UiO : University of Oslo

Lai, Xiaoran

Modelling, inference and simulation of personalised breast cancer treatment

Thesis submitted for the degree of Philosophiae Doctor

Oslo Centre for Biostatistics and Epidemiology (OCBE)
Faculty of Medicine

2021

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1234

ISSN 1234-5678

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Be approximately right rather than exactly wrong.
- John Tukey

Preface

This dissertation is submitted in partial fulfilment of the requirements for the degree of *Philosophiae Doctor* at the University of Oslo. The research presented here is conducted under the supervision of professor Arnoldo Frigessi and researcher Alvaro Köhn-Luque.

The dissertation is a collection of three papers, presented in chronological order. The common theme to them is personalised simulation of cancer therapy. The papers are preceded by an introductory chapter that relates them together and provides background information and motivation for the work.
Paper I is a joint effort of many researchers from Univeristy of Oslo (UiO) and Oslo University Hospital (OUS) in Oslo, Norway over the course of three and a half years. The ongoing work associated with Paper II has been produced by my co-authors and I from UiO and Aalto University in Helsinki, Finland. Paper III is the fruitful production of brilliant collaborators at UiO and Simula Research Laboratory in Lysaker, Norway.

This dissertation should be of interest to statisticians and mathematicians, in the field of Cancer Biology specifically, but also other individuals with a general background in modelling. It should also be of interest to cancer biologists and clinicians. This research was supported, in part, by Big Insight, Oslo, a centre for research-based innovation (sfi), funded by the Research Council of Norway and other fifteen partners.

Acknowledgements

I thank my supervisor professor Arnoldo Frigessi for his constant and selfless care and time. I also thank co-supervisor Alvaro Köhn-Luque for his parachuting appearance four years ago to make everything in this project possible.

The works would not be possible without heartfelt initiatives of their respective principle investigators. I deeply appreciate their encouraging discussion and their precious knowledge in their respective field. The fruitful results of the project cannot be realised without the collaborative efforts from my co-authors at OUS, in particular at the department of radiology, department of diagnostic physics, department of genetics and department of oncology. My gratitude to my co-authors at Simula Research Laboratory. It has been an humble and remarkable experience to be working at Simula. Your expertise and professionalism truly impress me. I also would like to thank my most distant co-authors at Aalto University for their instrumental contribution to the statistical aspect of this project.

Last but not least, my sincere appreciation goes to Håkon Taskén, for his magic hands to conjure up beautiful computing codes during his summer stay at

UiO.

This four and a half year journey would not be the same without the emotional support from my colleagues at Oslo Centre for Biostatistics (OCBE). I am eternally in debt to my parents. I always knew that you believed in me and wanted the best for me. Thank you for teaching me that my job in life was to learn, to be happy, and to know and understand myself; only then could I know and understand others. And to Bea, I am much more me when I am with you.

Lai, Xiaoran
Oslo, April 2021

List of Papers

Paper I

Lai X, Geier O, Fleischer T, Garred Ø, Borgen E, Funke SW, Kumar S, Rognes, ME, Seierstad T, Børresen-Dale, A, Kristensen, VN, Engebråten O, Köhn-Luque A and Frigessi, A. (2019). Towards personalized computer simulation of breast cancer treatment: a multi-scale pharmacokinetic and pharmacodynamic model informed by multi-type patient data. Cancer Research May 22 2019 DOI: 10.1158/0008-5472.CAN-18-1804

Paper II

Lai X, Pesonen H, Köhn-Luque A, Kaski S, Corander J, and Frigessi A. (2019). Likelihood-free inference for hybrid cellular automaton models for personalized simulation of breast cancer treatment. *In preparation.*

Paper III

Lai X, Taskén H, Funke SW, Frigessi A, Rognes ME, and Köhn-Luque A. (2019). Scalable solver for a multiscale model of personalized breast cancer therapy. *In prepartion.*

Contents

List of Figures

Chapter 1

Introduction

Despite many advances in the treatment of breast cancer, treatment failure incidence is still high, entailing a heavy burden upon the affected individuals and healthcare systems in general [1, 2]. Although each cancer patient receive the recommended treatment, this may not represent the optimal one. This includes patients who do not respond to standard treatment, develop resistance to drugs or relapse, and patients who do not qualify for standard-of-care treatment, because of age or health condition. It could be that a treatment that would benefit such a patient actually exists, as a non-standard combination of currently available drugs and compounds, in a novel dose and administration regimen. Personalised treatment requires searching for this optimal treatment, taking into account the individual genomic makeup, transcriptional profile, proteome, microenvironment, clinical condition and life expectancy. Medical science has already made important steps in this direction [3]. Modern development in cancer treatment is moving away from the 'one drug fits all' ideology to a more personalised approach motivated by advancement in biomarker technologies. However, this level of 'personalisation' is far from ideal: it assigns patients to one of a few classes, in which intra-patients' intra-heterogeneity can be observed [4].

Mathematical modeling and simulation tools are becoming an attractive and time- and cost-effective approach to determine optimal solution among numerous possibilities [5, 6]. Current models are capable of addressing pharmacokinetics and pharmacodynamics of anticancer medicine at various spatial and temporal scales. Simulations can then be performed to explore many treatment regimens at once to identify optimal plans with minimal toxicity. However the flexibility of the model comes at a cost; parameters for each individual patient requires separate tuning and validation, and the runtime of simulations remains unmatched with the state-of-the-art decision-aiding tool in routine clinical practice.

In this project, I investigate whether

- a mathematical model designed for a specific type of cancer, integrating routinely-collected data from a clinical trial is feasible,
- the model is robust enough to simulate and predict various responses using individual data, and
- the collected data are sufficient for the purpose of validation and personalisation of the model

We use data from a recently published neoadjuvant clinical phase II trial in patients with large breast tumours [7], where histological, magnetic resonance

imaging (MRI) and molecular data were collected before, during and at the end of neoadjuvant treatment. The enrolled patients were randomised to receive FEC with or without bevazicumab, an anti-angeogenic drug. Clinical observations show large heterogeneity between patients, however no predictive tools are currently available. I joined a team composed of oncologists, pathologists, molecular biologists, medical imaging physicists, statisticians and mathematicians. Together we examined whether a computational modelling approach can be used as a tool to predict individual outcomes of patients in the trial. In the first part of the project, presented in Paper I, we formalised procedures for clinical data preprocessing, model initialisation, personalisation and validation. we demonstrate that patient-specific (incorporating data unique to the individual) and multi-scale (combining processes and data of different length- and time-scales) modelling allows us to reproduce treatment outcome. In addition we investigate what alternative treatment protocols would have produced different outcomes. This is a first step towards virtual treatment comparison.

Based on the success of the first paper, we use modern statistical computation techniques known as approximate Bayesian computation (ABC) to infer uncertainty in the patient-specific parameters from the Paper I given available data. Additionally, we discussed whether additional data collection would lead to improvement in the accuracy of outcome prediction.

Finally in Paper III, we strive to improve the relevancy of the scale of the original model by developing novel efficient numerical methods, built upon modern cluster computing architecture with parallel capabilities to address heterogeneities of tumour microenvironment. This crucial step enables simulation of personalised therapy that are up to 500 times larger than our previous achievement.

Overall, our study demonstrates the effectiveness and the potential of simulation-based personal treatment optimisation. It lays the basis for future programme in delivering robust clinic companion diagnostic tool.

The remainder of this thesis is structured in the following manner: in Chapter 2, a general overview on breast cancer and development of clinical treatment are discussed. Chapter 3 introduces the role of mathematical modelling in cancer medicine and its application. In Chapter 4, the need and usage of numerical approximation of differential equations in the context of FEniCS are discussed. And finally Chapter 5 explains the rationale of bayesian inference and in particular, some aspects of likelihood-free inference on simulation-based models.

Chapter 2

Breast Cancer

2.1 Epidemiology

Breast cancer is the most common type of cancer for women and number two commonly occurring type of cancer. In Norway, there are approximately 3600 cases of newly diagnosed breast cancer patients, accounting for ever 25% of diagnosed cancer cases among women. Over the past 50 years, the incidence rate of breast cancer has increase by over 50% [8]. There are several factors considered to be associated with increasing risk of breast cancer and fall into two categories: hereditary and environmental factors. Hereditary factors including age, familial history and ethnicity account for approximately 27 percent of breast cancer cases. Notably, two hereditary breast cancer genes associated with high risk, BRCA1 and BRCA2 account for 5-10 percent of all cases. However environmental factors account for substantial proportion of breast cancer. Lifestyle and dietary choices have been associated with increased risk of getting the disease. Given that age is a risk factor for breast cancer, the increasing incidence rate can be partially explained by the overall increase in life expectancy. Additionally the increase in the screening programme also increased the probability to detect the disease.

2.2 Diagnostic and Prognostic Modality

2.2.1 Imaging

Early suspicion of breast cancer often arises from palpation, a form of physical examination, either by a medical examiner or by self-examination. Another form of effective detection is by mammography. It is a type of X-ray image of breast that enables earlier diagnoses [9]. This is usually carried out as part of a national screening programme [10]. Norway, amongst other nordic countries, UK and the US, have a well-established breast cancer screening programme. Women between the age of 50 and 69 are screened regularly every two years. Furthermore, women with familial history are offered screening earlier on due to higher breast cancer risk [11]. For more accurate diagnosis and better localisation, other imaging tools are used such as magnetic resonance imaging (MRI) and positron emission tumography (PET).

2.2.2 Histopathological biopsy

Once a patient is found to be positive to the test, a histological examination of fine needle aspiration biopsy (FNAB) or core needle biopsy usually follows to confirm the finding. The biopsy can also be preserved in formalin and paraffin, and take out later for further assessment by a trained histologist.

2.2.3 Molecular data; gene-expression

New technology, microarray gene expression profiling, has now started to be routinely collected. It gives additional insights of the complexity as well as the heterogeneity of the disease at molecular level. Treatment can be further selected and optimised for each patient. Multi-gene signatures are collected and identified to capture key molecular features of the tumor. Such signatures like OncoTypeDX (21-gene signature) [12, 13], and MammaPrint (70-gene signatures)[14, 15], provides invaluable prognostic tool in addition to clinical prognostic factors for both overall survival and recurrence of the disease and can be used to predict response to different treatment.

2.3 Taxonomy

Breast cancer is recognised as a heterogeneous disease with different characteristics, which leads to various response to one treatment. In this section breast tumour is separated by classification criteria, including histopathological, immunopathological and molecular separates breast into several groups.

2.3.1 Histopathological classification

Depending on the morphology of the tumour, breast cancer can be divided into several categories. Lobular and ductal breast carcinomas accounts for more than 95% of all breast cancers. The majority of the ductal breast carcinomas are invasive ductal carcinomas (IDCs) that grow in ducts or tubules and infiltrate the surrounding tissue. Approximately 75% of IDCs are not classified further and they are referred to as 'not otherwise specified' (NOS). Invasive lobular carcinoma (ILC) is the second most commonly diagnosed type of breast carcinoma (10-15%). The number of diagnosed cases are found to be increasing possibly due to an increase in prescription of hormonal replacement therapy for postmenopausal women [16]. Other types of breast carcinoma such as Inflammatory breast cancer (IBC) are named after the structure of the growth [17].

2.3.2 Predictive biomarkers

There are several biomarkers proposed for breast cancer. They are used to adjust or omit treatment according to the aggressiveness of the disease. In addition, they can also be used to estimate how likely the treatment would be successful for high-risk patients.

Histological grade A crucial and important guidance for diagnosis and prediction of IDC is histological grade. It is determined by a trained pathologist who studies the degrees of epithelial differentiation of tumour cells in a sample obtained by biopsy or surgery. Three features are evaluated for the estimation of histological grade: number of dividing cells, irregularity in shape and size and formation of tubules. Each feature is assigned a numeric value ranging from 1

to 3 by independent assessment. Grade 1 corresponds to a well-differentiated tumour, while grade 3 tumour is poorly differentiated. It has been shown that high grade tumours are often associated with poor prognosis [18, 19].

Stage - TNM classification In contrast to histological grade, staging is a systematic grading taking into account the biopsy of a tumour. This standard is globally recognised as a standard tool. The evaluation takes into account three aspects of a primary tumour: tumour size (T), lymph nodes status (N) and the presence of distant metastasis (M). Both T and N are indicative of the aggressiveness of the primary tumour, while M could indicate the disease is incurable at the time of diagnosis [20, 21].

Cellular receptor Estrogen receptor (ER), progesterone receptor (PR) and growth factor receptor (HER2) are the three important prognostic and predictive markers for breast cancer. [22, 23]. These three expressions forms part of the routine inspection of the immunohistochemical (IHC) staining of formalin-fixed,paraffin-embedded breast cancer tissues. ER and PR are receptors activated by estrogen and progesterone ligands binding. Such mechanism is responsible of the regulation of cellular functions such as proliferation, differentiation and angiogenesis. Positively expressed ER and PR status usually have a good response to endocrine therapy, while low/absent PR expression accompanying positive ER status indicates high proliferation and poor response to endocrine therapy. HER2 is a growth factor receptor tyrosine kinase than is involved in cell growth regulation. It is found to be overexpressed in approximately 20% of all breast tumours and is a marker for poor prognosis [24]. Three clinically relevant subtypes of breast cancer are defined from the levels of expression of ER, PR and HER2: hormone receptor positive, HER2 positive (regardless of ER and PR status) and triple negative (TN) breast cancer, meaning absence of all three receptors. 70-80% of the breast cancer population are classified as ER positive, while 12-17% are TN.

Ki-67 Ki-67 is another biomarker for proliferation. It is a nuclear protein expressed only in active cells. Its activity can be measured by IHC, called Ki-67 index, and it is particularly useful in determining proliferation status [25]. Despite recommendations from the Breast Cancer Working Group, the main difficulty in integrating Ki-67 proliferation score as part of the clinical practice is its high inter-observer variability. A systematic assessment in its reliabilities as a useful biomarker is needed [26].

2.3.3 Molecular classification; gene-expression signatures as classifiers

Recent advancement in microarray gene expression profiling made it possible to gather high-dimensional data and to classify breast cancer samples based on multiple genes or even the whole genome. The most prominent study proposed

five biologically distinctive subtypes: luminal A, luminal B, HER2-enriched, basal-like and normal-like. Initial study [27, 28] concluded the subtypes using approximated 500 genes because they varied the least within the same tumours, but differed the most between different tumours comparing expressions before and after doxorubicin treatment. Results were validated and refined further, shrinking the set of genes to a total number of 50, naming the collection PAM50 [29]. In addition to its usage in subtyping, it was also developed to assess prognostic risk of relapse, proliferation based on subtypes of gene related to cell cycle, and to aid in composite scores including tumour size with molecular phenotypes [30]. In light of seeking minimising misclassification and identifying patients with poor chemotherapy treatment outcome, prospective cohort studies have been set up to study the potential of PAM50 as predictive marker for risk of recurrence in Norway [31].

There are also several other multi-gene signatures that have been defined to capture features of tumours associated with treatment response, namely OncoTypeDX and MammaPrint. However their value for long-term outcome prediction have been questioned despite their prognostic value [32].

2.4 Treatment

After the patient has been detected of breast cancer by palpation or mammography, biopsy, and imaging of type ultrasound, MRI or PET are taken for accurate diagnosis and localisation. Treatment is then planned by considering aforementioned prognostic and predictive biomarkers for each patient. Treatment types can be roughly classified into two categories. Local therapy, meaning treatment without affecting other parts of the body, includes surgery and radiotherapy. On the other hand, systematic treatment uses drugs that targets cancer cells throughout the whole body. Depending on the types of drugs, the treatment, including chemotherapy, hormonal therapy, targeted therapy and immunotherapy, can be administered orally or intravenously. The increasing involvement of a multi-disciplinary team in the proposed treatment plan, which usually contains a combination of local and systematic treatment, has seen a significant improvement in breast cancer survival in the intervention. However advances in microarray-based gene profiling enables intrinsic intra- and inter- tumour heterogeneity. This explains many of the varying responses to standard treatment. Currently we are in a transition era from the traditional population-based oncology to the primarily molecular-marker based individualised assessment and programme adapted to personal care.

2.4.1 Local therapy

Surgery Operations used to treat breast cancer include lumpectomy, a breast-conserving surgery, and mastectomy, the removal of the entire breast. Lymph nodes dissection may proceed either of the procedure to determine if the cancer

has metastasised in the axilla. A less invasive technique called sentinel lymph node (SLN) procedure is used initially. If cancer cells are found in the lymph node of the primary site, extensive axillary lymph node dissection is performed. Surgery is thought to be the primary intervention of primary breast cancer of stage I-III before the tumour metastasises.

Radiotherapy Radiation therapy utilises highly energetic ionising radiation such as X-rays and protons to damage the DNA of the cells including normal and cancer cells. Since cancer cells are on average more proliferative than normal cells, less differentiated and with limited DNA-repair mechanism, it is more damaging to cancer cells compared to normal cells. Irradiation after breast-conserving surgery substantially reduces not only the risk of local recurrence but also improves the survival[33]. The procedure also prevents the need of mastectomy and improves the cosmesis among the breast cancer patients.

2.4.2 Systematic treatment

Systematic therapy can be given in addition to surgery (*adjuvant*) to prevent recurrence of tumour and or metastasis. It can, however, be given prior to surgery, called *neoadjuvant* therapy. The aim of this approach is to help shrink the tumor and make it easier to remove surgically. In both of these situations, the main aim is to kill or control undetected microscopic metastases before they grow. This reduces the risk of distant spread of breast cancer in the years after surgery.

Chemotherapy An oncologist will recommend a chemotherapy regimen entailing plans for chemotherapy. This usually includes the name of the receiving drugs, the order in which they will be administered, the amount of each drug (the dose) as well as how often and how long the treatment lasts. Most women with early-stage breast cancer will typically receive a regimen that last approximately three to six months. It was shown that combination of cytotoxic drugs targeting different cell cycle phases is more effective [34]. The most commonly prescribed chemotherapy regimen in Norway includes three drugs: 5-flurouracil, epirubicin and cyclophosphamide, collectively called FEC. For tumours expressing high level of Ki67, FEC are given in addition to taxanes.

Hormone therapy Hormone therapy targets the presence of ER signalling and block the binding of oestrogen to the ER or synthesis of oestrogen, inhibiting tumour cell proliferation. For patients who haven't gone through menopause, an estrogen receptor blocker such as tamoxifen is usually recommended [35]. In post-menopausal women, oestrogen is produced via aromatisation of ovarian and adrenal androgens in the liver, muscle and fatter tissue. Aromatase inhibitor is therefore commonly prescribed. It blocks estrogen production and thereby starves cancer cells of the hormones they need for growth.

Anti-angiogenic therapy A current focus of breast cancer research is to seek out medicine that targets specific molecules concerning breast cancer development.

It was observed that a tumour cannot grow beyond a size with diameter of 2mm without formation of functional vessels (angiogenesis). As it grows beyond 2mm in diameter, its need for oxygen and other nutrients outstrip its supply. Tumour enters hypoxic states, producing growth factors such as VEGF, TGF-β and PDGF signalling for new vessels. The vasculature of tumour is highly irregular, caused by the abnormally high levels of VEGF. The open gaps between endothelial cells on the surface of tumour vessels make delivery of nutrients and, in particular render systematic therapy ineffective. Bevacizumab is a humanised monoclonal antibody to Vascular Endothelial Growth Factor A(VEGF-A), a protein produced by cells that stimulates the formation of blood vessels (angiogenesis) and increases vessel permeability. It works by directly inhibiting VEGF and affect normalisation of existing vessels, removal of abnormal and non-functional vessels and inhibition of new vessel growth. It has shown clinically that, bevacizumab in combination with chemotherapy versus chemotherapy can lengthen progression-free survival and increases the response rate in first-line therapy for locally recurrent or metastatic breast cancer [36, 37]. However the benefits are transitory, and research shows that the effects and the possible resistence mechanisms are much more complex than initially thought [38].

Chapter 3

Mathematical Modelling of Cancer

3.1 Mathematics in medicine

Mathematics has been used to model biological phenomena for centuries. Mathematical model allows precise formulation of biological hypothesis on a mechanism and provide predictions that can be tested experimentally. Such mathematical model has to be simple yet complex enough to be able to approximate biological processes at work with high accuracy given each parameter in the model underlying biological processes.
Recently, medicine is becoming increasingly dependent on mathematics [39–41]. With the aid of computers, the involvement and usage of mathematics has led to the understanding and discovery of recent medical advances. For example: predicting sudden cardiac death [42, 43], in-silico simulation of human brain [44], developing cost-effective strategy for cancer screening and modelling potential drug targets for specific disease [45].

3.2 Mathematical oncology

Cancer modelling has been one of the fast-growing and challenging research topics gathering applied mathematicians and other scientist. Its popularity is not only raised by its scientific challenge, but also because of the growing number of death cases caused by cancer, superseded by cardiovascular diseases. Cancer biology involves complex intracellular and/or intercellular behaviours, such as DNA replication, cell growth, division and migration, with very specific interactions in different tissues parts. Although current mathematical tools can provide a solid foundation for the study of the biological processes, there are still lacking integrated models which can be used to investigate fully the interactions between different components. An integrated, multi-scale model addressing each characteristics of cancer, including processes such as proliferation, invasion, resistance and angiogenesis, can provide a mechanistic explanations for the observed behaviour, and could potentially contribute towards the advancement in the understanding of cancer and better treatment outcome.

3.3 Population-based model

Tumors are constantly evolving in time. Cancer cells can proliferate, be quiescent, or die, and other common variables controlling the cancer system, such as substrate concentration in cell's microenvironment and expression levels of functional molecules triggered by certain event varies. Population-based model, also known as continuous model uses differential equations to formulate changes

in quantities such as as cell number and substrate concentration at a global scale. Depending on its consideration for spatial dependence, two types of differential equations, ordinary differential equations (ODEs) and partial differential equations (PDEs) emerge as powerful tools in simulating tumour growth and treatment response.

3.3.1 Ordinary differential equation models

ODEs can not only be used to define changes in cell numbers and substrate quantities under normal conditions, but also to incorporate cells' response to various treatment.

Fundamentals of growth models The simplest tumour growth model is called exponential growth model. It can be formulated as: Given λ, a constant describing the net population growth rate, the cell population at time t, c_t can be formulated as

$$c_t = c_0 \exp(\lambda t) \tag{3.1}$$

where c_0 is the initial number of cells at $t = 0$. if $\lambda = 0$, the number of cells stays constant over time; while $\lambda < 0$, or $\lambda > 0$ corresponds to decreasing and increasing population at a constant pace over time respectively. The value of λ can be estimated from experiments by fitting observed cell count data to Equation (3.1).

While the exponential growth model is an simple model with closed-form solution, the assumption of constant net growth of cancer cells is biologically and physically unrealistic. A modification on the exponential growth model [46] proposes the growth model has an initial exponential phase followed by a linear phase, called exponential-linear model, and it was used to described tumour growth in nontreated animals. This model was extended further to include two more stages, namely stasis and second growth. It was used in the context of three-dimensional tumour growth and angiogensis under chemotherapy treatment [47].

While the exponential/exponential-linear model is a simple model with closed-form solution, the assumption of constant net growth of cancer cells is biologically and physically unrealistic. It ignores physical constraints such as the maximum capacity of the growing environment, the carrying capacity, and limits on resources. It is therefore natural to reformulate the changes in tumour growth as a function of the tumour size:

$$\frac{1}{dt}c = f(c) \cdot c \tag{3.2}$$

One of the classic tumour growth model is the logistic model, anticipating a slower growth as the population reaches its carrying capacity K:

$$\frac{1}{dt}c = \lambda c(1 - \frac{c}{K}). \tag{3.3}$$

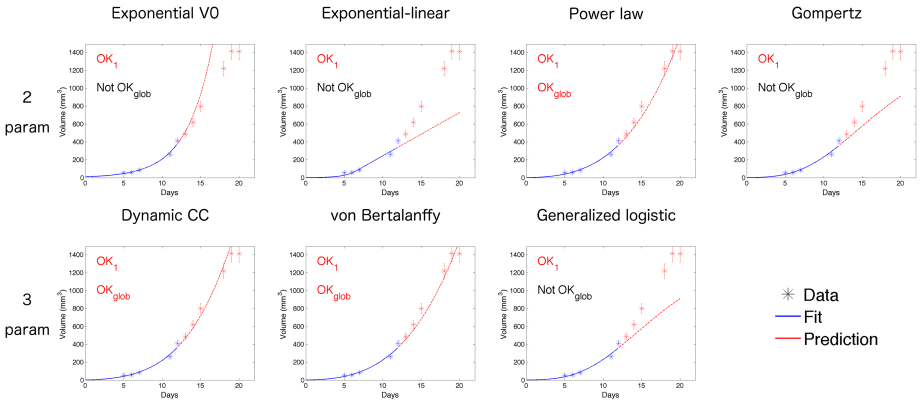

Figure 3.1: Representative examples of classic cancer growth models fit to lung data set. Figure by [54] (CC0)

This means that, when $c << K$, cell grows at approximately rate of λ, equating to exponential growth (c.f. Equation (3.1); when $c \to K$, cells are reaching carrying capacity, the growth is stalling and approaching rate of 0.
Other dynamic growth rate functions exist and have been discussed in the context of tumour growth [48, 49]. Another popular growth model has been shown to reproduce biological relevant growth pattern is the Gompertz growth [50]. It assumes that the rate of growth decreases as the population grows[51]. The restriction on constant carrying capacity can also be alleviated by adding a second differential equation, forming a system ODE mimicking a vascular cancer system under the stress of growing cells. (c.f. Von Bertalanffy [52] and power law models [53, 46]). See Figure 3.1 for an example of various models fitted to a lung dataset. I refer reader to [54] for a comprehensive treatment of comparison in different classical mathematical model for tumour growth.

It is relatively straight-forward to incorporate tumour treatment and it follows two schemes: the decrease in tumour size caused by treatment can be imposed by adding an extra cell death term, or by reducing the carrying capacity. The first scheme is suitable for demonstrating a continuous tumour cell kill such as chemotherapy or immunotherapy, while the effect of angiogenesis is best suited for the second scheme for its ability in regularising vascular network.

Usage in Pharmacokinetic(PK) model Pharmacokinetic model describes the rate of distribution of a drug to different tissues and the rate of elimination of the drug. The mathematical framework is developed using compartments, consisting of tissues with its own distribution characteristics. PK parameters such as clearance (CL) and volume of distribution (V) of each compartment and their variability are estimated from drug concentration-time data. Individual patient PK parameters are used to explain intra-patient variability while typically

average dynamics are used due to difficulties in obtaining individual data. They can be implemented into the model through individual patient characteristics such as weight and height.
A simple multi-compartment model contains three levels of information: the elimination process can be either linear or Michaelis-Menten-like; second is to consider the number of compartments, which typical ranges from one to three; And finally the route of administration, *e.g.* intravenous bolus, infusion, or oral, is approximated up to first order. The advantage of the compartment model is that the concentration of drugs can be obtained at any given time. However pharmacokineticist make certain assumptions on the body compartments which can be difficult to validate. Hence non-compartmental methods, which are model-independent and consistent, is favoured in establishing the initial exposure characteristics of a drug prior to entry into the clinic [55].

3.3.2 Partial differential equation (PDE) models

Despite ODE models have shown to be remarkably useful in simulating temporal evolution of cancer cells, they are unable to describe the internal spatial structure of the tumour population and uncover characteristics of distinct spatial patterns because of tumour heterogeneity. Invasion, angiogenesis and metastatic spread are three of the main hallmarks of cancer. These two properties are inherently spatial, and they can be captured by a PDE model. In this model, quantities such as cell density or fraction of tumour volume for a given population are captured at a location in either 2D or 3D. These quantities not only depend on time, but also space.

Reaction-Diffusion equation In their simplest form, a reaction diffusion equations describe the change of the concentration of a substrate in space and time. Reaction causes the transformation of substances, while diffusion causes the substances to spread out in space. The reaction-diffusion equation for diffusion of a substance can be derived from the mass conservation equation, dictating that the net in-flow and the net generation equal to the rate of change in the substance, relating the velocity of the substance (flux) to the concentration. In 2D, this gives:

$$\frac{\partial c}{\partial t} = D(\frac{\partial^2 c}{\partial^2 x} + \frac{\partial^2 c}{\partial^2 y}) + f(x, y, t), \tag{3.4}$$

which is often written in a more compact form:

$$\frac{\partial c}{\partial t} = D\nabla^2 c + f, \tag{3.5}$$

where the operator ∇^2 is defined as $\nabla^2 c = \frac{\partial^2 c}{\partial^2 x} + \frac{\partial^2 c}{\partial^2 y}$, called the Laplace operator. $f(\cdot, t)$ is the reaction term describing the generation and depletion of the substance in time. D is the diffusivity of the substance, with unit space-squared over time [56].

Reaction-diffusion describing tumour growth Here we present a few examples in the application of reaction-diffusion PDEs in describing tumour growth. Inspired by previous work, Gatenby and Gawlinski [57] developed the first spatial model of cancer invasion. The model takes into account the diffusion and proliferation of cancer cells in degrading local tissue casued by excces H^+ ions. The resulting system of PDEs in terms of cancer cell density, c, amount of H^+ ions, m, and extracellular matrix, v, can be written as:

$$\begin{aligned}\frac{\partial c}{\partial t} &= \nabla \cdot (D_c(1-v)\nabla c) + \rho c(1-c)\\ \frac{\partial m}{\partial t} &= \nabla^2 m + \delta(c-m)\\ \frac{\partial v}{\partial t} &= v(1-v) - \gamma m v,\end{aligned} \tag{3.6}$$

where D_c is the diffusion coefficient, ρ is the cancer cell proliferation, δ is the H^+ ion production rate and γ is the extracellular matrix degradation rate. Note that the cancer cells undergo a random cell migration through the diffusion term. This model has then been improved and used in similar context. Instead of cell migrating randomly, the so-called reaction-diffusion-taxis model, includes a key mechanism of cell migration, haptotaxis. The cells migrate in response to the gradients of extracellular matrix density. One of the notable work of utilising this type model was done by Anderson et al., where the influence of the extracellular matrix was explicitly modelled in a 2D setting. Other work using the continuous model include [58–61], focusing on cell-cell adhesion. See [62] for a recent review on mathematical modelling of tumour growth.

Reaction-diffusion describing tumour microenvironment Oxygen and nutrients supplied by the vasculature are crucial for cell function and survival [63]. The initial growth of the tumour is limited as it is dependent on diffusion as a way to receiving nutrients and discarding wastes. Reaction-diffusion equations can not only be used to model spatial spread of tumours, it is particularly useful in modelling the supply/consumption of substrates. In [64], Anderson and coworkers proposed a hybrid model detailing the interactions of tumour cells, matrix-degradative enzymes (MDE) and macromolecules (MM). In particular, cell migration and diffusion of oxygen, MDE, and MM form a system of reaction-diffusion-chemotaxis equations. The production, degradation, and diffusion of these substrates in the microenvironment were considered at the level of a single cell.

3.4 Cell-based model

Cell-based models, also known as discrete models, simulate individual cell behaviours within tissue environments. These models have several advantages. Cells can act independent of each other, and each can be parametrised individually to reflect heterogeneity in cancer. In addition, the model is flexible

enough to implement rules dictating cell behaviour and interactions. This allows swift testing of biologic hypotheses by generating simulations with different mathematical rules and compare the outcomes against existing observation.

Cell-based models, as illustrated in Figure 3.2 can be classified by restriction on the placement of cells: Lattice-based model contains cells moving along a rigid grid, while cells can move freely in off-lattice models.

3.4.1 Lattice-based model

Lattice-based models can be applied on regular meshes in 2D or 3D and irregular ones. On regular meshes, it is easy to visualise and to implement, however the dynamics of the cells are highly dependent on the grid. While irregular meshes eliminate the bias, they add complexity to the system.

Cellular automaton (CA) model CA is a type of regular mesh-based model that is discrete in time, space and state. Its earliest appearance can be traced back to modelling pattern formation in biological systems [65]. In its basic form, each lattice grid can hold at most one cell. At each time step, a discrete rule is executed to determine the fate of the cell at each lattice point: it can remain, move to a neighbouring point or divide and put a daughter cell in the neighbour. The formation of the grouping of the cells relies on the sequence of the updates. One way to reduce the effect of update is to carry out asynchronous update, meaning the cells are updated in a random order at each time step.

Lattice gas CA (LGCA) model Similar to CA model, it can model the interplay of cells with themselves and their heterogeneous environment. LGCA models were introduced to model gas and fluid flows through implementing simplistic local collisions [65]. In tradition synchronous-update CA model, implementation movement rules for cells in CA is difficult due to limitation of typically one cell per site. As a result, cells usually clash into each other when two want to move to the same empty site. Rather than tracking the movement of each cell, in LGCA, the number of cells travelling between different sites are tracked. Rest status can also be added for cells that do not move. It therefore allows simulation of very large numbers of cells for longer time frame. The main difficulty is to find rules that describe precisely the collective behaviour of the cells. Simulation of LGCA models allows characterising tumour growth and invasion [66–68].

Cellular Potts models (CPMs) As a type of unstructured lattice-based model, CPM uses multiple lattice grids to resemble the morphologies of individual cells. The model accepts or reject the update, consisting random swapping of two neighbouring sites, based on whether the move would reduce the global energy of the system. It is computationally intensive, however it is favoured for its usefulness in modelling cell morphology.

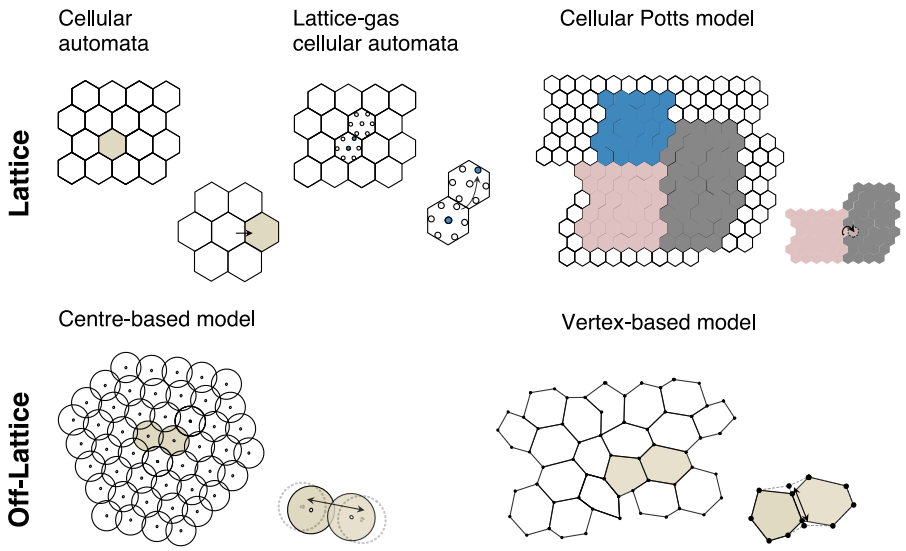

Figure 3.2: A schematic classification of cell-based modelling approaches. In each panel, the initial state of the model is shown on the left, and the dynamic of the update is shown on its immediate right.

3.4.2 Off-lattice model

Off-lattice models draw attention to the modelling of cell movements in space and how the cells evolve according to mechanical and chemical interactions. These models differentiate themselves by their ability of capturing cell shape and thus can be divided roughly into two types, one that focuses on cell volumes, and the other focuses on cell boundaries.

Centre-based models In its simplest form, cells are treated as spheres and the interaction in terms of mechanical forces are tracked. When a change in force exchanged between cell centres are detected, cells' positions are updated accordingly. The basic model can be further simplified by grouping cells into clusters or functional units. The movement of these new units are then simulated with assigned interactions. This reduces the computational overhead and allows incorporating heterogenous conditions into different clusters of cells. On the other hand, greater details can be achieved by modelling interactions between subcellular elements of individual cells. This model is superior in terms of the realism of cell biomechanics, albeit at a high computational cost.

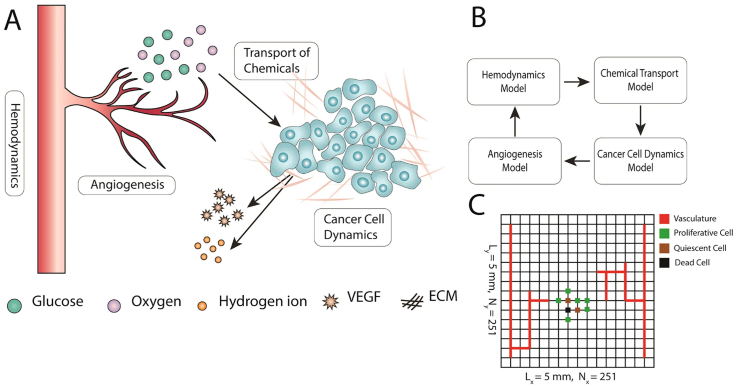

Figure 3.3: Overview of a multi-scale hybrid model of tumor microenvironment to investigate the effect of glucose uptake rate of cancer cells on tumour growth by [69]. Rearranged in layout under CC BY 2.0

Boundary-based models In contrast to centre-based models, boundary-based models track and compute forces that act on points located on the boundary of adjacent cells. Depending on the requirement of spatial resolution, the cell can be modelled as either polygons or bounded fluid. They are the most realistic model and most computationally intensive cell-based methods, however they are proved to be effective tools in linking cell mechanics to fluid and solid tissue mechanics.

3.5 Mathematical model of cancer growth under treatment

3.5.1 Hybrid multi-scale agent-based models

Hybrid models compose a large group of agent-based models that naturally link with tumour biology. They represent cells as individual discrete agents and often use continuous concentration to model extracellular and intracellular environments. It therefore allows for multi-scale integrative approach in modelling of the interaction of many variables both intrinsically and extrinsically. Typically, an on- or off-lattice agent-based model (c.f. Section 3.4), depending on the modelling requirement of cell number and geometry, serves as the basis. Various processes linking cells and their environment can be added in the models in a realistic way by coupling, for example ODEs and PDEs to describe signalling, metabolic pathways of chemical processes (c.f. Section 3.3). See Figure 3.3 for an illustration of this approach.

First appearance of hybrid model, addressing avascular tumour growth appeared in [67]. This model allows tracking of the fate of the cell (LGCA model), as well as explicit modelling of chemical diffusion (nutrient and necrotic signal). This type of square-lattice hybrid model has been used on simulating

avascular [70–73] and vascular tumour growth [74, 75], as well as tumour–microenvironmental interactions [76–79]. Examples on other types of on-lattice model, including hexagonal-lattice CA, Potts models, and off-lattice models can be found in this excellent review in [80].
In the following section, several previously proposed models that inspired this project are highlighted and discussed.

3.5.2 Previously proposed models

Among various complex models in tumour angiogenesis, Byrne and Chaplain [81] accounted for interactions between endothelial cells, angiogenic factors and other cell types and successfully simulated many features of angiogenesis that are observed *in-vivo*. However, many earlier works were unable to distinguish different vascular morphologies. Additionally, vascular remodelling and the impact of blood flow were often neglected. Anderson and Chaplain [82] developed a cellular automaton based model with a system of PDEs to describe the spatio-temporal interactions between tumour cells, endothelial cells, extracellular matrix (ECM) and chemical such as tumour angiogenic factor (TAF). In this continuum-discrete model, flow of blood in the vessel were included, and the movement of vessel tips was modelled via a biased random walk. However the effect of specific pro-angiogenic and anti-angiogenic factors such as VEGFA were not considered in the study due to lack of understanding of the underlying biochemistry.
Alarcón and coworkers [83] have developed a computation framework on a hexagonal lattice that couples functional processes at subcelluar, cellular, and tissue scales. This complex model was used to investigate the importance of interaction between different processes such as blood flood and tumour growth under heterogenous conditions. It was able to simulate irregular growth of tumour and supported the idea of vessel normalisation with anti-angiogenic treatment. Although tissue-vasculature coupling were included, new vessel formation was not explicitly modelled. (the model was later extended in [75] to account for changes in morphology of vessel network induced by regression and angiogenesis.) Additionally neither cell migration or immune response was considered. In [84], a mathematical model coupling tumour growth with immune systems was made for early stage of tumour development. The model includes spatial interactions between tumour and immune cells using chemotaxis terms in the governing PDEs. However the model was limited to the simulation results on radially symmetric spatial growth and evolution of temporal cell species. More recent work by Owen and coworkers [85] seeked to explain combinatory effects of treatment strategies involving standard chemotherapy and novel drug. Specifically the treatment relies on genetically-engineered macrophages to deliver cytotoxic drugs to hypoxic region. To date, hindered by limited availability of data directly linked to the input parameters, only limited number of studies [86] included personalised models that are capable of describing tumour dynamics and predict treatment outcome at an individual level.

Chapter 4

Numerical simulations of chemical concentration

4.1 Numerical approximation of differential equations

Despite of the usefulness of analytic techniques for solving differential equations, they are inherently limited to provide solutions to the simplest models. The rise of modern computing and development of reliable numerical methods facilitate general solving of differential equations. To convert the equations into a computable form understood by machines, a discretisation method is needed to transform the problem into its finite dimensional representation. In particular, several methods are currently in use in solving PDEs, for example the finite difference method (FDM), finite element method (FEM) and the spectral method. There are inevitable loss of information during the transformation due to the reduction of dimension. Different discretisation schemes attempt to address differently the task of replacing the partial differential equation system with algebraic ones while being stable, consistent, accurate yet efficient.

Finite difference method approximates the differential operators using neighbouring points. It can be further divided into explicit and implicit scheme depending on whether the current state of the system was involved during the calculation. In general, explicit method is less stable however easier to implement; while the opposite is true for the implicit ones. A most notable implicit-type scheme is the Crank-Nicolson scheme, it is known to be unconditionally stable at a cost of solving system of equation at each time step. In general, FDM gives optimal solution when it is limited to structured grids; While both FEM and spectral methods are well-suited for problems with complex geometries, for which unstructured grids are needed. These methods approximate the functions in terms of linear combination of basis functions. In addition to its ability to handle complicated geometries, FEM is usually preferred for its stability and convergence in situations where solution lacks smoothness. FEM can be considered as a generalisation of FDM: first order FEM is identical to FDM for Poisson's equation given the discretisation is carried out on a regular rectangular mesh with each rectangle divided into two triangles.

For time-dependent, initial-boundary problems defined in space and time, setups of FDM and FEM are different. While numerical treatment of space and time can be done similarly in FDM with different accuracies, obtaining numerical approximation using FEM requires four stages: pde problem (strong probem), variation formulation (weak form), finite element formulation and the algorithmic

implementation. Temporal approximation can be treated separately using FEM or FDM depending on the requirement of accuracy and stability. FDM and FEM listed above are far from being the only discretising scheme. Spectral method uses globally defined functions rather than piecewise polynomials as approximating functions.

The use of a finite element software library suite such as FEniCS enables the implementation of the problem in near identical notation to their mathematical formulation. It allows efficient computation of finite element matrices while still retaining geometric generality [87]. Its features include but are not limited to: automated solution of variational problems, automated error control and adaptivity, extensive library of finite elements, high performance linear algebra through existing libraries such as PETSc and Trilinos and existence of both python and cpp interfaces. The code written under FEniCS framework stays compact and assembles mathematical formulation when mathematical and computational complexity increases. In the following section, we will formulate solving of general problems discussed in this work in stages using FEniCS library. We refer reader to [88] for complete reading on solving various PDE problem using FEniCS.

4.2 Numerical approximation of PDEs in 2D

4.2.1 Poisson equation

PDE problem The poisson problem is the simplest and the most famous elliptic PDE. It can be described as:

$$-\nabla^2 u(\boldsymbol{x}) = f(\boldsymbol{x}), \qquad \boldsymbol{x} \text{ in } \Omega \tag{4.1}$$

$$u(\boldsymbol{x}) = u_D(\boldsymbol{x}), \qquad \boldsymbol{x} \text{ on } \partial\Omega \tag{4.2}$$

$u = u(\boldsymbol{x})$ is the unknown function, $f = f(\boldsymbol{x})$ is the source/sink function. ∇^2 is the Laplace operator, Ω is the spatial domain and $\partial\Omega$ is the boundary of Ω. Equations (4.1) and (4.2), the PDE and the boundary condition together are called boundary-value problem.

Variation Formulation Formulating the PDE problem into its variational form is the first step towards solving the problem. The variational form is obtained by multiplying the PDE by a test function v such that $v = 0 on \partial\Omega$ and integrating the resulting equation over Ω. In this case, we multiply Equation (4.1) by the test function v and perform integration by parts over Ω, obtaining:

$$-\left(\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x - \int_{\partial\Omega} \frac{\partial u}{\partial n} v \,\mathrm{d}s\right) = \int_\Omega f v \,\mathrm{d}x \tag{4.3}$$

where $\frac{\partial u}{\partial n} = \nabla u \cdot n$ is the derivative of u in the outward normal direction n on the boundary. For a no-flux boundary, meaning there is no flow of substrate in

the the outward direction normal to the boundary, $\frac{\partial u}{\partial n} = 0$. It therefore follows that

$$\int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x = \int_\Omega f v \, \mathrm{d}x \tag{4.4}$$

Given an appropriate choice of test and trial space $\hat{V}$ and V, the variational form can be states as: find $u \in V$ such that

$$a(v, u) = L(v) \tag{4.5}$$

for all $v \in \hat{V}$, where

$$a(v, u) = \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x \tag{4.6}$$

$$L(v) = \int_\Omega f v \, \mathrm{d}x \tag{4.7}$$

Having defined the variational problem, the finite element method for the Poisson equation seeks an approximate solution in a discrete (finite dimensional) subspace $V_h \subset V$ and $\hat{V}_h \subset \hat{V}$. Hence the discrete variational problem becomes: find $u_h \in V_h \subset V$ such that

$$a(v, u_h) = L(v) \quad \forall v \in \hat{V}_h \subset \hat{V} \tag{4.8}$$

This discrete variational problem, together with suitable function space V_h and $\hat{V}_h$, uniquely defines approximate numerical solution of Poisson equation.

FEniCS implementation To solve this linear PDE in FEniCS, one needs to perform two steps: Choose a finite element space V and $\hat{V}$, specifying the domain Ω including the mesh, and the type of function space in terms of polynomial degree and type. Then express the PDE as a diescrete variational problem as Equation (4.8).

4.2.2 Heat equation

PDE problem Heat equation can be seen as a natural extension to the Poisson equation, describing the stationary distribution of heat in a body to a time-dependent problem. The problem can be written as:

$$\frac{\partial u}{\partial t} = \nabla^2 u + f \qquad \text{in } \Omega \times (0, T] \tag{4.9}$$

$$u = u_D \qquad \text{on } \partial\Omega \times (0, T] \tag{4.10}$$

$$u = u_0 \qquad \text{at } t = 0. \tag{4.11}$$

u is an unknown function varies in space and time. The source/sink function f also vary in space and time. The initial condition u_0 is a function in space only.

Variation Formulation For a function that varies in both space and time, we need to consider which discretisation scheme is the best for this particular problem. The simplest approach is to first discretise the time derivative by a finite difference approximation, converting the problem into its corresponding stationary case, then transform the stationary problem into a variational formulation. We deploy θ-method for discretising time derivative: let u^n denotes u at time level n. For $0 \leq \theta \leq 1$ and u^n known from previous time-step, compute u^{n+1} by solving:

$$\frac{u^{n+1} - u^n}{\Delta t} = \theta \left(\nabla^2 u^{n+1} + f^{n+1}\right) + (1-\theta)\left(\nabla^2 u^n + f^n\right) \tag{4.12}$$

This is the time-discrete version of the heat equation Equation (4.11). With $\theta = 1$, we recover the implicit Euler discretisation, $\theta = 0$ gives the explicit Euler scheme, while $\theta = 0.5$ gives the Crank-Nicolson discretisation. For this example, we will derive the weak form with $\theta = \frac{1}{2}$. Rearranging the resulting equation by organising the unknown u^{n+1} and other terms on one side of the equation sign, assuming u^n is knwon from the previous time step:

$$u^0 = u_0, \tag{4.13}$$

$$u^{n+1} - \frac{1}{2}\Delta t \nabla^2 u^{n+1} - u^n + \frac{1}{2}\Delta t \nabla^2 u^n - \frac{1}{2}\Delta t (f^{n+1} + f^n) = 0 \tag{4.14}$$

We then multiply the equation by a test function $v \in \hat{V}$and integrate second-derivatives by parts to turn the equations into weak forms. The resulting weak form from Equation (4.14) can be written in an abstract formulation, letting $u = u^{n+1}$:

$$F_{n+1}(u; v) = 0 \tag{4.15}$$

where

$$F_{n+1}(u; v) = \int_\Omega \left(uv + \frac{1}{2}\Delta t \nabla u \cdot \nabla v - u^n v - \frac{1}{2}\Delta t \nabla u^n \nabla v - \frac{1}{2}\Delta t (f^{n+1} + f^n) v \right) \mathrm{d}x. \tag{4.16}$$

The initial condition also needs to be turned into a variational problem:

$$F_0(u; v) = 0 \tag{4.17}$$

where

$$F_0(u; v) = \int_\Omega (u - u_0) v \, \mathrm{d}x. \tag{4.18}$$

This initial condition can be computed either by projection or interpolation. Both operations are trivial to carry out in FEniCS with a single statement. Having found $u^0 \in V$ such that $F_0(u^0, v) = 0$ holds for all $v \in \hat{V}$, we then find $u^{n+1} \in V$ such that $F_{n+1}(u^{n+1}, v) = 0$ for all $v \in \hat{V}$ for $n = 0, 1, 2, \ldots$

Algorithm 1 Time-stepping algorithm for heat equation

```
Define the boundary condition
Compute u^0 as the projection/interpolation of the given initial value
Define the form of F
set stopping time T
t ← Δt
while t^n ≤ T do
    apply essential boundary conditions
    solve F = 0 for u and store in u
    t ← t + Δt
    u_1 ← u
end while
```

FEniCS implementation Here we detail the time-stepping algorithm for the heat equation, see Algorithm 1.

4.2.3 System of reaction-diffusion equations

PDE problem The following system of PDEs models the chemical reaction between two species A and B, $A + B \rightarrow C$ in some domain Ω:

$$\frac{\partial u_1}{\partial t} = \nabla \cdot (\epsilon_1 \nabla u_1) - K u_1 u_2 + f_1 \tag{4.19}$$

$$\frac{\partial u_2}{\partial t} = \nabla \cdot (\epsilon_2 \nabla u_2) - K u_1 u_2 + f_2 \tag{4.20}$$

$$\frac{\partial u_1}{\partial t} = \nabla \cdot (\epsilon_3 \nabla u_3) + K u_1 u_2 - K u_3 + f_3 \tag{4.21}$$

We assume the reaction is first-order, such that the reaction rate, K, is proportion to the product of concentration of A and B. We also assume that the newly-formed compound C decays with a rate proportional to the concentration of C. $u_1, u_2, u3$ in the PDE above denotes the concentration of A, B and C respectively. A, B and C are also assumed to diffuse with diffusivity $\epsilon_1, \epsilon_2, \epsilon_3$ respectively. Finally, we define the Neumann boundary condition on the field where the chemical reaction takes place, *i.e.* $\frac{\partial u_i}{\partial n} = 0, i = 1, 2, 3$ The initial values of u_1, u_2, u_3 are set to 0 at $t = 0$, and f_1, f_2, f_3 are the respective source/sink terms.

Variation Formulation When forming the variational formulation of system, each of the equation is multiplied by a test function. All equations are then integrated and summed up. As we did in the numerical treatment of heat equation, I first introduce the discretisation in time (using backward Euler in this demonstration.) and approximate the time derivative by $(u_i^{n+1} - u_i^n)/\Delta t$. Let $t_1, t_2,$ and t_3 be the component of a test function. The variational form can

be written as:

$$F(u_1, u_2, u_3; v) = 0 \tag{4.22}$$

where

$$F(u_1, u_2, u_3; v) = \int_\Omega (\Delta t^{-1}(u_1^{n+1} - u_1^n)v_1 + \epsilon_1 \nabla u_1^{n+1} \cdot \nabla v_1)\,\mathrm{d}x \tag{4.23}$$

$$+ \int_\Omega (\Delta t^{-1}(u_2^{n+1} - u_2^n)v_2 + \epsilon_2 \nabla u_2^{n+1} \cdot \nabla v_2)\,\mathrm{d}x \tag{4.24}$$

$$+ \int_\Omega (\Delta t^{-1})(u_3^{n+1} - u_3^n)v_3 + \epsilon_3 \nabla u_3^{n+1} \cdot \nabla v_3)\,\mathrm{d}x \tag{4.25}$$

$$- \int_\Omega (f_1 v_1 + f_2 v_2 + f_3 v_3)\,\mathrm{d}x \tag{4.26}$$

$$- \int_\Omega (-K u_1^{n+1} u_2^{n+1} v_1 - K u_1^{n+1} u_2^{n+1} v_2 + K u_1^{n+1} u_2^{n+1} v_3 - K u_3^{n+1} v_3)\,\mathrm{d}x \tag{4.27}$$

The time-stepping algorithm follows the same setup as heat equation, detailed in Algorithm 1. This simultaneous solving demonstrates the ease of implementing a coupled PDE system fully implicitly using FEniCS.

4.3 Parallelisation of numerical simulations

The multi-core structure of modern computing architecture is designed to provide a practical way of boost computing power. While it promises to bring improvements, it also imposes many programming challenges. Before multi-core processor became the norm, improvement in program performance was dependent on the increase in processor frequency, thus required little effort. Nowadays, scientific computing is usually carried out on a computing node in a cluster, containing many multi-core processors with simpler modules and lower clock speeds. It is therefore vital to master new programming techniques to fully exploit its potential.

Developing parallel programs requires agile thinking, one needs to coordinate work in a parallel way by considering many tasks including resource sharing and task scheduling. Furthermore, investigating the performance of parallel programs is much more difficult than simply implementing a correct one. There exists many paradigms in parallel computations, in which distributed memory, called Message Passing Interface (MPI) and shared memory, called Open Multiprocessing (OpenMP) are most popular. MPI mostly targets a distributed memory system. Each processor can only see memory available to them, as supposed to shared memory, where all processing units can see the memory system. Its main advantage over OpenMP is its scalability per cost provided the programme is correctly implemented. Thus it has become the dominating programming model for highly scalable scientific application over the years.

4.3.1 Parallelising the Finite Element Method

A large scale finite element simulation can be decomposed into four components: mesh partitioning, assembly, solution of linear system and input/output of data. FEniCS seeks to automate these processes by having a general code to solve a large class of problems.
FEniCS handles the mesh partitioning by a method called nodal grid partitioning. It first constructs a graph from the mesh, then the graph is partitioned by methods such as ParMETIS or PSCotch (parallel version of METIS and Scotch). This results in a shared band of border elements. Consequently nodal grid partitioning makes solver such as algebraic multigrid much easier since the matrix row associated with a node is in one place.
Parallel assembly requires little care, as the matrices can be assembled locally; and the input and output of the data can be stored locally then gathered post-process.
Parallelisation for systems requiring precondition remains a problem. Linear algebra of the solver requires some consideration regarding updates of ghost-node (ref), but not all preconditioing schemes can be easily parallelised. Algebraic multigrid shows great promise as a parallell preconditioner for coupled system [89].

Chapter 5

Bayesian Inference on Simulator-based Model

The widespread of affordable computing power and tremendous improvements in computational inference for statistical models open up many opportunities for ambitious researchers to address complex datasets using advanced models.
In Bayesian inference, Markov chain Monte Carlo (MCMC) has been the universal machinery for Bayesian inference for nearly 70 years since early 1990's [90] It has provided a guaranteed convergence to the quantities of interest with minimal assumptions on the targeted distribution compared to the most standard Monte Carlo methods that require direct simulation of the target. It has been continuing to evolve from random walk to Hamiltonian Monte Carlo, each proposing modifications to accelerate MCMC algorithms in the face of high dimensionality in the parameter space. Approximate models and algorithms are at the centre of statistical analysis as they reduce dimension and size while capturing the important aspect of the data.
In this chapter the concept of bayesian inference is introduced, followed by highlighting the progress and issues associated with MCMC. A collection of algorithms called approximate bayesian computation (ABC) is then described in detail among other solutions built upon approximate models and/or summarised versions of the data. Finally applications of ABC in the realm of biostatistics are discussed.

5.1 Bayesian inference with Markov chain Monte Carlo (MCMC)

5.1.1 Bayesian Inference

The likelihood $p(\mathcal{D}|\theta)$ describes how likely the data $\mathcal{D}$ are given fixed parameters θ under the model $m(\theta, \mathcal{D})$. In constrast to frequentist approach, both parameter model and the data are treated as random variables, allowing individuals to specify prior distribution, $p(\theta)$, reflecting personal knowledge and belief on the parameter prior to exposure of data. Bayes theorem (Bayes, Price, 1763) is then applied to get the posterior $p(\theta|\mathcal{D})$:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} \tag{5.1}$$

where $p(\mathcal{D})$ is the normalising constant and it is the integral of the following form over Θ:

$$p(\mathcal{D}) = \int_{\Theta} p(\mathcal{D}|\theta)p(\theta)\,\mathrm{d}\theta \tag{5.2}$$

The level of difficulty in obtaining samples from the posterior distribution $p(\theta|\mathcal{D})$ lies in the complexity of the model $m(\theta, \mathcal{D})$. While a conjugate prior to the likelihood function would guarantee an explicit form for the posterior [91], this choice is seldom used outside of context of toy examples in textbooks. One approach is to focus on implementing efficient procedures in computing $p(\mathcal{D})$ numerically, which does not carry an exact form in relatively high dimension of parameter space. Some numerical integration scheme such as quadrature [92], trapezoidal or Simpson's rule have shown success in relatively low dimension parameter space [93]. Since $p(\mathcal{D})$ is a constant, one can estimate the posterior distribution without having to compute the integral. One of the popular sampling-based approach is Markov chain Monte Carlo (MCMC) algorithms. This approach will be briefly described in the following section.

5.1.2 Basics of MCMC

MCMC procedures [94] involve simulating a Markov chain that explores the posterior target without knowing the shape of the true density. The output generates a correlated output that requires some burnin time to 'forget' the impact of the initial distribution, and such Markov chain is *ergodic*, meaning that it will converge to the posterior distribution no matter where it started at time $t = 0$ [94] .
Here the algorithm of Metropolis-Hasting [95–98] is demonstrated as an example showing the working of MCMC. Given a computable density π (up to a normalising constant) on the paramter space Θ, a proposal, $q(\cdot|\cdot)$, as known as a Markov kernel, a Markov chain can be generated with the proper stationary distribution.

Algorithm 2 Metropolis-Hasting algorithm for generating a stationary distribution

1: Choose initial value $\theta^{(0)}$
2: **for** $n = 1 \dots N$ **do**
3: Generate θ' from the proposal $q(\cdot|\theta^{n-1})$
4: compute the acceptance probability

$$a^{(n)} = min(1, \frac{\pi(\theta')q(\theta^{(n-1)}|\theta')}{\pi(\theta^{n-1})q(\theta'|\theta^{(n-1)})})$$

5: Accept $\theta^n = \theta'$ if $u_n \leq a^{(n)}, u_n \sim Unif(0,1), \theta^n = \theta^{(n-1)}$ otherwise
6: **end for**

5.1.3 Various techniques on improving upon standard MCMC

A large part of the research on MCMC has been on the choice of the proposal $q(\cdot, \cdot)$, as the convergence performance of the algorithm is dependent on the efficiency of q to explore the parameter space.

Another aspect of the Metropolis-Hastings is the ever-growing complexity in the evaluation of the likelihood during the accept-reject step even though the normalising constant is not required. In the advent of 'big data', density can be either too expensive to compute or intractable for complicated models and large datasets. Many researches have been done on designing method using approximations or appropriate estimators to circumvent the problem.

The progress on approaches to accelerate MCMC algorithms is summarised here, and they can be divided roughly in a few categories, some improve knowledge about the target distribution, while other modify the proposal in the algorithm.

Hamiltonian Monte Carlo (HMC) [99], popularised by [100, 101], is a hybrid method that attempts to explore the geometry of the target density before constructing the algorithm. This has now become central in statistical software such as STAN [102]. Scalable MCMC methods attempt to address the burden of large datasets on a single machine. This includes divide-and-conquer approaches [103–105], partitioning the whole dataset into batches and run separate MCMC algorithms on each data batch independently; as well as sub-sampling approaches, aiming at reducing the number of individual datapoint likelihood evaluations. Some efforts are devoted to exact subsampling methods [106, 107], while others use approximate subsampling methods aiming at the target distribution [108, 109] or the acceptance probability [110, 111].

There are also other types of MCMC schemes exploiting the architecture of modern parallel processor. Naive implementation, meaning running several MCMC chains in parallel is theoretically sounding, however achieving stationarity is difficult. There are several attempts found in the literature [112–114]. Other more sophisticated schemes assign tasks of evaluating different parts of the density in different processor and multiplied together at each MCMC step [115–118].

Lastly, to resolve issues with multi-modality in MCMC in high dimensions, meaning Markov chain being 'trapped' due to locally optimal proposals, an array of methods were invented, including but not limited to: tempering techniques [119–125], adaptive MCMC [126–130], multiple try MCMC [131–133] Readers are referred to Robert 2018 for a comprehensive survey on various techniques of MCMC acceleration.

5.2 Approximate Bayesian Computation (ABC) method and others

There has been an increasing interest in applying statistical methods to models that are easy to simulate but impossible to calculate transition probabilities or likelihoods. ABC, along other techniques including variational Bayes [134], empirical likelihood [135] and integrated nested Laplace approximation (INLA)

[136] have shown to be indispensable in the analysis of complex stochastic models. While traditional tools such as Monte Carlo-type simulation became impractical for posterior distribution given infinite computing power [137], these techniques are computational feasible at a cost of introducing bias. This balance between improvement in precision and loss of information is currently at the centre of discussion within the statistics community [138].

5.2.1 ABC

ABC is the approximation of posterior distribution by finding values that produces simulations that are sufficiently close to the observation. It was first introduced in the context of population genetics [139, 140]. ABC only requires a generative model, *i.e*, a model from with data are assumed to be generated, however no assumption are made on the probabilistic features of the model components.

In its simplest form, it can be formulated as follows:

Algorithm 3 Basic rejection sampling of ABC algorithm

1: **for** $n = 1 \dots N$ **do**
2: **repeat**
3: Generate θ' from the prior distribution $\pi(\cdot)$
4: Generate y' from the model $f(\cdot|\theta')$
5: Compute the distance $\rho(\eta(y^0), \eta(y'))$
6: **until** $S(\eta(y^0), S(y')) < \epsilon$
7: Set $\theta_i = \theta'$
8: **end for**

Where $f(\cdot)$ denotes the generative model, $\rho(\cdot|\cdot)$ is a distant metric, such as L_2 norm, and Kullbeck-Leibler divergence, measuring the discrepancy between the two data sets. $S(\cdot)$ is called *summary statistics*, replacing the data with much smaller-dimension summaries. This allows sensible comparison of particular aspects of data instead of expecting a close match between all components of the observation and the simulated data. However finding a low-dimensional sufficient statistics is nearly impossible, and it is a central topic in ABC literature. In practice, domain knowledge and heuristics are used to construct a set of summary statistics that are hoped to be approximately sufficient. For choice of summary statistics, it is referred to see the discussion in the work of [141, 142]. Another difficulty in ABC is to specify the threshold ϵ. This threshold strikes the balance between degrees of bias in the approximations and the speed of convergence. The choice of threshold is achieved by experimenting with precomputed pool of simulations, or tuning the acceptance rate [143]. (see also [144] for a discussion on the use of coverage property for choosing the threshold.)

Beyond the simple rejection sampling scheme, there are three pother popular algorithms that improve upon the performance of the basic ABC algorithm. MCMC ABC algorithm, based on Metropolis-Hastings MCMC and Sequential

Monte Carlo (SMC) ABC, an adaptation of importance sampling [94] attempt to construct proposals that are closer to the posterior; Post-sampling Correct Methods seeks to 'correct' the sample obtained from ABC to be close to the posterior [143, 145]. While different approximation schemes are being developed and tested, with arguments ranging from efficient programming, to avoiding simulations, to having an ability to deal with more complex structures, their drawback is the overall incapacity to assess the amount of approximation involved.

5.3 Likelihood-Free Inference for simulator-based models

Simulator-based models, also called implicit models [146], by definition, are any function that map parameters and some random variables to the data. Parameters in the model are usually expressed as input of a computer programs, governing the interests of the generated data. On the other hand, the random variables are generated by a random number generator, expressing the stochastic variation during the simulation process.

The mapping of the function is flexible in terms of complexity, usually motivated by the need of researchers seeking realistic representation of the data without being hindered by over-simplification of the model.

As the random variable is stochastic, the simulators are not guaranteed to generate the same outcome given the same input parameters. In other words, the distribution of the random variable is implicitly expressed. The probability distribution of this random variable can be formally defined, however the computation of this probability *analytically* is impossible for a complex model. Instead, the outcome of the random variable for a particular simulation can be compared empirically.

There exists a large body of literature on non-Bayesian approaches in addition to Bayesian methods. For examples, readers are referred to methods of simulated moments [147, 148] and indirect inference [149, 150] originated from econometrics.

In the following section one particular approach in recent development of ABC is introduced. Full details of this work can be found in Gutmann and Corander 2016.

5.3.1 Bayesian Optimisation for Likelihood-Free Inference (BOLFI)

BOLFI, proposed [151] aims to improve performance of ABC. In particular, it overcomes the obstacles of continuous evaluation of the (dis)similarity between the simulated outcome to the observation. Under BOLFI, a probabilistic model, *e.g.* Gaussian process, is used as a surrogate model to learn about the relation between the parameters values and the distance. Once the model is learnt, the knowledge is then used to approximate the likelihood for any threshold without further runs of the simulator. It leverages techniques from Bayesian optimisation

[152, 153] to efficiently acquire data in the regions of parameters space where the distance of discrepancy is small.

Chapter 6

Summary of Papers

Paper I – "Towards personalized computer simulation of breast cancer treatment: a multi-scale pharmacokinetic and pharmacodynamic model informed by multi-type patient data" constitutes the foundation of the whole project. This work is to our knowledge the first of its kind in designing a multi-scale mathematical model integrating individual patient data collected in a clinical trial, including histopathology, magnetic resonance imaging, and molecular profiling, and successfully simulating various response of tumour under chemotherapeutic and anti-angiogenic treatments.
The model accounts for dynamics of tumour on a cross-section at three levels: extracellular, intravascular and intracellular. Specifically the growth and death of cancer cells are modelled by a multi-scale hybrid cellular automaton model (c.f. Section 3.5.1) controlled by intracellular and environmental factors described by ordinary and partial differential equations (Sections 3.3.1 and 3.3.2). Vessel dynamic is modelled using a birth-death process whereby the respective probabilities are dependent of the local VEGF concentration, which is controlled internally by hypoxia mechanism and externally by the concentration of anti-angiogenic agent.
Equipped with the model, we successfully simulated and validated against clinical outcome of four patients, selected for their various treatment response on a $200\,\mu m \times 300\,\mu m$ 2D tumour section. In addition, we tested *in-silico* alternative treatment for better outcome. Consequently, possible mechanistic explanations of their treatment outcomes were suggested: cell-cycle specific drugs are highly-effective on highly proliferative tumours, while for tumour with slow cell-cycles, a more frequent but lower dosage would be advantageous; moderate dose of anti-angiogenetic agent would improve outcome of patients with severe hypoxia.
An important contribution of this paper to the literature of personalised cancer medicine is the identification of key parameters that cannot be accurately estimated from available clinical data, namely the chemosensitivity of tumour cells and sensitivity of vessels to VEGF concentration. The quantification of uncertainties in these parameters becomes the central theme of Paper II.

Paper II – "Likelihood-free inference for hybrid cellular automaton models for personalized simulation of breast cancer treatment" is a natural extension of Paper I. It addresses uncertainty quantification in key parameters found in Paper I and prediction in the context of likelihood-free Bayesian inference. One of the major limitation of developing mathematical models for clinical medicine is the accurate tuning of many parameters given observed patients data (c.f. Section 3.5.1). likelihood-free inference methods for simulation-based models, discussed in Section 5.3 were developed for such case which likelihood can not

be written out explicitly. In particular we favour the methods under Bayesian framework for its ease of integrating relevant prior knowledge and interpretability of uncertainties in parameter estimation (Section 5.1.1). Using BOLFI, a variant of approximate Bayesian computation (ABC) technique, we demonstrated that parameters defining the response of individual patient to chemotherapies can be reliably estimated given clinical data hypothetically collected at various relatively sparse interval in early stage of a treatment. Moreover, the distributions for the cancer cell density at end of treatment were able to be predicted, hence inferring the treatment outcome of the patients. This is our attempt at answering the second and the third aim of this project.

Paper III – "Scalable solver for a multiscale model of personalized breast cancer therapy" revisited the second aim of the paper in relation to robust modelling of tumour heterogeneity. It tackles the numerical difficulties faced in Paper I, where only small sections of patient's histological biopsy were simulated. This is especially relevant in precise integration of MRI data, as the spatial-resolution of a voxel is of size $1\,\mathrm{mm}^3$. Compared to the simulation domain in Paper I it is approximately 20 times larger. While continuous models can be solved and scaled efficiently (Chapter 4), hybrid models combining stochastic cellular automaton and continuous models require tailored treatment. In this paper, an efficient and scalable algorithm, built upon our first attempt in Paper I is presented. The algorithm exploits paralleling processing power of cluster computing. It was able to minimise communication between processors during discrete updates and preserve reproducibility of the model. We paid particular attention to a patient with heterogeneous perfusion condition from previous study whose outcome was difficult to simulate. The size of the simulation domain is comparable to core needle biopsy of 1 cm in length. The complex effect of heterogenous perfusion condition gave rise to stark cell growth pattern in the simulated biopsy: in area with low perfusion, cells were hypoxic and seldom killed by chemotherapy; in high perfusion area, chemotherapies were delivered efficiently and rapidly kill cell in local area. This observation coincides with patient's MRI at week 0 and 1, where dense tumour core was found with high cell activity on the edge of the tumour. This work is a significant instrument in providing general framework for future development of software in *in-silico* guided clinical trial for personalized cancer medicine.

Chapter 7

Discussion

7.1 Significance

This project has been undertaken to design a novel mathematical approach for personalised cancer therapy and to evaluate (i) its validity by simulating clinically relevant biopsy sites and (ii) its accuracy by predicting treatment outcome using individual patient data. Three papers have been produced to reflect these aims in the viewpoint of *modelling*, *inference* and *simulation*, thus complementing each other in delivering *in-silico* guided personalised therapy.

Pilot study (Paper I) represents the most important work of this thesis. It demonstrates that, in first principle, a mechanistic mathematical model can capture non-linear and multi-scale behaviours of a vascular tumour under treatment. [154, 41, 155–164] A crucial difference between our approach to existing attempts in personalised treatment optimisation is the incorporation of individual patient data into relevant biologically-specific mechanisms. This establishes a quantitative framework, through *mathematical modelling*, for identification of patient-specific parameters. Some parameters cannot be estimated before start of therapy and are calibrated to meet the observed outcome.
The second aim of the study (Paper II) sets out to bridge the gap between formal *statistical inference* and parameter calibration through simulation. Results from this investigation confirm our findings in the pilot study that (the patient-specific parameters in the models are necessary for predicting individual outcomes of treatment. The research has also illustrated the importance of rich longitudinal data in the very early stage of the trial to improve accuracy of outcome prediction. This new understanding helps to strength our framework in quantifying the impact of uncertainties in parameters of complex multi-scale mechanstic models on cancer treatment prediction. It also adds to the rapidly expanding field of approximate bayesian computation for simulation-based models ([165, 166]) with a new perspective in the context of hybrid cellular automaton model.
Final study (Paper III) extends and refines the framework, by focusing on the computational aspects of our simulator. Improved algorithm can simulate dynamics of cancer cells, in combination with other continuous models, in domains that are hundreds of times larger than in the pilot study, thanks to distributed computing. We take on the challenge of *simulating* a large section of patient's histological biopsy with realistic heterogenous perfusion condition. Results from these simulations advocate the need for efficient implementation of hybrid cell-based model in describing tumour heterogeneity. Our algorithm contributes to the existing computational frameworks of multi-scale model simulation [167–173]. It is our first step towards creating scalable computing framework for multi-scale

hybrid cell-based models based on open-source softwares.

7.2 Limitations and future directions

There are several means of extending works presented in this dissertation, many of which are described in detail in the corresponding discussion section of the articles. I shall instead look beyond individual works and address key issues in relation to the objectives of my research, discuss their impact and provide possible directions for future work.

Finding optimal therapeutical treatment regime One source of weakness in Paper II which could have affected the quantification of uncertainty in the parameters is that, it was based on simulated data. This issue is two-fold: One is the limited availability and resolution of data, the other is the construction of sufficient summary statistics based on the given data. Although the requirement of likelihoods of ABC method is relaxed so that it can be applied to models of arbitrary complexity, successful application of ABC in the past have been centred on population genetics models [143, 174, 175]. In such application, simulation of a few crucial underlying processes are often relatively simple so that simple summary statistics are usually sufficient. In addition, simulation data can be easily validated against observation from *in-vitro* experiments. In contrast, in cancer modelling scenarios, the models are usually complex and computationally demanding even with the help of large computing resources. To our knowledge, we found very few application of ABC methods for inference in agent-based models [166]. Compared to [166], although the current study is based on simulated data, our summary statistics are intuitive and offers insights into the impact of parameters on treatment outcome. Considerably more work needs to be done to explore better summary statistics on the depth of data, such as perfusion, collected at sparse interval.

Generalisation and improvement of numerical framework It is unfortunate that Paper I did not include results of the full biopsy up to the end of the 12 week treatment period. In this work, up to 200 thousand cells were solved sequentially on a system of 3 equations in addition to 4 more pdes. The simulation is computationally expensive and requires parallelisation for large meshes. It would have not been possible without having access to powerful computer cluster. Approximately 20000 CPU hours used for the production of the scalability and simulation results notwithstanding debug and testing time. Furthermore, after running the simulation, the challenge of large data set and post-processing requires a lot of memory and becomes a significant part of the work. Although the results were simulated for the first three weeks of the treatment period, it did substantiate the potential in accelerating personalised therapy simulations.

To speed up the simulations and limit the size of the output there is clearly a need for improved numerical methods, faster algorithms, and efficient strategies for saving only relevant output.

Beyond the 'proof-of-concept' model As a 'proof-of-concept' model, an issue that was not addressed in the study was the omission of certain tumour mechanisms such as cell migration [176], cell clonal expansion, complex mode of actions of antiangeogenic drugs and drug synergy revealed by the molecular essays [7, 177]. Despite its exploratory nature, the current model is capable of capturing fundamental biological processes at an acceptable level of approximation, while offering some insight into possible mechanistic explanation of individual treatment outcomes. Currently, the scope of this work is limited to the analysis of five patients due to limited time and amount of available resources. A natural progression is to validate our model on more patients, and analyse if, given the available data from the patients, a more complex and biologically realistic model would provide more insights for more patients and better predictions at individual level.
Taken together, the presented framework could then be expanded to probe alternative therapies, and the knowledge learnt from these outcomes can be used to set up *in-silico* clinical trials and to serve as a blueprint for future trial designs.

Bibliography

[1] American Cancer Society. *Cancer Treatment & Survivorship Facts & Figures 2016-2017*. Tech. rep. Atlanta: American Cancer Society, 2016.

[2] Cancer Registry of Norway. *Cancer in Norway 2015 - Cancer incidence, mortality, survival and prevalence in Norway*. Tech. rep. Oslo: Cancer Registry of Norway, 2016.

[3] La Thangue, N. B. and Kerr, D. J. "Predictive biomarkers: a paradigm shift towards personalized cancer medicine". In: *Nat Rev Clin Oncol* vol. 8, no. 10 (Aug. 2011), pp. 587–596.

[4] Zardavas, D., Irrthum, A., Swanton, C., and Piccart, M. "Clinical management of breast cancer heterogeneity". In: *Nat Rev Clin Oncol* vol. 12, no. 7 (July 2015), pp. 381–394.

[5] Altrock, P. M., Liu, L. L., and Michor, F. "The mathematics of cancer: integrating quantitative models". In: *Nat Rev Cancer* vol. 15, no. 12 (Dec. 2015), pp. 730–745.

[6] Barbolosi, D., Ciccolini, J., Lacarelle, B., Barlesi, F., and Andre, N. "Computational oncology–mathematical modelling of drug regimens for precision medicine". In: *Nat Rev Clin Oncol* vol. 13, no. 4 (2016), pp. 242–255.

[7] Silwal-Pandit, L., Nord, S., Lippe Gythfeldt, H. von der, Møller, E. K., Fleischer, T., Rødland, E., et al. "The longitudinal transcriptional response to neoadjuvant chemotherapy with and without bevacizumab in breast cancer." In: *Clin Cancer Res* (2017), clincanres–0160.

[8] Norway, C. R. of. *Cancer in Norway 2017 - Cancer incidence, mortality, survival and prevalence in Norway*. Tech. rep. Cancer Registry of Norway, 2018.

[9] Smith, R. A., Duffy, S. W., Gabe, R., Tabar, L., Yen, A. M., and Chen, T. H. "The randomized trials of breast cancer screening: what have we learned?" In: *Radiologic Clinics* vol. 42, no. 5 (2004), pp. 793–806.

[10] Perry, N., Broeders, M., De Wolf, C., Törnberg, S., Holland, R., Von Karsa, L., et al. "European guidelines for quality assurance in breast cancer screening and diagnosis Fourth Edition". In: *Luxembourg: Office for Official Publications of the European Communities* (2006).

[11] Group, N. B. C. *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft*. Tech. rep. Norsk Bryst Cancer Gruppe, 2019.

[12] Sparano, J. A., Paik, S., et al. “Development of the 21-gene assay and its application in clinical practice and clinical trials”. In: *Journal of Clinical Oncology* vol. 26, no. 5 (2008), pp. 721–728.

[13] Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., et al. “A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer”. In: *New England Journal of Medicine* vol. 351, no. 27 (2004), pp. 2817–2826.

[14] Van’t Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., et al. “Gene expression profiling predicts clinical outcome of breast cancer”. In: *nature* vol. 415, no. 6871 (2002), p. 530.

[15] Van De Vijver, M. J., He, Y. D., Van’t Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., et al. “A gene-expression signature as a predictor of survival in breast cancer”. In: *New England Journal of Medicine* vol. 347, no. 25 (2002), pp. 1999–2009.

[16] Bertos, N. R. and Park, M. “Breast cancer—one term, many entities?” In: *The Journal of clinical investigation* vol. 121, no. 10 (2011), pp. 3789–3796.

[17] Weigelt, B. and Reis-Filho, J. S. “Histological and molecular types of breast cancer: is there a unifying taxonomy?” In: *Nature reviews Clinical oncology* vol. 6, no. 12 (2009), p. 718.

[18] Thoresen, S. “Histological grading and clinical stage at presentation in breast carcinoma.” In: *British journal of cancer* vol. 46, no. 3 (1982), p. 457.

[19] Hopton, D., Thorogood, J., Clayden, A., and MacKinnon, D. “Histological grading of breast cancer; significance of grade on recurrence and mortality.” In: *European journal of surgical oncology: the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* vol. 15, no. 1 (1989), pp. 25–31.

[20] Haybittle, J., Blamey, R., Elston, C., Johnson, J., Doyle, P., Campbell, F., et al. “A prognostic index in primary breast cancer”. In: *British journal of cancer* vol. 45, no. 3 (1982), p. 361.

[21] Theissig, F., Kunze, K., Haroske, G., and Meyer, W. “Histological grading of breast cancer: interobserver, reproducibility and prognostic significance”. In: *Pathology-Research and Practice* vol. 186, no. 6 (1990), pp. 732–736.

[22] EBCTG et al. “Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials”. In: *The Lancet* vol. 365, no. 9472 (2005), pp. 1687–1717.

[23] EBCTG et al. “Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials”. In: *The lancet* vol. 378, no. 9793 (2011), pp. 771–784.

[24] Wolff, A. C., Hammond, M. E. H., Hicks, D. G., Dowsett, M., McShane, L. M., Allison, K. H., et al. "Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update". In: *Archives of Pathology and Laboratory Medicine* vol. 138, no. 2 (2013), pp. 241–256.

[25] Scholzen, T. and Gerdes, J. "The Ki-67 protein: from the known and the unknown". In: *Journal of cellular physiology* vol. 182, no. 3 (2000), pp. 311–322.

[26] Polley, M.-Y. C., Leung, S. C., McShane, L. M., Gao, D., Hugh, J. C., Mastropasqua, M. G., et al. "An international Ki67 reproducibility study". In: *Journal of the National Cancer Institute* vol. 105, no. 24 (2013), pp. 1897–1906.

[27] Perou, C. M., Sørlie, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Rees, C. A., et al. "Molecular portraits of human breast tumours". In: *nature* vol. 406, no. 6797 (2000), p. 747.

[28] Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications". In: *Proceedings of the National Academy of Sciences* vol. 98, no. 19 (2001), pp. 10869–10874.

[29] Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., et al. "Supervised risk predictor of breast cancer based on intrinsic subtypes". In: *Journal of clinical oncology* vol. 27, no. 8 (2009), p. 1160.

[30] Norum, J., Andersen, K., and Sørlie, T. "Lessons learned from the intrinsic subtypes of breast cancer in the quest for precision therapy". In: *British Journal of Surgery* vol. 101, no. 8 (2014), pp. 925–938.

[31] Ohnstad, H. O., Borgen, E., Falk, R. S., Lien, T. G., Aaserud, M., Sveli, M. A. T., et al. "Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up". In: *Breast Cancer Research* vol. 19, no. 1 (2017), p. 120.

[32] Drukker, C., Van Tinteren, H., Schmidt, M., Rutgers, E. T., Bernards, R., Vijver, M. van de, et al. "Long-term impact of the 70-gene signature on breast cancer outcome". In: *Breast cancer research and treatment* vol. 143, no. 3 (2014), pp. 587–592.

[33] EBCTG et al. "Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials". In: *The Lancet* vol. 378, no. 9804 (2011), pp. 1707–1716.

[34] EBCTG et al. "Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials". In: *The Lancet* vol. 379, no. 9814 (2012), pp. 432–444.

[35] Sasson, S. and Notides, A. C. "Mechanism of the estrogen receptor interaction with 4-hydroxytamoxifen". In: *Molecular Endocrinology* vol. 2, no. 4 (1988), pp. 307–312.

[36] Jain, R. K. "Normalizing tumor vasculature with anti-angiogenic therapy: a new paradigm for combination therapy". In: *Nature medicine* vol. 7, no. 9 (2001), p. 987.

[37] Jain, R. K. "Normalizing tumor microenvironment to treat cancer: bench to bedside to biomarkers". In: *Journal of Clinical Oncology* vol. 31, no. 17 (2013), p. 2205.

[38] Bergers, G. and Benjamin, L. E. "Angiogenesis: tumorigenesis and the angiogenic switch". In: *Nature reviews cancer* vol. 3, no. 6 (2003), p. 401.

[39] Keener, J. P. and Sneyd, J. *Mathematical physiology*. Vol. 1. Springer, 1998.

[40] Noble, D. "The rise of computational biology". In: *Nature Reviews Molecular Cell Biology* vol. 3, no. 6 (2002), p. 459.

[41] Byrne, H., Alarcon, T., Owen, M., Webb, S., and Maini, P. "Modelling aspects of cancer dynamics: a review". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* vol. 364, no. 1843 (2006), pp. 1563–1578.

[42] Trayanova, N. A. and Boyle, P. M. "Advances in modeling ventricular arrhythmias: from mechanisms to the clinic". In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* vol. 6, no. 2 (2014), pp. 209–224.

[43] Clayton, R. and Panfilov, A. "A guide to modelling cardiac electrical activity in anatomically detailed ventricles". In: *Progress in biophysics and molecular biology* vol. 96, no. 1-3 (2008), pp. 19–43.

[44] Markram, H. "The human brain project". In: *Scientific American* vol. 306, no. 6 (2012), pp. 50–55.

[45] Wang, Z. and Deisboeck, T. S. "Mathematical modeling in cancer drug discovery". In: *Drug discovery today* vol. 19, no. 2 (2014), pp. 145–150.

[46] Simeoni, M., Magni, P., Cammia, C., De Nicolao, G., Croci, V., Pesenti, E., et al. "Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents". In: *Cancer research* vol. 64, no. 3 (2004), pp. 1094–1101.

[47] Tang, L., Ven, A. L. van de, Guo, D., Andasari, V., Cristini, V., Li, K. C., et al. "Computational modeling of 3D tumor growth and angiogenesis for chemotherapy evaluation". In: *PloS one* vol. 9, no. 1 (2014), e83962.

[48] Norton, L. "Cancer stem cells, self-seeding, and decremented exponential growth: theoretical and clinical implications". In: *Breast disease* vol. 29, no. 1 (2008), pp. 27–36.

[49] Sachs, R., Hlatky, L., and Hahnfeldt, P. "Simple ODE models of tumor growth and anti-angiogenic or radiation treatment". In: *Mathematical and Computer Modelling* vol. 33, no. 12-13 (2001), pp. 1297–1305.

[50] Gompertz, B. “XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c”. In: *Philosophical transactions of the Royal Society of London*, no. 115 (1825), pp. 513–583.

[51] Winsor, C. P. “The Gompertz curve as a growth curve”. In: *Proceedings of the National Academy of Sciences of the United States of America* vol. 18, no. 1 (1932), p. 1.

[52] Von Bertalanffy, L. “Quantitative laws in metabolism and growth”. In: *The quarterly review of biology* vol. 32, no. 3 (1957), pp. 217–231.

[53] Hart, D., Shochat, E., and Agur, Z. “The growth law of primary breast cancer as inferred from mammography screening trials data”. In: *British journal of cancer* vol. 78, no. 3 (1998), p. 382.

[54] Benzekry, S., Lamont, C., Beheshti, A., Tracz, A., Ebos, J. M. L., Hlatky, L., et al. “Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth”. In: *PLOS Computational Biology* vol. 10, no. 8 (Aug. 2014), pp. 1–19.

[55] DiStefano 3rd, J. “Noncompartmental vs. compartmental analysis: some bases for choice”. In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* vol. 243, no. 1 (1982), R1–R6.

[56] Herrer, M. A. “Reaction-diffusion systems: A mathematical biology approach”. In: *Cancer Modelling and Simulation*. CRC Press, 2003, pp. 383–436.

[57] Gatenby, R. A. and Gawlinski, E. T. “A reaction-diffusion model of cancer invasion”. In: *Cancer research* vol. 56, no. 24 (1996), pp. 5745–5753.

[58] Holder, A. B., Rodrigo, M. R., and Herrero, M. A. “A model for acid-mediated tumour growth with nonlinear acid production term”. In: *Applied Mathematics and Computation* vol. 227 (2014), pp. 176–198.

[59] Chaplain, M. A. and Lolas, G. “Mathematical modelling of cancer cell invasion of tissue: The role of the urokinase plasminogen activation system”. In: *Mathematical Models and Methods in Applied Sciences* vol. 15, no. 11 (2005), pp. 1685–1734.

[60] Frieboes, H. B., Zheng, X., Sun, C.-H., Tromberg, B., Gatenby, R., and Cristini, V. “An integrated computational/experimental model of tumor invasion”. In: *Cancer research* vol. 66, no. 3 (2006), pp. 1597–1604.

[61] Gerisch, A. and Chaplain, M. “Mathematical modelling of cancer cell invasion of tissue: local and non-local models and the effect of adhesion”. In: *Journal of Theoretical Biology* vol. 250, no. 4 (2008), pp. 684–704.

[62] Enderling, H. and AJ Chaplain, M. “Mathematical modeling of tumor growth and treatment”. In: *Current pharmaceutical design* vol. 20, no. 30 (2014), pp. 4934–4940.

[63] Hanahan, D. and Weinberg, R. A. “The hallmarks of cancer”. In: *cell* vol. 100, no. 1 (2000), pp. 57–70.

[64] Anderson, A. R. "A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion". In: *Mathematical medicine and biology: a journal of the IMA* vol. 22, no. 2 (2005), pp. 163–186.

[65] Deutsch, A., Dormann, S., et al. *Cellular automaton modeling of biological pattern formation*. Springer.

[66] Böttger, K., Hatzikirou, H., Voss-Böhme, A., Cavalcanti-Adam, E. A., Herrero, M. A., and Deutsch, A. "An emerging Allee effect is critical for tumor initiation and persistence". In: *PLoS computational biology* vol. 11, no. 9 (2015), e1004366.

[67] Dormann, S. and Deutsch, A. "Modeling of self-organized avascular tumor growth with a hybrid cellular automaton". In: *In silico biology* vol. 2, no. 3 (2002), pp. 393–406.

[68] Hatzikirou, H., Breier, G., and Deutsch, A. "Cellular automaton modeling of tumor invasion". In: *Computational Complexity: Theory, Techniques, and Applications* (2012), pp. 456–464.

[69] Shamsi, M., Saghafian, M., Dejam, M., and Sanati-Nezhad, A. "Mathematical modeling of the function of warburg effect in tumor microenvironment". In: *Scientific reports* vol. 8, no. 1 (2018), p. 8903.

[70] Spencer, S. L., Gerety, R. A., Pienta, K. J., and Forrest, S. "Modeling somatic evolution in tumorigenesis". In: *PLoS computational biology* vol. 2, no. 8 (2006), e108.

[71] Enderling, H., Hlatky, L., and Hahnfeldt, P. "Migration rules: tumours are conglomerates of self-metastases". In: *British journal of cancer* vol. 100, no. 12 (2009), p. 1917.

[72] Enderling, H., Anderson, A. R., Chaplain, M. A., Beheshti, A., Hlatky, L., and Hahnfeldt, P. "Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics". In: *Cancer research* vol. 69, no. 22 (2009), pp. 8814–8821.

[73] Zhang, L., Strouthos, C. G., Wang, Z., and Deisboeck, T. S. "Simulating brain tumor heterogeneity with a multiscale agent-based model: Linking molecular signatures, phenotypes and expansion rate". In: *Mathematical and computer modelling* vol. 49, no. 1-2 (2009), pp. 307–319.

[74] Stephanou, A., McDougall, S. R., Anderson, A. R., and Chaplain, M. A. "Mathematical modelling of flow in 2D and 3D vascular networks: applications to anti-angiogenic and chemotherapeutic drug strategies". In: *Mathematical and Computer Modelling* vol. 41, no. 10 (2005), pp. 1137–1156.

[75] Owen, M. R., Alarcón, T., Maini, P. K., and Byrne, H. M. "Angiogenesis and vascular remodelling in normal and cancerous tissues". In: *J Math Biol* vol. 58, no. 4-5 (2009), pp. 689–721.

[76] Gatenby, R., Smallbone, K., Maini, P., Rose, F., Averill, J., Nagle, R. B., et al. “Cellular adaptations to hypoxia and acidosis during somatic evolution of breast cancer”. In: *British journal of cancer* vol. 97, no. 5 (2007), p. 646.

[77] Gerlee, P. and Anderson, A. “Evolution of cell motility in an individual-based model of tumour growth”. In: *Journal of theoretical biology* vol. 259, no. 1 (2009), pp. 67–83.

[78] Silva, A. S., Yunes, J. A., Gillies, R. J., and Gatenby, R. A. “The potential role of systemic buffers in reducing intratumoral extracellular pH and acid-mediated invasion”. In: *Cancer research* vol. 69, no. 6 (2009), pp. 2677–2684.

[79] Anderson, A. R., Hassanein, M., Branch, K. M., Lu, J., Lobdell, N. A., Maier, J., et al. “Microenvironmental independence associated with tumor progression”. In: *Cancer research* vol. 69, no. 22 (2009), pp. 8797–8806.

[80] Rejniak, K. A. and Anderson, A. R. “Hybrid models of tumor growth”. In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* vol. 3, no. 1 (2011), pp. 115–125.

[81] Byrne, H. and Chaplain, M. “Mathematical models for tumour angiogenesis: numerical simulations and nonlinear wave solutions”. In: *Bulletin of mathematical biology* vol. 57, no. 3 (1995), pp. 461–486.

[82] Anderson, A. R. and Chaplain, M. “Continuous and discrete mathematical models of tumor-induced angiogenesis”. In: *Bulletin of mathematical biology* vol. 60, no. 5 (1998), pp. 857–899.

[83] Alarcón, T., Byrne, H. M., and Maini, P. K. “A multiple scale model for tumor growth”. In: *Multiscale Modeling & Simulation* vol. 3, no. 2 (2005), pp. 440–475.

[84] Owen, M. R., Byrne, H. M., and Lewis, C. E. “Mathematical modelling of the use of macrophages as vehicles for drug delivery to hypoxic tumour sites”. In: *Journal of theoretical biology* vol. 226, no. 4 (2004), pp. 377–391.

[85] Owen, M. R., Stamper, I. J., Muthana, M., Richardson, G. W., Dobson, J., Lewis, C. E., et al. “Mathematical modeling predicts synergistic antitumor effects of combining a macrophage-based, hypoxia-targeted gene therapy with chemotherapy”. In: *Cancer Res* vol. 71, no. 8 (2011), pp. 2826–2837.

[86] Kolokotroni, E., Dionysiou, D., Veith, C., Kim, Y.-J., Sabczynski, J., Franz, A., et al. “In silico oncology: quantification of the in vivo anti-tumor efficacy of cisplatin-based doublet therapy in non-small cell lung cancer (NSCLC) through a multiscale mechanistic model”. In: *PLoS computational biology* vol. 12, no. 9 (2016), e1005093.

[87] Kirby, R. C., Knepley, M., Logg, A., and Scott, L. R. “Optimizing the evaluation of finite element matrices”. In: *SIAM Journal on Scientific Computing* vol. 27, no. 3 (2005), pp. 741–758.

[88] Langtangen, H. P., Logg, A., and Tveito, A. *Solving PDEs in Python: The FEniCS Tutorial I*. Springer International Publishing, 2016.

[89] Gerstenberger, A. and Tuminaro, R. S. "An algebraic multigrid approach to solve extended finite element method based fracture problems". In: *International Journal for Numerical Methods in Engineering* vol. 94, no. 3 (2013), pp. 248–272.

[90] Gelfand, A. E. and Smith, A. F. "Sampling-based approaches to calculating marginal densities". In: *Journal of the American statistical association* vol. 85, no. 410 (1990), pp. 398–409.

[91] Schlaifer, R. and Raiffa, H. *Applied statistical decision theory*. 1961.

[92] Naylor, J. C. and Smith, A. F. "Applications of a method for the efficient computation of posterior distributions". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* vol. 31, no. 3 (1982), pp. 214–225.

[93] Ausín, M. C. "Quadrature and numerical integration". In: *Wiley StatsRef: Statistics Reference Online* (2014).

[94] Robert, C. and Casella, G. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[95] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. "Equation of state calculations by fast computing machines". In: *The journal of chemical physics* vol. 21, no. 6 (1953), pp. 1087–1092.

[96] Hastings, W. K. "Monte Carlo sampling methods using Markov chains and their applications". In: (1970).

[97] Besag, J., Green, P., Higdon, D., Mengersen, K., et al. "Bayesian computation and stochastic systems". In: *Statistical science* vol. 10, no. 1 (1995), pp. 3–41.

[98] Robert, C. and Casella, G. "A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data". In: *Statistical Science* (2011), pp. 102–115.

[99] Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. "Hybrid monte carlo". In: *Physics letters B* vol. 195, no. 2 (1987), pp. 216–222.

[100] Neal, R. M. *Bayesian learning for neural networks*. Vol. 118. Springer Science & Business Media, 2012.

[101] MacKay, D. J. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[102] Team, S. D. et al. "Stan: A C++ library for probability and sampling". In: *Online: http://mc-stan. org* (2014).

[103] Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., and McCulloch, R. E. "Bayes and big data: The consensus Monte Carlo algorithm". In: *International Journal of Management Science and Engineering Management* vol. 11, no. 2 (2016), pp. 78–88.

[104] Neiswanger, W., Wang, C., and Xing, E. "Asymptotically exact, embarrassingly parallel MCMC". In: *arXiv preprint arXiv:1311.4780* (2013).

[105] Wang, X., Guo, F., Heller, K. A., and Dunson, D. B. "Parallelizing MCMC with random partition trees". In: *Advances in neural information processing systems*. 2015, pp. 451–459.

[106] Andrieu, C., Roberts, G. O., et al. "The pseudo-marginal approach for efficient Monte Carlo computations". In: *The Annals of Statistics* vol. 37, no. 2 (2009), pp. 697–725.

[107] Quiroz, M., Villani, M., and Kohn, R. "Exact subsampling MCMC". In: *arXiv preprint arXiv:1603.08232* (2016).

[108] Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., et al. "Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals". In: ().

[109] Fielding, M., Nott, D. J., and Liong, S.-Y. "Efficient MCMC schemes for computationally expensive posterior distributions". In: *Technometrics* vol. 53, no. 1 (2011), pp. 16–28.

[110] Bardenet, R., Doucet, A., and Holmes, C. "Towards scaling up Markov chain Monte Carlo: an adaptive subsampling approach". In: *International Conference on Machine Learning (ICML)*. 2014, pp. 405–413.

[111] Bardenet, R., Doucet, A., and Holmes, C. "On Markov chain Monte Carlo methods for tall data". In: *The Journal of Machine Learning Research* vol. 18, no. 1 (2017), pp. 1515–1557.

[112] Mykland, P., Tierney, L., and Yu, B. "Regeneration in Markov chain samplers". In: *Journal of the American Statistical Association* vol. 90, no. 429 (1995), pp. 233–241.

[113] Guihenneuc-Jouyaux, C. and Robert, C. P. "Discretization of continuous Markov chains and Markov chain Monte Carlo convergence assessment". In: *Journal of the American Statistical Association* vol. 93, no. 443 (1998), pp. 1055–1067.

[114] Jacob, P. E., O'Leary, J., and Atchadé, Y. F. "Unbiased Markov chain Monte Carlo with couplings". In: *arXiv preprint arXiv:1708.03625* (2017).

[115] Jacob, P., Robert, C. P., and Smith, M. H. "Using parallel computation to improve independent Metropolis–Hastings based estimation". In: *Journal of Computational and Graphical Statistics* vol. 20, no. 3 (2011), pp. 616–635.

[116] Calderhead, B. "A general construction for parallelizing Metropolis- Hastings algorithms". In: *Proceedings of the National Academy of Sciences* vol. 111, no. 49 (2014), pp. 17408–17413.

[117] Martino, L., Elvira, V., Luengo, D., Corander, J., and Louzada, F. "Orthogonal parallel MCMC methods for sampling and optimization". In: *Digital Signal Processing* vol. 58 (2016), pp. 64–84.

[118] Terenin, A., Simpson, D., and Draper, D. "Asynchronous gibbs sampling". In: *arXiv preprint arXiv:1509.08999* (2015).

[119] Geyer, C. J. and Thompson, E. A. "Annealing Markov chain Monte Carlo with applications to ancestral inference". In: *Journal of the American Statistical Association* vol. 90, no. 431 (1995), pp. 909–920.

[120] Marinari, E. and Parisi, G. "Simulated tempering: a new Monte Carlo scheme". In: *EPL (Europhysics Letters)* vol. 19, no. 6 (1992), p. 451.

[121] Neal, R. M. "Sampling from multimodal distributions using tempered transitions". In: *Statistics and computing* vol. 6, no. 4 (1996), pp. 353–366.

[122] Earl, D. J. and Deem, M. W. "Parallel tempering: Theory, applications, and new perspectives". In: *Physical Chemistry Chemical Physics* vol. 7, no. 23 (2005), pp. 3910–3916.

[123] Xie, Y., Zhou, J., and Jiang, S. "Parallel tempering Monte Carlo simulations of lysozyme orientation on charged surfaces". In: *The Journal of chemical physics* vol. 132, no. 6 (2010), 02B602.

[124] Mohamed, L., Calderhead, B., Filippone, M., Christie, M., and Girolami, M. "Population MCMC methods for history matching and uncertainty quantification". In: *Computational Geosciences* vol. 16, no. 2 (2012), pp. 423–436.

[125] Carter, J. and White, D. "History matching on the Imperial College fault model using parallel tempering". In: *Computational Geosciences* vol. 17, no. 1 (2013), pp. 43–65.

[126] Liang, F., Liu, C., and Carroll, R. J. "Stochastic approximation in Monte Carlo computation". In: *Journal of the American Statistical Association* vol. 102, no. 477 (2007), pp. 305–320.

[127] Robert, C. P., Casella, G., and Casella, G. *Introducing monte carlo methods with r*. Vol. 18. Springer, 2010.

[128] Douc, R., Guillin, A., Marin, J.-M., Robert, C. P., et al. "Convergence of adaptive mixtures of importance sampling schemes". In: *The Annals of Statistics* vol. 35, no. 1 (2007), pp. 420–448.

[129] Roberts, G. O. and Rosenthal, J. S. "Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms". In: *Journal of applied probability* vol. 44, no. 2 (2007), pp. 458–475.

[130] Saksman, E., Vihola, M., et al. "On the ergodicity of the adaptive Metropolis algorithm on unbounded domains". In: *The Annals of applied probability* vol. 20, no. 6 (2010), pp. 2178–2203.

[131] Liu, J. S., Liang, F., and Wong, W. H. "The multiple-try method and local optimization in Metropolis sampling". In: *Journal of the American Statistical Association* vol. 95, no. 449 (2000), pp. 121–134.

[132] Bédard, M., Douc, R., and Moulines, E. "Scaling analysis of multiple-try MCMC methods". In: *Stochastic Processes and their Applications* vol. 122, no. 3 (2012), pp. 758–786.

[133] Martino, L. “A review of multiple try MCMC algorithms for signal processing”. In: *Digital Signal Processing* vol. 75 (2018), pp. 134–152.

[134] Jaakkola, T. S. and Jordan, M. I. “Bayesian parameter estimation via variational methods”. In: *Statistics and Computing* vol. 10, no. 1 (2000), pp. 25–37.

[135] Owen, A. B. *Empirical likelihood*. Chapman and Hall/CRC, 2001.

[136] Rue, H., Martino, S., and Chopin, N. “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. In: *Journal of the royal statistical society: Series b (statistical methodology)* vol. 71, no. 2 (2009), pp. 319–392.

[137] Green, P. J., Dryden, I., and Kent, J. “MAD-Bayes matching and alignment for labelled and unlabelled configurations”. In: *Geometry Driven Statistics* vol. 121 (2015), p. 377.

[138] Marin, J.-M., Pillai, N. S., Robert, C. P., and Rousseau, J. “Relevant statistics for Bayesian model choice”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* vol. 76, no. 5 (2014), pp. 833–859.

[139] Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. “Inferring coalescence times from DNA sequence data”. In: *Genetics* vol. 145, no. 2 (1997), pp. 505–518.

[140] Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. “Population growth of human Y chromosomes: a study of Y chromosome microsatellites.” In: *Molecular biology and evolution* vol. 16, no. 12 (1999), pp. 1791–1798.

[141] Blum, M. G., Nunes, M. A., Prangle, D., Sisson, S. A., et al. “A comparative review of dimension reduction methods in approximate Bayesian computation”. In: *Statistical Science* vol. 28, no. 2 (2013), pp. 189–208.

[142] Fearnhead, P. and Prangle, D. “Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* vol. 74, no. 3 (2012), pp. 419–474.

[143] Beaumont, M. A., Zhang, W., and Balding, D. J. “Approximate Bayesian computation in population genetics”. In: *Genetics* vol. 162, no. 4 (2002), pp. 2025–2035.

[144] Prangle, D., Fearnhead, P., Cox, M. P., Biggs, P. J., and French, N. P. “Semi-automatic selection of summary statistics for ABC model choice”. In: *Statistical applications in genetics and molecular biology* vol. 13, no. 1 (2014), pp. 67–82.

[145] Blum, M. G. “Approximate Bayesian computation: A nonparametric perspective”. In: *Journal of the American Statistical Association* vol. 105, no. 491 (2010), pp. 1178–1187.

[146] Diggle, P. J. and Gratton, R. J. “Monte Carlo methods of inference for implicit statistical models”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* vol. 46, no. 2 (1984), pp. 193–212.

[147] McFadden, D. “A method of simulated moments for estimation of discrete response models without numerical integration”. In: *Econometrica: Journal of the Econometric Society* (1989), pp. 995–1026.

[148] Pakes, A. and Pollard, D. “Simulation and the asymptotics of optimization estimators”. In: *Econometrica: Journal of the Econometric Society* (1989), pp. 1027–1057.

[149] Gourieroux, C., Monfort, A., and Renault, E. “Indirect inference”. In: *Journal of applied econometrics* vol. 8, no. S1 (1993), S85–S118.

[150] Heggland, K. and Frigessi, A. “Estimating functions in indirect inference”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* vol. 66, no. 2 (2004), pp. 447–462.

[151] Gutmann, M. U. and Corander, J. “Bayesian optimization for likelihood-free inference of simulator-based statistical models”. In: *The Journal of Machine Learning Research* vol. 17, no. 1 (2016), pp. 4256–4302.

[152] Jones, D. R. “A taxonomy of global optimization methods based on response surfaces”. In: *Journal of global optimization* vol. 21, no. 4 (2001), pp. 345–383.

[153] Brochu, E., Cora, V. M., and De Freitas, N. “A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning”. In: *arXiv preprint arXiv:1012.2599* (2010).

[154] Sandström, M., Freijs, A., Larsson, R., Nygren, P., Fjällskog, M., Bergh, J., et al. “Lack of relationship between systemic exposure for the component drug of the fluorouracil, epirubicin, and 4-hydroxycyclophosphamide regimen in breast cancer patients.” In: *Journal of clinical oncology* vol. 14, no. 5 (1996), pp. 1581–1588.

[155] You, B., Meille, C., Barbolosi, D., Tranchand, B., Guitton, J., Rioufol, C., et al. “A mechanistic model predicting hematopoiesis and tumor growth to optimize docetaxel+ epirubicin (ET) administration in metastatic breast cancer (MBC): Phase I trial”. In: *J Clin Oncol* vol. 25, no. 18_suppl (2007), pp. 13013–13013.

[156] Kronik, N., Kogan, Y., Elishmereni, M., Halevi-Tobias, K., Vuk-Pavlovi, S., and Agur, Z. “Predicting outcomes of prostate cancer immunotherapy by personalized mathematical models”. In: *PLoS One* vol. 5, no. 12 (Dec. 2010), e15482.

[157] Pascal, J., Bearer, E. L., Wang, Z., Koay, E. J., Curley, S. A., and Cristini, V. “Mechanistic patient-specific predictive correlation of tumor drug response with microenvironment and perfusion measurements”. In: *Proc Natl Acad Sci U S A* vol. 110, no. 35 (Aug. 2013), pp. 14266–14271.

[158] Alfonso, J., Köhn-Luque, A., Stylianopoulos, T., Feuerhake, F., Deutsch, A., and Hatzikirou, H. "Why one-size-fits-all vaso-modulatory interventions fail to control glioma invasion: in silico insights". In: *Sci Rep* vol. 6 (2016), p. 37283.

[159] Robertson-Tessi, M., Gillies, R. J., Gatenby, R. A., and Anderson, A. R. "Impact of metabolic heterogeneity on tumor growth, invasion, and treatment outcomes". In: *Cancer Res* vol. 75, no. 8 (2015), pp. 1567–1579.

[160] Weis, J. A., Miga, M. I., Arlinghaus, L. R., Li, X., Abramson, V., Chakravarthy, A. B., et al. "Predicting the Response of Breast Cancer to Neoadjuvant Therapy Using a Mechanically Coupled Reaction-Diffusion Model". In: *Cancer Res* vol. 75, no. 22 (Nov. 2015), pp. 4697–4707.

[161] Prokopiou, S., Moros, E. G., Poleszczuk, J., Caudell, J., Torres-Roca, J. F., Latifi, K., et al. "A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation". In: *Radiat Oncol* vol. 10 (July 2015), p. 159.

[162] Jackson, P. R., Juliano, J., Hawkins-Daarud, A., Rockne, R. C., and Swanson, K. R. "Patient-specific mathematical neuro-oncology: using a simple proliferation and invasion tumor model to inform clinical practice". In: *Bull Math Biol* vol. 77, no. 5 (May 2015), pp. 846–856.

[163] Cook, L. M., Araujo, A., Pow-Sang, J. M., Budzevich, M. M., Basanta, D., and Lynch, C. C. "Predictive computational modeling to define effective treatment strategies for bone metastatic prostate cancer". In: *Sci Rep* vol. 6 (2016), p. 29384.

[164] Mollard, S., Ciccolini, J., Imbs, D. C., El Cheikh, R., Barbolosi, D., and Benzekry, S. "Model driven optimization of antiangiogenics + cytotoxics combination: application to breast cancer mice treated with bevacizumab + paclitaxel doublet leads to reduced tumor growth and fewer metastasis". In: *Oncotarget* vol. 8, no. 14 (Apr. 2017), pp. 23087–23098.

[165] Jagiella, N., Rickert, D., Theis, F. J., and Hasenauer, J. "Parallelization and high-performance computing enables automated statistical inference of multi-scale models". In: *Cell systems* vol. 4, no. 2 (2017), pp. 194–206.

[166] Sottoriva, A. and Tavaré, S. "Integrating Approximate Bayesian Computation with Complex Agent-Based Models for Cancer Research". In: *Proceedings of COMPSTAT 2010*. Physica-Verlag HD, 2010, pp. 57–66.

[167] Starruß, J., Back, W. de, Brusch, L., and Deutsch, A. "Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology". In: *Bioinformatics* vol. 30, no. 9 (2014), pp. 1331–1332.

[168] Swat, M. H., Thomas, G. L., Belmonte, J. M., Shirinifard, A., Hmeljak, D., and Glazier, J. A. "Multi-scale modeling of tissues using CompuCell3D". In: *Methods in cell biology*. Vol. 110. Elsevier, 2012, pp. 325–366.

[169] Ghaffarizadeh, A., Heiland, R., Friedman, S. H., Mumenthaler, S. M., and Macklin, P. "PhysiCell: an open source physics-based cell simulator for 3-D multicellular systems". In: *PLoS computational biology* vol. 14, no. 2 (2018), e1005991.

[170] Hoehme, S. and Drasdo, D. "A cell-based simulation software for multicellular systems". In: *Bioinformatics* vol. 26, no. 20 (2010), pp. 2641–2642.

[171] Mirams, G. R., Arthurs, C. J., Bernabeu, M. O., Bordas, R., Cooper, J., Corrias, A., et al. "Chaste: an open source C++ library for computational physiology and biology". In: *PLoS computational biology* vol. 9, no. 3 (2013), e1002970.

[172] Kang, S., Kahan, S., McDermott, J., Flann, N., and Shmulevich, I. "Biocellion: accelerating computer simulation of multicellular biological system models". In: *Bioinformatics* vol. 30, no. 21 (2014), pp. 3101–3108.

[173] Cytowski, M. and Szymanska, Z. "Large-scale parallel simulations of 3d cell colony dynamics". In: *Computing in Science & Engineering* vol. 16, no. 5 (2014), pp. 86–95.

[174] Marjoram, P. and Tavaré, S. "Modern computational approaches for analysing molecular genetic variation data". In: *Nature Reviews Genetics* vol. 7, no. 10 (2006), p. 759.

[175] Sipola, A., Corander, J., and Marttinen, P. "Bacmeta: simulator for genomic evolution in bacterial metapopulations". In: *Bioinformatics* vol. 34, no. 13 (Feb. 2018), pp. 2308–2310.

[176] Chaplain, M. A. and Lolas, G. "Mathematical modelling of cancer invasion of tissue: dynamic heterogeneity." In: *NHM* vol. 1, no. 3 (2006), pp. 399–439.

[177] Kristian, A., Revheim, M. E., Qu, H., Mælandsmo, G. M., Engebråten, O., Seierstad, T., et al. "Dynamic 18F-FDG-PET for monitoring treatment effect following anti-angiogenic therapy in triple-negative breast cancer xenografts". In: *Acta Oncol* vol. 52, no. 7 (2013), pp. 1566–1572.

Papers

Paper I

Towards personalized computer simulation of breast cancer treatment: a multi-scale pharmacokinetic and pharmacodynamic model informed by multi-type patient data

Xiaoran Lai, Oliver M. Geier, Thomas Fleischer, Øystein Garred, Elin Borgen, Simon W. Funke, Surendra Kumar, Marie E. Rognes, Therese Seierstad, Anne-Lise Børresen-Dale, Vessela N. Kristensen, Olav Engebråten, Alvaro Köhn-Luque, Arnoldo Frigessi

Cancer Research, to appear. 2019 DOI: 10.1158/0008-5472.CAN-18-1804

Paper II

Likelihood-free inference for hybrid cellular automaton models for personalized simulation of breast cancer treatment

Xiaoran Lai, Henri Pesonen, Alvaro Köhn-Luque, Samuel Kaski, Jukka Corander, Arnoldo Frigessi

in preparation

II

Paper III

Scalable solver for a multi-scale model of personalized breast cancer therapy

Xiaoran Lai, Håkon Taskén, Simon W. Funke, Arnoldo Frigessi, Marie E. Rognes, Alvaro Köhn-Luque

in prepartion

III