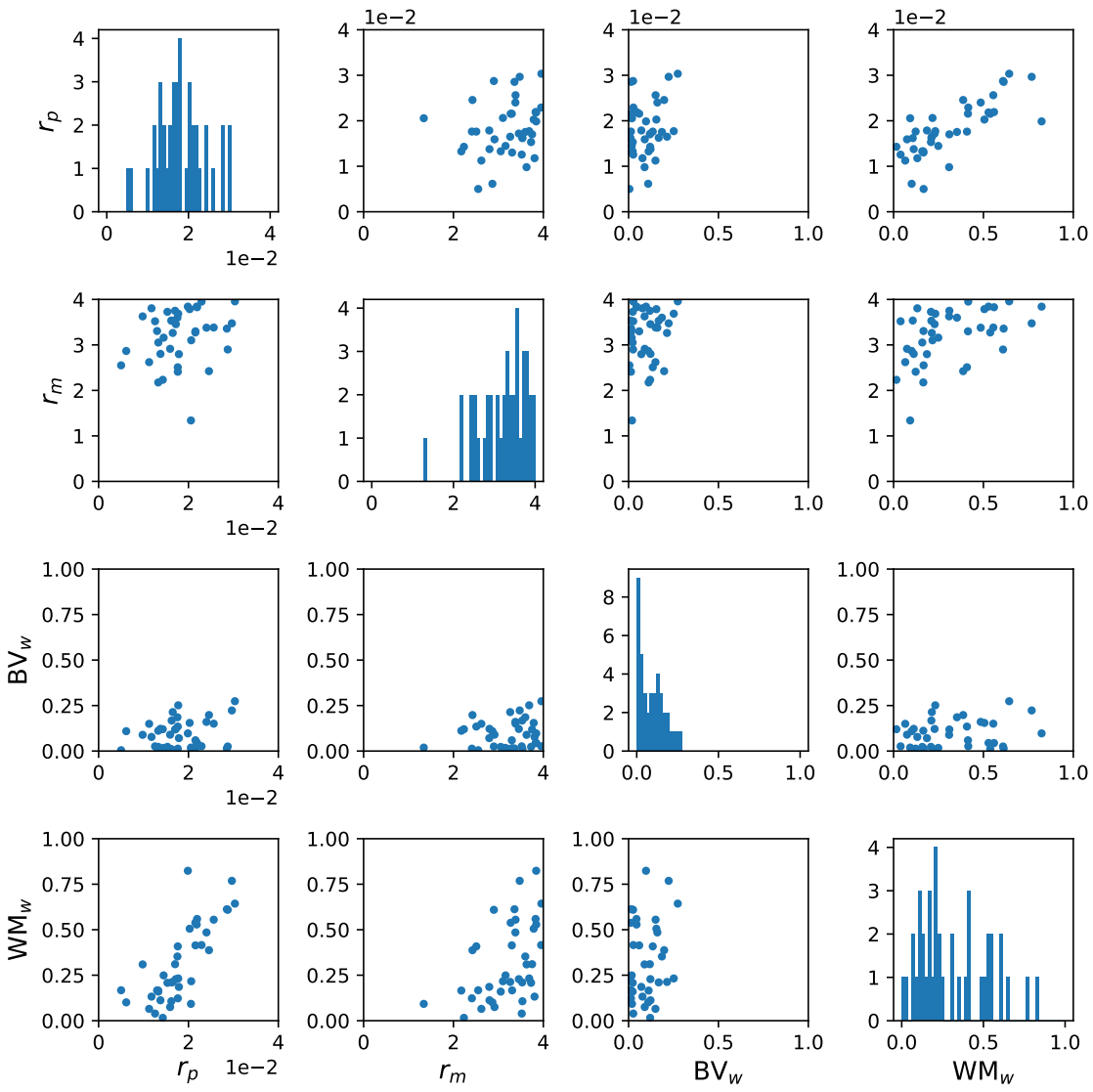


Figure 4.11: Cell line-specific parameter identification of brain 3, the four-dimensional posterior distribution is represented by histograms and marginal distributions of all parameter combinations.

In figure 4.12 the rescaled and mapped representation of the experimental results, brain 3 are presented. Here all three coronal slices are shown where the pixel intensity describes the number of cells. In total, 4000 simulations have been compared to the experimental result and the simulation with the smallest required relaxation of $\varepsilon_{exp}^{[1,2,3]}$ is presented in figure 4.13. As mentioned in the previous paragraph the posterior seems to capture the higher tumour density exhibited by brain 3, this is also shown for the best simulation where especially coronal slice no. 90 displays a high cancer cell density. However, the tumour boundary is sharper in the experimental case, suggesting increased migration preference for some tissue structure where the best estimation exhibits a migration close to homogeneous, with parameters $\mathrm{BV}_w = 0.03$ and $\mathrm{WM}_w = 0.04$.

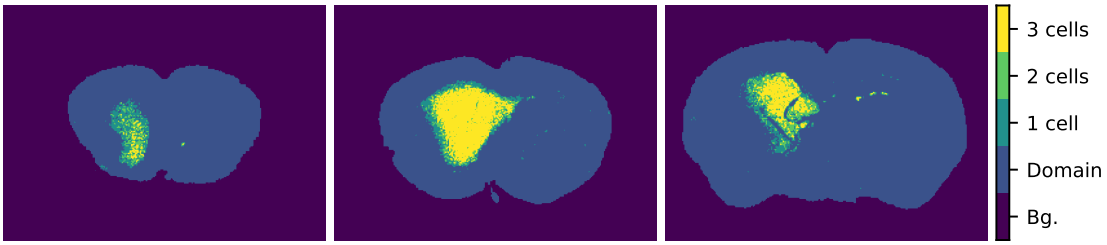


Figure 4.12: Experimental result, CS no. 70, 90 and 140 from left to right, the images have been rescaled and mapped to the intensity range of simulations. Each figure shows a coronal slice where the pixel intensity describes the cancer cell density. Furthermore, the brain domain is accentuated to be distinguished from the background (Bg.).

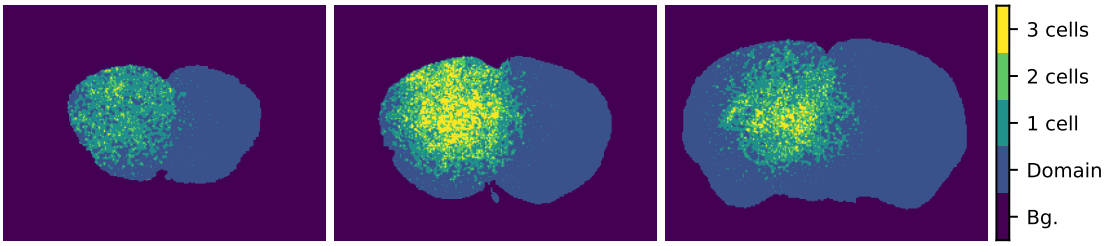


Figure 4.13: Best simulation result among all performed simulations in the estimation of brain 3, CS no. 70, 90 and 140 from left to right, derived with the least required relaxation of $\varepsilon_{exp}^{[1,2,3]}$.

5. Discussion

In this chapter, the results will be discussed with respect to the aim of the study. Particularly, the performance of the method in terms of model consistency, advantages of the current implementation and suggestions for future development will be considered.

5.1 Model consistency

The initial investigations of model consistency, by estimation of one parameter individually, with the approach to compare images directly, seemed promising with relatively precise parameter estimation of both BV_p, figure 4.1, and BV_w. The estimation of the migration preference for white matter, WM_w, was less precise but still successful by being able to estimate the parameter by mean value closely. The large variance in estimation of WM_w could be related to the parameter value of BV_w, leading to the tumour shape to be dominated by the preference for blood vessels, with a parameter value fixed to 0.51. Hence, a small change in parameter value has a limited impact on the final tumour shape. In contrast to the preference for blood vessels where a small change in a parameter value results in simulations that are rejected based on the similarity measures. This assumption is supported by comparable values of relaxation coefficient needed in single parameter estimation, mentioned in appendix A. This information, gained from the synthetic parameter estimation describes the characteristics of simulations with respect to the parameter value. By extension, this aids in the interpretation of the parameter estimation of experimental data.

However, when parameters were estimated pairwise, it was noted that the direct comparison technique had too severe drawbacks to be used for accurate parameter estimation. In the estimation of proliferation- and migration rates we could see that vastly different simulations were accepted. As briefly suggested previously in the report, the direct image approach disregarded the cancer cell density by the thresholding approach resulting in the loss of crucial information. As an attempt to solve this problem the three-levelled approach was developed. There was an evident improvement, seen in figure 4.2, where the new approach managed to estimate the rate of proliferation and migration properly. The effect of the findings was that exclusively the three-levelled approach was used in the estimation of parameters from experimental results.

Addressing the aim to investigate different measurements describing tumour similarity, we turn to the box plot of different posterior distributions showed in figure 4.3. Here we see the improvement of using all measures in combinations, rather than us-

ing one measure individually. In the estimation of both proliferation and migration rate the individual measures over- respectively underestimates the true parameter value. The combination of all measures results in a good balance between different aspects of tumour similarity. The presented case, estimation of proliferation and migration rate in isotropic tissue, was the most evident result. In some cases of parameter estimation the Jaccard index, mean squared error and the windowed Wasserstein distance performed at a similar level to the combined measures. This was the case for single parameter estimations of BV_p, BV_w and WM_w.

In the estimation of parameters originating from synthetic data with the three-levelled image comparison technique and the mean threshold, $\boldsymbol{\varepsilon}_{mean}^{[1,2,3]}$, it was noted that the parameter acceptance was very limited. One reason could be the high bias in the model. Although the model is stochastic, biases are introduced by directed migration and increased proliferation in the vicinity of blood vessels. This results in realizations of the model having a relatively low variance. Hence, the mean threshold is restrictive by nature, rejecting also many simulations with the correct parameter value. An additional difficulty arises in the transition from one to two parameters to be estimated, progressing from a one-dimensional parameter space to two dimensions requires a squared number of simulations to achieve the same sampling density. The simulations are computationally costly to execute and it would arguably be necessary to achieve a higher sampling density to accept a sufficient number of parameters with the mean threshold.

The low acceptance rate led to the implementation of the method where the least necessary relaxation, r_i, was computed as an alternative to the mean threshold. In figure 4.4, the estimated posterior distribution of migration preference with the max threshold is compared to the posterior distribution derived with the least relaxation of the mean threshold. The posterior distributions show similar success in the task to estimate the parameter combination, suggesting approval of the alternative method. This means that there is no evident problem to combine the vector of summary statistics into one representation. In fact, the relationship between the summary statistics is included since the method is based on the values of the mean threshold. Further support for the method of least required relaxation is presented in appendix A, where the least required relaxation results in a more restrictive acceptance and the simulations closer to the true parameter value are among the best percentage.

The three-levelled approach of image comparison is not without disadvantages, in figure 4.5 the estimation of the base rates in anisotropic tissue shows a difficulty to determine the rate of migration accurately. The hypothesis for this weakness is related to the implementation of the model. Namely the fact that migration is aborted if the potential voxel is occupied by the maximum possible number of cells. In the case of heterogeneous migration, the bias in the model is high and cells move, to a large extent, in the same direction. This was found also in the derivation of thresholds, where the variance of simulations with fixed, identical, parameters was found to be limited. If all cells move similarly the carrying capacity is soon reached

and many migration events will be aborted according to the construction of the mathematical model. Hence, it is possible that a simulation does not exhibit the assigned rate, instead, it is possible that a lower effective rate governs the simulated growth pattern. Therefore, simulations with different parameter values may result in a similar effective migration rate and therefore also similar tumour results. It is important to remember this inaccuracy when it comes to the estimation of experimental parameters.

However, according to the hypothesis, this problem occurs in cases of compact tumours, for a low-density tumour the carrying capacity is not reached to the same extent with many voxels below the full capacity. Therefore it is expected that there is less difference between the effective and assigned migration rate. This argument is supported by the posterior distribution of migration rate in the case of homogeneous migration, figure 4.2 (left), where the estimation is improved compared to the compact tumour found in the case of heterogeneous migration.

5.2 Cell line-specific parameter estimation

In the following sections, the results regarding the characterization of the three experimental results will be discussed with respect to the parameter identification and the best simulation outcome introduced in section 4.2.

5.2.1 Brain 1

From figure 4.6, it is possible to conclude an indication of parameter identification for all parameters regarded in experimental estimations. Although the prior distribution for r_m was chosen restrictively, it is possible to see a limitation in the parameter value governing the rate of proliferation. Furthermore, there is a notable restriction in the migration preference for the tissue structures considered. Specifically for brain 1, it is possible to conclude a tumour growth governed by homogeneous migration, to a large extent. This is something possible to conclude directly from the experimental result. It is however asserting that the ABC procedure with the developed mathematical model succeeds to reach this conclusion.

Unfortunately, the variance in the results is quite high and the best percentage of simulations exhibits large visual difference. The best simulation, shown in the right of figure 4.7, and a few other simulations with the lowest required relaxation show reasonable visual conformity. This is, however, not the case for all simulations accepted in the posterior distribution. Simulations that require a larger relaxation for acceptance show less visual similarity to the experimental result. The result is nonetheless promising, we can conclude the success in the method to use the approach of least required relaxation to find simulation results with the best visual comparison. However, a higher sampling density is needed for greater accuracy in the parameter estimation across all simulations of the best percentage. Furthermore, we have seen in the generation of synthetic data that the stochasticity in the model is low where simulations with the same parameter combination exhibit results with

high similarity. This suggests that the results of a parameter estimation could be as good as the best simulation shown in figure 4.7, if a sufficient sampling density is reached.

5.2.2 Brain 2

In the parameter estimation of the parameters governing the cell lines from brain 2, one of the main sources of error becomes evident. The experimental results show a high preference for the well-known white matter structure *corpus callosum*, recognizable in figure 4.9 as a horizontal wave in the coronal plane where the cancer cells have migrated along. This white matter bundle is however not present in the available medical data. For example, in coronal slice 90 shown in the centre of figure 5.1 the specific horizontal structure is entirely absent. Badea *et al.* reports a low difference of *corpus callosum* between different individuals [28]. Hence the absence of *corpus callosum* in the medical data is unlikely to originate from the fact that different individuals are regarded in data and experiments. Instead, it is possibly caused by the transformation to take the brain vasculature and white matter tracts to the same spatial domain.

Moreover, coronal slice 150 in figure 5.1 show some indication of the known structure with a horizontal white matter tract in the expected region of the brain. However, many, expectedly smaller, white matter structures are equally present in this coronal slice. This may relate to the method of tract-derivation where an equal thickness is assigned to all detected white matter presence above a certain threshold. There are multiple ways to derive white matter tracts and the approximation done in this study was essentially found to reject too much information and unfortunately influence the possibility to estimate the growth pattern observed in the second experiment.

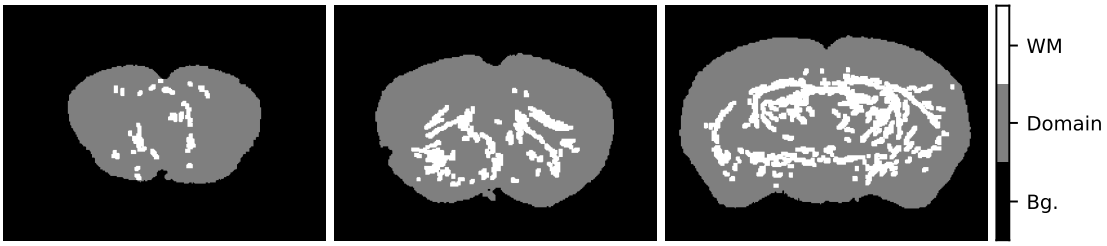


Figure 5.1: Representation of white matter tracts in a coronal view, CS no. 70, 95 and 150 from left to right.

The absence of known structures in the white matter data is not suspected to be the sole reason for the high visual difference between experiments and simulations. In the histogram representation of parameter estimation, figure 4.8, it is possible to observe a requirement of a migration rate at the very top of the prior distribution. This shows evidence for a too restrictive prior distribution. However, the posterior distribution for the migration rate corresponds to the probability of an event in the range $[0, 1]$, hence the time step is required to be reduced for an extended prior

range. This problem is also shown in figure 4.10 where the simulation have been performed with a migration rate $r_m = 6.9\,\mathrm{d}^{-1}$, regardless, resulting in a tumour with high density compared to the experimental result. Hence there is evidence for insufficient time resolution in the estimation of the growth patterns observed in brain 2.

5.2.3 Brain 3

The experimental results in brain 3 display the highest tumour density and are significantly different to the other two cases. Here it could be important to remember findings in the synthetic parameter estimation. The compact tumour simulations are expected to show a difference in effective and assigned migration rate as observed in the synthetic estimation of r_m. This could be one of the reasons for the varying parameter values observed in the accepted parameters of r_m shown in figure 4.11. The simulation with the lowest required relaxation, figure 4.13, show mere visual agreement with the experimental results. The tumour density seems reasonably estimated but the simulation fails to show the distinct tumour shape expressed in the experiments.

Turning the attention to coronal slice no. 140 in figure 4.12, it is possible to observe some unexpected tumour irregularities. The main tumour in this section shows unforeseen regions of zero cancer cell density in the middle of an otherwise compact tumour. It is important to remember that xenotransplantation is a large surgical intervention where there is a high risk of damaging the brain tissue. This could potentially lead to growth patterns impossible to observe in a healthy brain. Moreover, the slicing procedure after 40 weeks of tumour growth is a manual action with a risk of damaging the brain composition, including tumours. The slicing procedure is also uncertain in terms of the precision and relation to the corresponding coronal slice in the simulation domain. In this study, the experimental results have been assigned to a coronal slice by visual comparison of experiment and simulation domain, which of course introduces a source of error in the estimation of tumour characteristics from all brains.

5.3 Social and Ethical Aspects

The experimental data originating from xenotransplantations used in this study requires ethical reflection. To produce this data, cancer tissue from a patient diagnosed with grade IV glioblastoma was transplanted into another living creature. It is undoubtedly questionable to create a deadly disease in this way. However, all experiments are approved by the regional ethical review board and have been done with the intention to gain additional knowledge of a severe disease. The information have the potential to improve the treatment and in extension lead to longer survival times for patients diagnosed with glioblastoma. This study has used the medical images to develop a mathematical model to describe the spread and growth of glioblastoma tumours, aiming for a reduced need for future animal experiments and ideally it is possible to replace the experimental studies with reliable *in silico* models. In con-

clusion, it is possible to argue that the potential gain in knowledge from the medical images exceeds the caused harm, it is however important to acknowledge the forced damage and assure that it is not done in a wasteful manner.

6. Conclusion

We conclude this study by returning to the aims of the study. More specifically we repeat the sub-task which this study has treated.

The first task was to find a way to compare simulation outcomes with experimental data representing tumour growth patterns. Here the issue was approached with the means of Approximate Bayesian Computation. This poses the requirement for a generative mathematical model and summary statistics to compare simulations with the experimental data. The model suggested for the task was implemented as a stochastic, agent-based mathematical model. With the important detail of the inclusion of white matter tracts and brain vasculature to account for the brain heterogeneity. A selection of five summary statistics was introduced for the task of image comparison, which was conducted in two ways. It was shown that the direct image comparison technique was insufficient, mainly due to the neglection of the pixel intensity. The problem was overcome by a second, three-levelled approach. Furthermore, the advantage of using multiple summary statistics in combination rather than a single measure individually was determined.

It is possible to regard the model to be consistent, to a large extent, based on the parameter identification conducted for synthetic data. Here the approach showed success in many cases of individual and pairwise parameter identification with relatively precise posterior distributions. The results support both the use for the three-levelled image comparison and the least required relaxation of a more restrictive threshold. The estimation of white matter presence and base rates with heterogeneous migration produced less precise results, highlighting two potential deficiencies of the model. There were however reasonable interpretations for the reason of the weakness providing information for the use of the model in case of unknown parameters.

Lastly, we can state that the model and the framework of parameter inference show indications of parameter identification when it comes to estimating model parameters governing experimental growth patterns. We do not reach the aim, formulated in terms of distinct and reliable results describing tumour development, entirely. The results in this study showed high variability and limited visual correspondence to the experimental results. One of the main issues is related to the scarce sampling density with relatively few performed simulations. It is clear that the number of simulations needs to be increased for more reliable results. With additional time for simulations, it would also be beneficial to reduce the time step in all simulations, providing a higher time resolution and giving the possibility to recreate observed growth patterns with additional detail. It is also possible to increase the detail of

the number of cells, by increasing the carrying capacity the the model would result in more specific simulation and reducing the simplification in relation to the voxel size.

Furthermore, for future implementations is worth investigating different techniques of derivations of white matter tracts. All experimental results show growth patterns with close correspondence to well-known white matter structures. If it is possible to recreate the white matter tracts more accurately from the DTI data, there is a good possibility to improve the parameter estimation and significantly advance in terms of visual similarity to experimental results.

References

[1] E. C. Holland, "Glioblastoma multiforme: The terminator," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, National Academy of Sciences, 2000-06, pp. 6242–6244. DOI: 10.1073/pnas.97.12.6242. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC33993/.

[2] H. G. Wirsching, E. Galanis, and M. Weller, "Glioblastoma," in *Handbook of Clinical Neurology*, vol. 134, Elsevier, 2016-01, pp. 381–397, ISBN: 9780128029978. DOI: 10.1016/B978-0-12-802997-8.00023-2.

[3] M. Preusser, S. de Ribaupierre, A. Wöhrer, *et al.*, "Current concepts and management of glioblastoma," *Annals of Neurology*, vol. 70, no. 1, pp. 9–21, 2011-07, ISSN: 03645134. DOI: 10.1002/ana.22425. [Online]. Available: http://doi.wiley.com/10.1002/ana.22425.

[4] H. Zong, R. G. Verhaak, and P. Canoll, "The cellular origin for malignant glioma and prospects for clinical advancements," *Expert Review of Molecular Diagnostics*, vol. 12, no. 4, pp. 383–394, 2012-05, ISSN: 1473-7159. DOI: 10.1586/erm.12.30. [Online]. Available: http://www.tandfonline.com/doi/full/10.1586/erm.12.30.

[5] A. Ray-Chaudhury, "Pathology of Glioblastoma Multiforme," in *Glioblastoma: Molecular Mechanisms of Pathogenesis and Current Therapeutic Strategies*, Springer, New York, NY, 2010, pp. 77–84. DOI: 10.1007/978-1-4419-0410-2{_}3. [Online]. Available: https://link-springer-com.proxy.lib.chalmers.se/chapter/10.1007/978-1-4419-0410-2_3.

[6] D. N. Louis, A. Perry, G. Reifenberger, *et al.*, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, 2016-06, ISSN: 14320533. DOI: 10.1007/s00401-016-1545-1. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27157931/.

[7] Q. Xie, S. Mittal, and M. E. Berens, "Targeting adaptive glioblastoma: An overview of proliferation and invasion," *Neuro-Oncology*, vol. 16, no. 12, pp. 1575–1584, 2014-12, ISSN: 15235866. DOI: 10.1093/neuonc/nou147. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4232088/.

[8] V. A. Cuddapah, S. Robel, S. Watkins, and H. Sontheimer, "A neurocentric perspective on glioma invasion," *Nature Reviews Neuroscience*, vol. 15, no. 7, pp. 455–465, 2014-06, ISSN: 14710048. DOI: 10.1038/nrn3765. [Online]. Available: www.nature.com/reviews/neuro.

[9] H. J. Scherer, "The forms of growth in gliomas and their practical significance," *Brain*, vol. 63, no. 1, pp. 1–35, 1940-03, ISSN: 0006-8950. DOI: 10.1093/BRAIN/63.1.1. [Online]. Available: https://academic.oup.com/brain/article/63/1/1/263743.

[10] S. S. Margulies and D. F. Meaney, "Brain tissues," in *Handbook of Biomaterial Properties*, Springer US, 1998, pp. 70–80. DOI: 10.1007/978-1-4615-5801-9{_}8. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4615-5801-9_8.

[11] M. Esmaeili, A. L. Stensjøen, E. M. Berntsen, O. Solheim, and I. Reinertsen, "The direction of tumour growth in glioblastoma patients," *Scientific Reports*, vol. 8, no. 1, pp. 1–6, 2018-12, ISSN: 20452322. DOI: 10.1038/s41598-018-19420-z. [Online]. Available: www.nature.com/scientificreports.

[12] M. Cayre, P. Canoll, and J. E. Goldman, "Cell migration in the normal and pathological postnatal mammalian brain," *Progress in Neurobiology*, vol. 88, no. 1, pp. 41–63, 2009-05, ISSN: 0301-0082. DOI: 10.1016/J.PNEUROBIO.2009.02.001.

[13] A. Farin, S. O. Suzuki, M. Weiker, J. E. Goldman, J. N. Bruce, and P. Canoll, "Transplanted glioma cells migrate and proliferate on host brain vasculature: A dynamic analysis," *GLIA*, vol. 53, no. 8, pp. 799–808, 2006-06, ISSN: 08941491. DOI: 10.1002/glia.20334. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16541395/.

[14] K. Sasaki, S. Natsugoe, S. Ishigami, *et al.*, "Expression of CXCL12 and its receptor CXCR4 in esophageal squamous cell carcinoma," *Oncology Reports*, vol. 21, no. 1, pp. 65–71, 2009. DOI: 10.3892/OR{_}00000190.

[15] Y. Jiang and G. A. Johnson, "Microscopic diffusion tensor atlas of the mouse brain," *NeuroImage*, vol. 56, no. 3, pp. 1235–1243, 2011-06, ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2011.03.031.

[16] M. I. Todorov, J. C. Paetzold, O. Schoppe, *et al.*, "Machine learning analysis of whole mouse brain vasculature," *Nature Methods*, vol. 17, no. 4, pp. 442–449, 2020-03, ISSN: 1548-7105. DOI: 10.1038/s41592-020-0792-1. [Online]. Available: https://www.nature.com/articles/s41592-020-0792-1.

[17] H. Hatzikirou, A. Deutsch, C. Schaller, M. Simon, and K. Swanson, "Mathematical modelling of glioblastoma tumour development: A review," *Mathematical Models and Methods in Applied Sciences*, vol. 15, no. 11, pp. 1779–1794, 2005-11. DOI: 10.1142/S0218202505000960.

[18] A. Roniotis, K. Marias, V. Sakkalis, *et al.*, "Diffusive modelling of glioma evolution: a review," *Journal of Biomedical Science and Engineering*, vol. 3, no. 5, pp. 501–508, 2010-05. DOI: 10.4236/JBISE.2010.35070. [Online]. Available: http://www.scirp.org/Html/1791.html.

[19] K. R. Swanson, E. C. Alvord, and J. D. Murray, "A quantitative model for differential motility of gliomas in grey and white matter," *Cell Proliferation*, vol. 33, no. 5, pp. 317–329, 2000-10, ISSN: 0960-7722. DOI: 10.1046/j.1365-2184.2000.00177.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2184.2000.00177.x.

[20] R. K. Kwan, A. C. Evans, and B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1085–1097, 1999-11. DOI: 10.1109/42.816072. [Online]. Available: https://ieeexplore.ieee.org/document/816072.

[21] K. J. Painter and T. Hillen, "Mathematical modelling of glioma growth: The use of Diffusion Tensor Imaging (DTI) data to predict the anisotropic pathways of cancer invasion," *Journal of Theoretical Biology*, vol. 323, pp. 25–39, 2013-04, ISSN: 0022-5193. DOI: 10.1016/J.JTBI.2013.01.014.

[22] A. Swan, T. Hillen, J. C. Bowman, and A. D. Murtha, "A Patient-Specific Anisotropic Diffusion Model for Brain Tumour Spread," *Bulletin of Mathematical Biology 2017 80:5*, vol. 80, no. 5, pp. 1259–1291, 2017-05, ISSN: 1522-9602. DOI: 10.1007/S11538-017-0271-8. [Online]. Available: https://link.springer.com/article/10.1007/s11538-017-0271-8.

[23] H. Hatzikirou and A. Deutsch, "Cellular Automata as Microscopic Models of Cell Migration in Heterogeneous Environments," *Current Topics in Developmental Biology*, vol. 81, pp. 401–434, 2008. DOI: 10.1016/S0070-2153(07)81014-3.

[24] T. Alarcón, H. M. Byrne, and P. K. Maini, "A cellular automaton model for tumour growth in inhomogeneous environment," *Journal of Theoretical Biology*, vol. 225, no. 2, pp. 257–274, 2003-11, ISSN: 0022-5193. DOI: 10.1016/S0022-5193(03)00244-3.

[25] L. Naumov, A. Hoekstra, and P. Sloot, "Cellular automata models of tumour natural shrinkage," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 12, pp. 2283–2290, 2011-06, ISSN: 0378-4371. DOI: 10.1016/J.PHYSA.2011.02.006.

[26] G. J. Baker, V. N. Yadav, S. Motsch, *et al.*, "Mechanisms of Glioma Formation: Iterative Perivascular Glioma Growth and Invasion Leads to Tumor Progression, VEGF-Independent Vascularization, and Resistance to Antiangiogenic Therapy," *Neoplasia*, vol. 16, no. 7, pp. 543–561, 2014-07, ISSN: 14765586. DOI: 10.1016/j.neo.2014.06.003.

[27] P. J. Basser, J. Mattiello, and D. Lebihan, "MR Diffusion Tensor Spectroscopy and Imaging," *Biophysical Journal*, vol. 66, no. 1, pp. 259–267, 1994-01. DOI: 10.1016/S0006-3495(94)80775-1. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/8130344/.

[28] A. Badea, S. Gewalt, B. B. Avants, J. J. Cook, and G. A. Johnson, "Quantitative mouse brain phenotyping based on single and multispectral MR protocols," *NeuroImage*, vol. 63, no. 3, pp. 1633–1645, 2012-11, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2012.07.021.

[29] L. J. O'Donnell and C. F. Westin, "An introduction to diffusion tensor image analysis," 2011-04, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3163395/.

[30] P. N. Sen and P. J. Basser, "A model for diffusion in white matter in the brain," *Biophysical Journal*, vol. 89, no. 5, pp. 2927–2938, 2005-11, ISSN: 00063495. DOI: 10.1529/biophysj.105.063016.

[31] C. A. Combs, "Fluorescence Microscopy: A Concise Guide to Current Imaging Methods," *Current Protocols in Neuroscience*, vol. 50, no. 1, pp. 1–2, 2010-01, ISSN: 1934-8576. DOI: 10.1002/0471142301.NS0201S50. [Online]. Available: https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142301.ns0201s50.

[32] D. Siolas and G. J. Hannon, "Patient-derived tumor xenografts: Transforming clinical samples into mouse models," *Cancer Research*, vol. 73, no. 17, pp. 5315–5319, 2013-09, ISSN: 00085472. DOI: 10.1158/0008-5472.CAN-13-1069. [Online]. Available: www.aacrjournals.org.

[33] C. R. Ireson, M. S. Alavijeh, A. M. Palmer, E. R. Fowler, and H. J. Jones, "The role of mouse tumour models in the discovery and development of anticancer drugs," *British Journal of Cancer*, vol. 121, no. 2, pp. 101–108, 2019-06, ISSN: 1532-1827. DOI: 10.1038/s41416-019-0495-5. [Online]. Available: https://www.nature.com/articles/s41416-019-0495-5.

[34] H. A. Alturkistani, F. M. Tashkandi, and Z. M. Mohammedsaleh, "Histological Stains: A Literature Review and Case Study," *Global Journal of Health Science*, vol. 8, no. 3, p. 72, 2016-06. DOI: 10.5539/GJHS.V8N3P72. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4804027/.

[35] Y. Dodge, "Inference," in *The Concise Encyclopedia of Statistics*, New York, NY: Springer, New York, NY, 2008, pp. 263–263, ISBN: 978-0-387-32833-1. DOI: 10.1007/978-0-387-32833-1{_}196. [Online]. Available: https://link-springer-com.proxy.lib.chalmers.se/referenceworkentry/10.1007/978-0-387-32833-1_196.

[36] Y. Dodge, "Bayesian Statistics," in *The Concise Encyclopedia of Statistics*, Springer, New York, NY, 2008, pp. 32–36, ISBN: 978-0-387-32833-1. DOI: 10.1007/978-0-387-32833-1{_}23. [Online]. Available: https://link-springer-com.proxy.lib.chalmers.se/referenceworkentry/10.1007/978-0-387-32833-1_23.

[37] K. Csilléry, M. G. Blum, O. E. Gaggiotti, and O. François, "Approximate Bayesian Computation (ABC) in practice," *Trends in Ecology and Evolution*, vol. 25, no. 7, pp. 410–418, 2010-07, ISSN: 01695347. DOI: 10.1016/j.tree.2010.04.001.

[38] M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz, "Approximate Bayesian Computation," *PLoS Computational Biology*, vol. 9, no. 1, 2013-01, ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1002803. [Online]. Available: www.ploscompbiol.org.

[39] J. Pritchard, M. Seielstad, A. Perez-Lezaun, and M. Feldman, "Population growth of human Y chromosomes: a study of Y chromosome microsatellites," *Molecular biology and evolution*, vol. 16, no. 12, pp. 1791–1798, 1999-12, ISSN: 0737-4038. DOI: 10.1093/OXFORDJOURNALS.MOLBEV.A026091. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/10605120/.

[40] L. Zhu, Z. Xie, L. Liu, B. Tao, and W. Tao, "IoU-uniform R-CNN: Breaking through the limitations of RPN," *Pattern Recognition*, vol. 112, p. 107 816, 2021-04, ISSN: 0031-3203. DOI: 10.1016/J.PATCOG.2021.107816.

[41] F. Hou, W. Lei, S. Li, J. Xi, M. Xu, and J. Luo, "Improved Mask R-CNN with distance guided intersection over union for GPR signature detection and segmentation," *Automation in Construction*, vol. 121, p. 103 414, 2021-01, ISSN: 0926-5805. DOI: 10.1016/J.AUTCON.2020.103414.

[42] A. A. Taha and A. Hanbury, "An Efficient Algorithm for Calculating the Exact Hausdorff Distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 11, pp. 2153–2163, 2015-11, ISSN: 01628828. DOI: 10.1109/TPAMI.2015.2408351.

[43] V. M. Panaretos and Y. Zemel, "Statistical Aspects of Wasserstein Distances," *Annual Review of Statistics and Its Application*, vol. 6, pp. 405–431, 2019-03. DOI: 10.1146/ANNUREV-STATISTICS-030718-104938. [Online]. Available: https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-030718-104938.

[44] V. Horsley and R. H. Clarke, "The structure and functions of the cerebellum examined by a new method," *Brain*, vol. 31, no. 1, pp. 45–124, 1908-05, ISSN: 0006-8950. DOI: 10.1093/BRAIN/31.1.45. [Online]. Available: https://academic.oup.com/brain/article/31/1/45/460098.

[45] A. L. Stensjøen, O. Solheim, K. A. Kvistad, A. K. Håberg, Ø. Salvesen, and E. M. Berntsen, "Growth dynamics of untreated glioblastomas in vivo," *Neuro-Oncology*, vol. 17, no. 10, pp. 1402–1411, 2015-10, ISSN: 15235866. DOI: 10.1093/neuonc/nov029. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25758748/.

[46] F. Winkler, Y. Kienast, M. Fuhrmann, *et al.*, "Imaging glioma cell invasion in vivo reveals mechanisms of dissemination and peritumoral angiogenesis," *GLIA*, vol. 57, no. 12, pp. 1306–1315, 2009-09, ISSN: 08941491. DOI: 10.1002/glia.20850. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19191326/.

A. Complementary results of parameter estimations

In section 4.1 the single parameter estimation is omitted, except for the estimation of BV_p derived from a direct image comparison. The direct image comparison technique proved to be an inadequate comparison technique, most evident in the estimation of homogeneous migration with the estimation of the posterior distribution shown in figure 4.2. Hence it was decided to exclusively use the three-levelled image comparison method in the estimation of experimental results. In the following three figures the estimations of synthetic data are shown. The results were derived with the three-levelled approach to show the sufficiency of the method. Furthermore, the technique to accept the best percentage according to the least necessary relaxation is presented for each parameter estimation to provide an argument for the possibility to use this method in estimations of parameters from experimental results.

In figure A.1, the estimation of the increased proliferation probability due to the growth factors excreted by blood vessels is shown. To the left, the histogram of the posterior distribution, derived with the three-levelled threshold, $\boldsymbol{\varepsilon}_{max}^{[1,2,3]}$, is shown. Furthermore, the known parameter value is denoted by a dashed line. To the right, the posterior distribution is derived according to the least required relaxation of $\boldsymbol{\varepsilon}_{mean}^{[1,2,3]}$ needed for acceptance. Among all simulations, the best percentage was selected. This results in 54 parameters in the posterior distribution, compared to 710 accepted simulations with the maximum threshold. The more restrictive approach succeed to produce a close estimation of the known parameter value. Hence, we have support for the possibility to combine the vector of observed summary statistics to one measure, namely the relaxation coefficient.

In figure A.2 corresponding results are shown concerning the estimation of the preference in migration for blood vessels. The parameter is sightly overestimated in both methods of thresholds, to the left $\boldsymbol{\varepsilon}_{max}^{[1,2,3]}$ is used which results in 844 accepted parameters. The posterior distribution derived from the best percentage of simulation presented to the right in figure A.2, consists of 73 parameters.

The estimation of the migration preference for white matter, WM_w, demonstrate a new type of posterior distribution. The ABC procedure with the maximum threshold, $\boldsymbol{\varepsilon}_{max}^{[1,2,3]}$, results in 773 accepted parameters while only 2838 simulations were performed. The least number of total simulations in the case of synthetic parameter identification. The histogram of this estimation is shown in the left of figure A.3, where a larger range of the prior distribution is accepted by the procedure. To the right of figure A.3, the estimation derived from the least necessary relaxation is

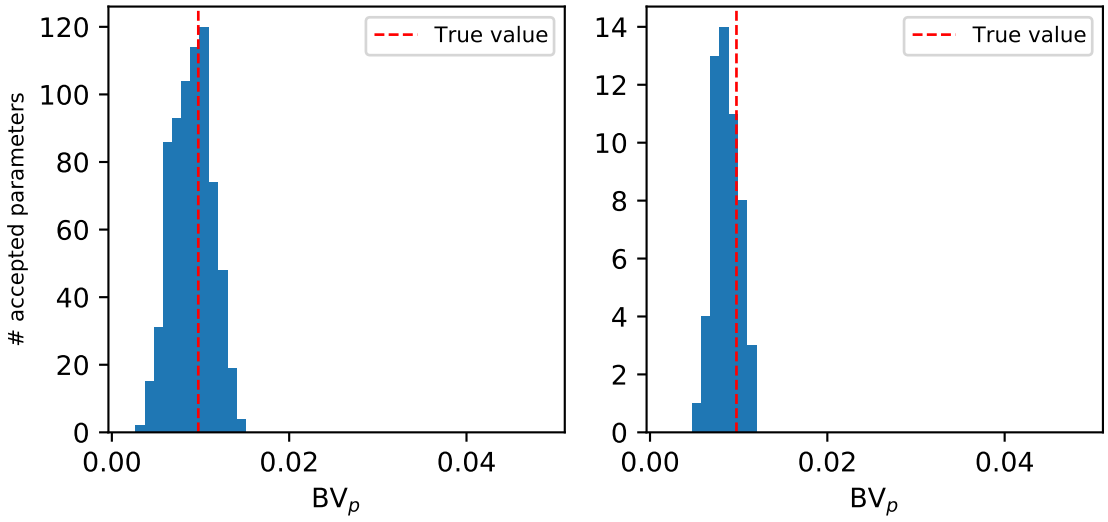


Figure A.1: Histograms representing the estimated posterior distribution of increased proliferation rate due to blood vessels presence, BV_p, derived with two different thresholds, $\varepsilon_{max}^{[1,2,3]}$ (left) and relative deviation from mean threshold (right). The three-levelled approach of image comparison was used in both cases.

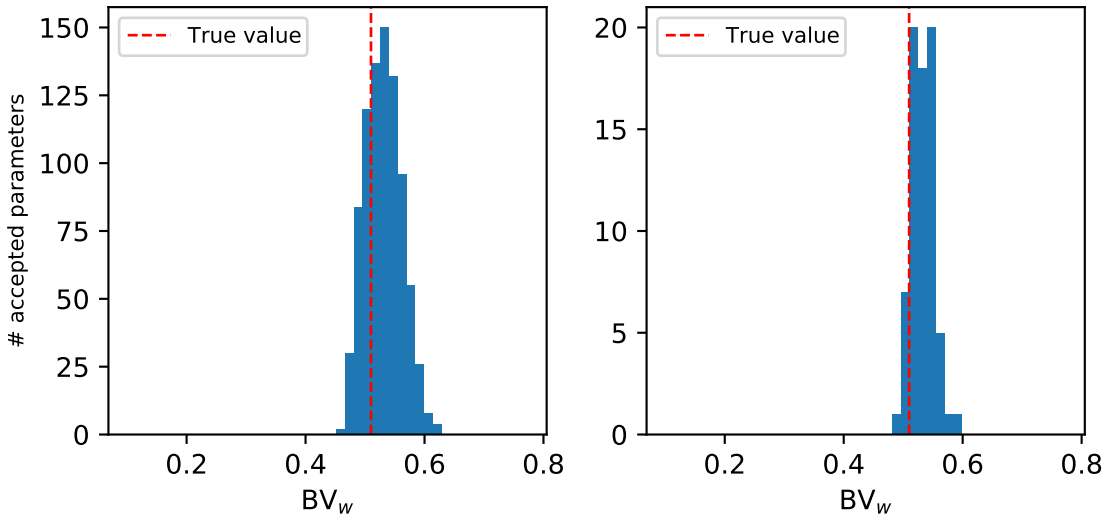


Figure A.2: Histograms representing the estimated posterior distribution of migration preference for blood vessels, BV_w, derived with two different thresholds, $\varepsilon_{max}^{[1,2,3]}$ (left) and relative deviation from mean threshold (right). The three-levelled approach of image comparison was used in both cases.

shown. Neither approach succeeds in terms of precise parameter estimation. However, the relaxation coefficient of the parameter included in the posterior is in a similar range as for the estimation of BV_p and BV_w, suggesting that a small change in the parameter value of WM_w results in a limited tumour difference. I.e. the effect of WM_w is smaller compared to BV_p or BV_w.

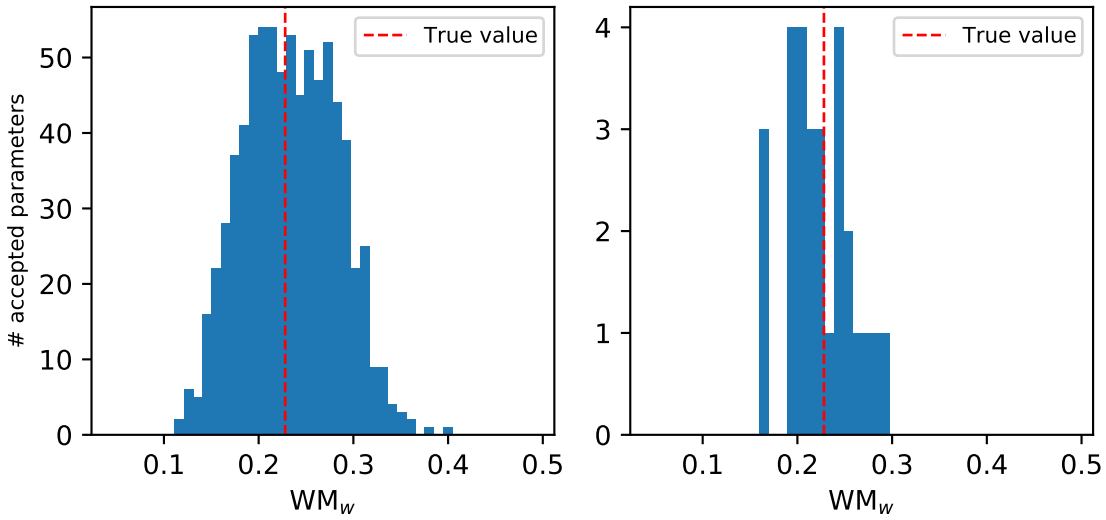


Figure A.3: Histograms representing the estimated posterior distribution of migration preference for white matter tracts, WM_w, derived with two different thresholds, $\varepsilon_{max}^{[1,2,3]}$ (left) and relative deviation from mean threshold (right). The three-levelled approach of image comparison was used in both cases.

B. Preliminary examination

To determine the necessary migration speed and the limits of the prior distributions in the context of estimating parameters governing specific experimental results a preliminary examination was performed. Here different time steps were investigated and compared to the invasion found in the experimental results.

The migration velocity in simulations is limited by the migration probability, $P_m = 1 \Rightarrow r_m = \frac{24}{\delta t} \Rightarrow v = \frac{24}{\delta t} \cdot \Delta x$, by (3.6) and (3.7). Accordingly, the migration velocity is inversely proportional to δt, hence an increased time resolution results in the possibility to simulate faster tumour invasion. Unfortunately, it is highly computationally costly to reduce the time step. Even if all estimations of experimental results would benefit from high time resolution we had to limit the computational time for this study. For efficiency, we decide the necessary time step for each experiment. In figure 3.1-figure 3.3 the different characteristics of the experimental results are shown, here it is possible to see a difference in tumour invasion. Most evident in the contrast between brain 1 and -3, in brain 1 cancer cells have reached the majority of the cerebrum in the available slice while in brain 3 we have a more compact tumour. This observation suggests the possibility to use different time resolutions to save computational time.

For fixed values of δt, tumours were simulated in isotropic tissue for different proliferation rates to get an idea of the necessary time resolution for corresponding experimental results. Here simulations were performed at the maximum allowed migration velocity to get an upper boundary of the possible invasion. Continuing, simulations were evaluated with respect to the minimum relaxation necessary for acceptance by threshold $\varepsilon_{exp}^{[1,2,3]}$, determined according to (3.11).

In figure B.1 the relaxation, r_i is reported for proliferation rates in the range $[0, 0.024]$, with the time step $\delta t = 2$h. Here we find the optimal proliferation rate in relation to the experiment brain 1 to be $r_p = 0.011\text{day}^{-1}$, we investigate this simulation in detail by a comparison to the experimental data. In figure B.2 the simulated tumour is shown, divided into image levels representing 1, 2 or 3 cancer cells per voxel, in figure B.3 the corresponding data is shown from the experimental tumour in brain 1. We conclude for a visual comparison that the tumour invasion and density are at a similar level resulting in fixation of the time step to 2 hours to estimate parameters of brain 1. The investigation was performed analogously for the other experimental results, here a lower temporal resolution was found to be sufficient. Namely 3 and 6 hours for brain 2 and -3 respectively.

In addition, also the proliferation rate has a negative impact on computational

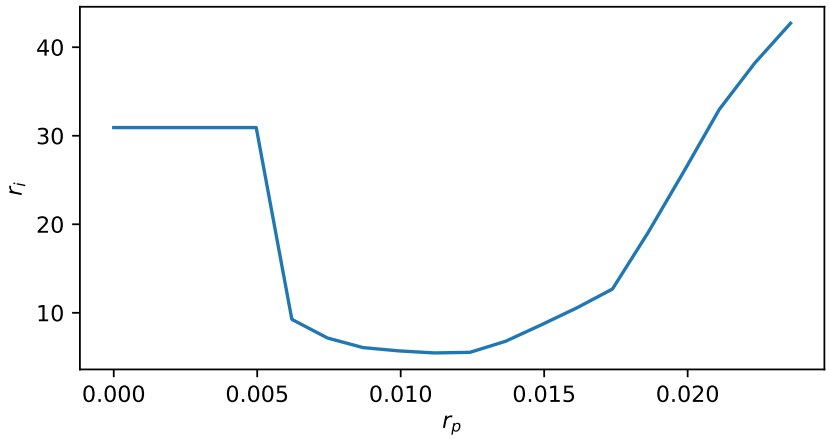


Figure B.1: Minimum required relaxation for acceptance of $\varepsilon_{exp}^{[1,2,3]}$ for different proliferation rates at time step, $\delta t = 2\text{h}$. Simulations have been evaluated in comparison to the experimental results of brain 1.

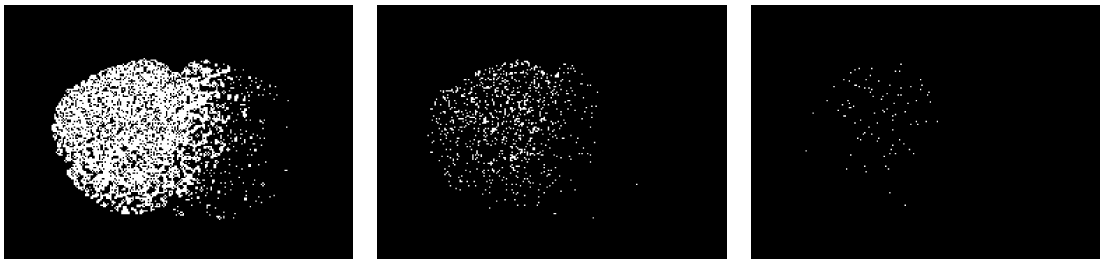


Figure B.2: Optimal simulation from preliminary investigation, $r_p = 0.011\,\text{day}^{-1}$, divided in image level 1, 2 and 3 from left to right.

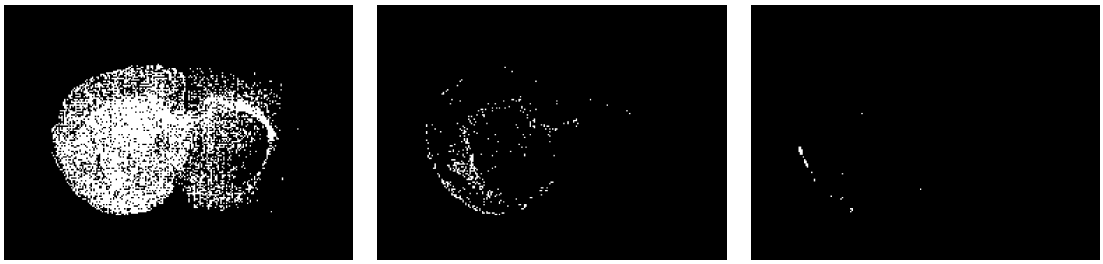


Figure B.3: Experimental result of brain 1, coronal slice no. 90, divided in image level 1, 2 and 3 from left to right.

time. With an increased proliferation rate simulations result in larger tumours with an additional amount of proliferation and migration events to be computed. Hence, there is time to save by selecting the prior boundaries relatively restricted. However, we still need to assure that simulations with proliferation rates outside of the prior limit not will be accepted. We return to figure B.1, here we see the trend of the relaxation needed for acceptance. For increasing proliferation rates the cell density will increase and the tumour presence level 1 will shift towards level 2 and -3 to a greater extent. This is seen in the required relaxation which rises for values of $r_p > 0.020\,\mathrm{day}^{-1}$, and there is no return from this point. In this way, the prior boundaries for the proliferation rate were selected for all simulation configurations in advance of the ABC procedure.

In conclusion, we arrive at the values for fixed parameters given in table 3.5, where the time step was derived by means of a visual comparison explained above, the 40 week simulation time was set to correspond to the length of the experiments. The migration rate is a subject of estimation and the prior distribution were including all feasible values of proliferation probability, i.e. $P_m \in [0, 1]$. An equivalent approach was adopted for the weight parameters governing migration preference, all allowed values were included in the prior distribution.

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY