Effects of combination treatment on a heterogeneous cancer cell population by computational and mathematical modelling

Van Thuy Truong

The University of Leeds

Department of Mathematics

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

February 2024

The candidate confirms that the work submitted is her own except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below.

The candidate confirms that appropriate credit has been given where reference has been made to the work of others.

Joint publications

The work in chapter 3 is based on the publication:

- Truong VT, Baverel PG, Lythe GD, Vicini P, Yates JW, Dubois VF. Step-by-step comparison of ordinary differential equation and agent-based approaches to pharmacokinetic-pharmacodynamic models. CPT: pharmacometrics & systems pharmacology. 2022 Feb;11(2):133-48.

The work in chapter 4 is based on the publication:

- Truong VT, Lythe GD, Vicini P, Yates JW, Dubois VF. Analysis of stochastic pharmacodynamic models: time to extinction of a heterogeneous cell population (submitted to the Journal of Pharmacokinetics and Pharmacodynamics, February 2024)

The work in chapter 5 is based on the publication:

- Truong VT, Lythe GD, Vicini P, Yates JW, Dubois VF. A hybrid PKPD agent-based model of the tumour immune interaction: effects of anti-cancer combination therapy with PD1 antibody, docetaxel, radiation, and DNA damage response inhibitors (submitted to the Journal of Pharmacokinetics and Pharmacodynamics, February 2024)

The work in chapter 6 is based on the publication:

- Truong VT, Lythe GD, Vicini P, Yates JW, Dubois VF. Combining a hybrid PKPD agent-based model with a reinforcement learning algorithm to optimize cancer combination treatment (submitted to CPT: pharmacometrics & systems pharmacology, February 2024)

Acknowledgements

This project was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764698.

First and foremost, I would like to thank my current and past academic supervisors Grant Lythe, Martín López-García and Carmen Molina-París for their continuous support and inspiring enthusiasm for mathematical biology. I am very appreciative of their patience and their guidance has been invaluable throughout my PhD studentship. Coming from a pharmacology background, I am grateful for the opportunity to learn about different mathematical methods and their application to model biological systems.

I want to give thanks to my past and current industrial supervisors Vincent Dubois, James Yates, Paolo Vicini, and Paul Baverel for their guidance and advice throughout the years.

I would like to express my gratitude to Luis Fonseca, Helen Moore and Reinhard Laubenbacher at the University of Florida for letting me have a secondment at their group and for teaching me methods to analyze agent based models.

Thank you for all the opportunities I was given during the PhD, the QuanTII program, the conferences and the secondment.

I want to thank the current and past PhD students of the mathematical biology group and the School of Maths in Leeds for their technical help, friendship and emotional support. It was a blessing to have them. They make my PhD journey and stays at Leeds enjoyable. I am also grateful to Jessica Brennan for organizing QuanTII meetings and my travels to Leeds.

Finally, I would like to thank my family and friends for their support. Without them, this wouldn't be possible.

Abstract

Mathematical models are widely used in cancer drug development to aid understanding the mechanism of action by simulating administration, distribution, metabolism, elimination, and beneficial and adverse effects. In pharmacometrics, ordinary differential equations are commonly used. However, this approach can have limitations when it comes to simulating heterogeneity, spatial distribution, and stochasticity. For example, in oncology understanding the spatial distribution and heterogeneity of the cells is crucial for successful treatment. Therefore, in this thesis different methods such as agent-based models and stochastic models and their combination with ordinary differential equation, and partial differential equations are explored and applied to model heterogeneity of cancer cell populations, the effect of different therapeutic options such as mono and combination therapy and the tumour immune cell interaction. In addition, the reinforcement algorithm is employed to optimise a complex system and to find an optimal treatment given a specific tumour with its parameters and immune micro-environment.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Dose prediction and optimisation have a rich historical background. In 1847 in England, Buchanan conducted calculations on the inhalation, exhalation, and retention of the short-acting anesthetic ether during induction (Buchanan 1847). In 1913, Michaelis and Menten described enzyme kinetics in Germany. Their equation, known as the Michaelis-Menten equation, is widely used in pharmacokinetics to model the elimination kinetics of different drugs (Michaelis, Menten et al. 1913). Significant contributions continued in the early 20th century. In 1924, Widmark and Tandberg published equations about the accumulation of drug in a one-compartment system (Widmark and Tandberg 1924). Later, in 1937, Teorell presented two papers on a two-compartment model (Teorel 1937; Teorell 1937).

A pivotal figure in the advancement of pharmacokinetics was Friedrich Hartmut Dost, a German pediatrician who utilised mathematical calculations to optimise drug doses in children. In 1953, he introduced the term "pharmacokinetics" in his book, 'Der Blutspiegel: Kinetik der Konzentrationsablaufe in der Kreislaufflussigkeit' (Gladtke 1985). According to Gibaldi and Levy (Levy and Gibaldi 1972), the term is defined as follows:

> Pharmacokinetics is concerned with the study and characterization of the time course of drug absorption, distribution, metabolism, and excretion, and with the relationship of these processes to the intensity and time course of therapeutic and adverse effects of drugs. It involves the application of mathematical and biochemical techniques in a physiologic and pharmacologie context. Pharmacokinetic information can provide a reasonable basis for the design of

> dosage regimens; it can indicate if dosage adjustment may be necessary in renal impairment, and it may also provide preliminary indications of the likelihood and types of drug interactions that may be encountered.

Between 1950 and 1970, significant progress occurred in the field of pharmacokinetics, primarily driven by advancements in bioanalytical techniques used to analyze drug concentrations in plasma, urine, and other tissues. Several notable researchers made valuable contributions during this period. Among them were Sidney Riegelman, Bernard B. Brodie, Eino Nelson, Gerhard Levy, and John Wagner from the United States, as well as Ekkehard Kruger-Thiemer, E.J. Ariens, and Jacques van Rossum from Europe (Paalzow 1995). During this time, the concept of pharmacokinetics/pharmacodynamics (PK/PD) emerged, connecting measured drug exposure data from laboratory experiments to the steady-state pharmacological responses. This linkage provided insights into the relationship between drug concentration and its effects (pharmacodynamics or PD).

In the subsequent decade, the field witnessed the development of econometric and biometric methods, along with the introduction of mixed effect modelling, which aided in quantifying patterns observed in observational data. Lewis Sheiner and Stuart Beal played a significant role in this period by creating the NONMEM software system, which enabled population pharmacokinetic studies (Sheiner and Beal 1983). They are widely regarded as the pioneers of pharmacometrics, a scientific discipline that aims to interpret and describe pharmacology in a quantitative manner (Usman et al. 2023; Gobburu 2010; Ette and Williams 2007). During the period of 1980-1990, regulatory agencies began endorsing the utilization of pharmacometrics in practice. The United States Food and Drug Administration (FDA) specifically called for pharmacological screening of drugs intended for use in elderly patients, aiming to understand the various sources of variability. In 1999, regulatory guidance was issued regarding population analysis, followed by another guidance in 2003 focusing on exposure-response relationships. Those guidelines can be found on www.fda.gov. Due to these milestones, pharmacometrics has a high-level impact on decision-making processes related to clinical trial design, drug development, approval, labeling, and the design of drug dosage regimens, and therefore enhance the rationalization of pharmacotherapy in clinical settings. At this point in time, labeling statements related to intrinsic and extrinsic factors started to gain support from pharmacometric analyses (Gobburu 2010).

A classical PK model typically consists of a central compartment representing plasma, which is connected to one or more peripheral compartments through rate constants. The model parameters, when expressed in terms of rate constants, generally lack physiological significance. However, they can be transformed to more interpretable PK descriptors such as clearance and volume of distribution. Clearance refers to the volume of plasma from which the drug is eliminated per unit time through metabolic or excretion processes. Volume of distribution is defined as the volume of plasma required to contain the total amount of drug in the body at the observed concentration in the plasma. Both clearance and volume of distribution are used to calculate the effective half-life of a drug. These PK models offer a concise and standardised representation of both preclinical and clinical experimental findings. In preclinical settings, PK parameters derived from various *in vivo* studies serve multiple purposes. They can aid in the ranking of compounds for further investigation and be linked to physicochemical, *in vitro*, or structural properties to optimise the PK characteristics of new compounds. In clinical settings, a comparison of PK parameters among different subjects can potentially reveal correlations with demographic characteristics. In addition, PK parameters obtained from one study design can be employed to simulate plasma concentrations for different doses or dosing regimens. Population PK models are frequently employed to describe the variability in drug concentrations and PK parameters within the target patient population. These models play a crucial role in guiding initial dose selection and personalised dosing for specific patient subgroups. Demographic and physiological variables, such as body weight and metabolic functions, are often considered as covariates due to their influence on dose-concentration relationships (Rowland and Tozer 2005).

However, it is important to note that these empirical approaches have limitations in fully incorporating all available prior information regarding the drug and physiology. Consequently, their ability to predict PK for a similar drugs or extrapolate the PK to different physiological conditions may be limited. In more recent times, physiologically-based pharmacokinetic (PBPK) models have emerged as a complement to PKPD models. These models utilise a similar mathematical framework but are parameterised based on known physiology, incorporating a larger number of compartments that represent various organs or tissues in the body. These compartments are interconnected by flow rates that mimic the circulation of blood. Similar to empirical models, PBPK models estimate common PK parameters, e.g., clearance, volume of distribution, and effective half-life.

Due to the more detailed physiological representation, PBPK models offer a quantitative mechanistic framework that enables the prediction of plasma and tissue concentration-time profiles for new drugs after intravenous or oral administration. This is achieved by utilizing scaled drug-specific parameters through *in vitro-in vivo* extrapolation techniques. Another advantage of PBPK models is their ability to extrapolate doses from healthy volunteers to individuals in disease populations, provided the relevant physiological properties of the target population are available. For instance, if data reveals reduced cytochrome P450 expression in patients with chronic kidney disease, such information can be incorporated into a PBPK model, along with other affected parameters. This allows to predict dose adjustments for a drug in a special population relative to that in healthy volunteers (H. Jones and Rowland-Yeo 2013).

Due to decline in productivity in pharmaceutical research and development, there is a need to enhance the characterisation, understanding, and prediction pharmacological modulation of biological targets in a quantitative manner (Graaf 2012). This is emphasised in a White Paper by the National Institutes of Health Quantitative Systems Pharmacology workshop group, which highlights the increasing interest in Systems Pharmacology among academic, regulatory, and industrial researchers (Sorger et al. 2011). Quantitative Systems Pharmacology (QSP) aims to integrate comprehensive knowledge of biological processes into pharmacokinetic/pharmacodynamic and physiologically-based pharmacokinetic models. By applying concepts from systems engineering, computational modelling, and mathematical modelling to describe the dynamic interactions among various components of a biological system, QSP establishes the interface between Pharmacometrics and Systems Biology. QSP models provide a mechanistically-oriented form of drug and disease modelling by integrating features of the drug (dose, dosing regimen, exposure or concentration at target site, potency, or a full pharmacokinetic sub-model) with target biology, downstream effectors at molecular, cellular and pathophysiological levels, and possibly functional effector(s) of interest, such as a physiologically-based pharmacodynamic study endpoint. QSP models are often used to generate hypotheses and support a quantitative understanding of novel compound mechanism(s) of action, in a specific tissue, disease, non-clinical experimental or clinical patient population context. QSP may further be used in optimizing doses and dosing regimens, or in support of dose sequencing decisions for drug combinations, given a QSP model typically contains multiple effectors and at least one PD marker of interest – often the PD endpoint in a given study – down-

stream of the drug or compound target. By understanding the behavior of the system as a whole, rather than focusing solely on individual constituents, it is a holistic approach rather than a reductionist one (Graaf 2012; Knight-Schrijver et al. 2016; Helmlinger et al. 2019). As Anderson suggest in this paper 'More is Different' (P. W. Anderson 1972), a complex system might show behaviour that cannot be explained by the law that governs its microscopic components. This indicates that in complex systems emergent behaviour exist. Hence, even if a theory could explain all microscopic interactions, additional insights are required to understand the macroscopic level.

Cancer progression and treatment modelling present significant challenges due to the complexities of neoplastic disease. The hallmarks of cancer, such as sustained proliferative signalling, resistance to growth suppressors, evasion of cell death, replicative immortality, angiogenesis, and invasion/metastasis, genome instability, resulting heterogeneity, reprogramming of energy metabolism and inflammation, provides cancer cells the ability to evade immune surveillance and archieve resistance to drug treatment. In addition to cancer cells, tumours consist of a diverse array of recruited normal cells that contribute to the tumour micro environment (Hanahan and Weinberg 2011). Available models to describe cancer growth are the exponential model, the Mendelsohn model, the logistic model, the linear model, the surface model, the Gompertz model and the Bartalanffy model (Murphy et al. 2016; Tabassum et al. 2019). Those models consist of an ordinary differential equation (ODE) describing the change in tumour population size V with different terms for example for growth, growth limitation and death. The exponential model provides the description of a cancer cell splitting into two daughter cells with a growth depending proportionally on the cancer population size. The Mendelsohn model is a generalization of the exponential model with the growth term being proportional to a power of the population size. The logistic model has the addition of growth being limited by a carrying capacity. The Gompertz model is able to describe the decrease of tumour growth rate once the cancer size has increased by being a generalisation of logistic model with a sigmodial curve. In contrast, the Bertalanffy model relates the tumour volume and growth and cell death to the surface area (Murphy et al. 2016; Tabassum et al. 2019).

All those models generalise the tumour development by only describing the size of the cancer mass but exclude complex processes for example interaction with the tumour micro environment (TME), immune system, intracellular processes, or interplay of different therapies. The emergence of immunotherapies, such as cytotoxic T-lymphocyte-associated

Model name	Equation
Exponential model	$\frac{dV}{dt} = aV$
Mendelsohn model	$\frac{dV}{dt} = aV^b$
Logistic model	$\frac{dV}{dt} = aV(1 - \frac{V}{b})$
Linear model	$\frac{dV}{dt} = \frac{aV}{V+b}$
Surface model	$\frac{dV}{dt} = \frac{aV}{(V+b)^{\frac{1}{3}}}$
Gompertz model	$\frac{dV}{dt} = aVln\frac{b}{(V+c)}$
Bertalanffy model	$\frac{dV}{dt} = aV^{\frac{2}{3}} - bV$

Table 1.1: ODE based cancer growth models (Murphy et al. 2016; Tabassum et al. 2019)

protein 4, programmed cell death protein 1 (PD-1), or programmed death-ligand 1 (PD-L1) immune checkpoint inhibitors, has introduced additional mechanistic challenges. These therapies involve multiple immune biology events occurring at molecular and cellular levels, interacting dynamically with tumour biology and growth in various bodily compartments, including tumour tissue. Additionally, there are numerous potential targets in immune biology that can potentially induce anti tumour effects. The complex and dynamic nature of these interactions creates a highly dynamic, multi compartment, multi scale environment, making immuno-oncology an attractive field for QSP modelling. A systems approach is necessary to address critical questions regarding dosage, scheduling, and sequencing of treatments, particularly in combination therapies. Furthermore, predicting anti tumour responses, both through indirect (immune mediated) and direct tumour cell killing, and identifying drug combinations that can overcome intrinsic or acquired treatment resistance are crucial challenges in this context (Helmlinger et al. 2019). To address those challenges, various immuno-oncology QSP models exist which usually consist of a system of ODEs with a large number of equations. Examples are the following: The model by Coletti et al describes prostate cancer therapy with the tumour, key components of immune system, and seven treatment options included. 19 ODEs are employed to simulate the dynamics of 14 variables related to the tumour and the immune system, and 5 treatment-related variables (Coletti et al. 2020). Ma et al developeda a QSP model exploring the efficacy of combination therapy with an anti-PD-L1 monoclonal antibody and bispecific T cell engagers in patients with colorectal cancer. The model contain cancer cells, T cells, immune checkpoints, antibody pharmacokinetics, antigen presentation and T cell engagers. Several compartments are included: the central (blood), peripheral (other tissues and organs), tumour and tumour-draining lymph node compartments. 73 ordinary differential

equations and 105 algebraic equations were used to model the biological processes (Ma et al. 2020). A transcriptome-informed QSP model of metastatic triple-negative breast cancer to identify predictive biomarkers for PD-1 inhibition was build by Arulraj et al. The QSP model is composed of 641 equations and 737 parameters to represent the central, peripheral, primary tumour and primary tumour-draining lymph node, metastatic tumour and associated lymph node compartments (Arulraj et al. 2023).

In summary, QSP modelling offers a novel *in silico* approach to build a more detailed representation of underlying biological processes with the goal to gain a better understanding of the disease mechanism and effects of drug treatments (Lemaire et al. 2023). But often a large system of equations is needed for a detailed description.

1.2 Objectives of this thesis

Cancer, as a complex and heterogeneous disease, encompasses multiple scales of biological organization, ranging from intracellular molecular pathways to interactions between cells and organ and tissues scales. As shown in the previous section, traditional approaches in pharmacometrics and QSP often employ ODE models, which assume a homogeneous cell population and model the time-course of biological measurements continuously.

However, these models overlook crucial aspects of cancer, such as its heterogeneity, spatial characteristics, and the presence of random events like mutations or emergent behavior. Furthermore, modelling the interactions between cancer cells and the TME, including immune cells in space requires large number of parameters and compartments in the ODE approach. To address these limitations, this thesis aims to introduce alternative approaches for modelling the complexity of tumours and their response to drug treatments and combine them with existing PKPD and PBPK frameworks. By exploring alternative modelling strategies, this research seeks to enhance our understanding of tumour behavior and improve the design and optimization of cancer treatments as mono and combination therapy. It acknowledges the need to capture the dynamic nature of cancer and its micro environment, accounting for the inherent complexity and heterogeneity observed in cancer biology.

In chapter two, the aim is to introduce the mathematical and computational concepts employed in our research. To meet this objective, we

- explore the concept of agent based models (ABMs) and their programming, highlighting the underlying Markov process and the Gillespie algorithm,
- introduce ordinary and partial differential equations,
- delve into reinforcement learning, including the Finite Markov process, Bellman equation, the underlying algorithm involving the discount factor, learning rate, discovery rate, epsilon greedy algorithm, and reward formula,
- explore principle component analysis and kmeans clustering algorithm as tools to analyse outcomes of the ABM.

The third chapter has the purpose to introduce and compare an ODE based model and an ABM simulating cancer growth during drug treatment. Hence, in this chapter we

- conduct a case study focused on MEK inhibitors within the Ras-Raf-MEK-ERK pathway,
- implement a PKPD model for the MEK inhibitor cobimetinib,
- examine the treatment of cancer cells using both ODE and ABM approaches,
- investigate single and multiple dose treatments, analyzing individual cell fate, their elimination by the drug, division, and mutation within the population,
- highlight the differences, strengths, and weaknesses of both methods by comparing them.

The objective of chapter four is to present stochastic equations as a bridge connecting ODE models and ABMs. Therefore, we

- simplify the model introduced in chapter three,
- describe how it allows for a stochastic description of a single cell's fate under continuous drug treatment,
- extend this approach to model a population of heterogeneous cells,
- calculate the survival probability of each cell and the whole population to obtain the time to population extinction,
- identify a relation between the population size and time to extinction,

- simulate multiple dose treatments, considering cell recovery and division between cycles,
- identify critical division and death rates for uncontrolled growth and successful treatment.

The fifth chapter aims to apply the learning from the previous chapters by combining the advantages of ABMs, ODEs, and partial differential equations (PDEs) in a hybrid model to study the effects of mono and combination treatments on tumour-immune interactions. Hence, we

- simulate tumour and immune cells using an ABM incorporating rules governing interactions, emergence of behavior, and spatial heterogeneity with simplification of the immune system, considering effector and suppressor cells, PD1-PDL1 interaction, and effector cell exhaustion,
- explore detailed insights into PD1 antibody treatment by using a PBPK model, accounting for drug distribution through different physiological compartments,
- simulate chemotherapy and DNA damage response inhibitors using an ODE based PKPD model,
- incorporate a PDE to simulate diffusion and the concentration level of oxygen and drug in the 3D space,
- include oxygen gradients driven tumour cell behavior, including active division, quiescence, and the presence of a necrotic core,
- model oxygen supply and its impact on radiation treatment, employing a modified linear square model,
- incorporate cell cycle phase specific respond to treatment,
- simulate radiation mediated infiltration of immune cells, PDL1 mutation, and subsequent immune effector exhaustion,
- analyze the effects of different treatments on the interactions between various cell types.

The chapter six has the objective to explore the application of reinforcement learning as a method to automate the optimization of combination treatment schedules. Thus, we

- employ Q learning, where a learning algorithm aims to find the optimal action based on the previous state by collecting rewards and avoiding punishments,
- define the state space by considering the amount of infiltrated immune cells and the presence of cancer cells in specific cell cycle phases,
- allow various treatment options and combinations are considered as possible actions,
- define rewards and punishment as the number of eliminated cancer cells and immune cells,
- set constraints such as toxicity levels and treatment schedules are taken into account during the optimization process,
- train the algorithm during the learning phase, where the environment is randomly explored to discover the most effective strategies.

Chapter 2

Mathematical background

This chapter provides a background of the mathematical and computational methods and statistical techniques which will be used in this thesis. Here I will introduce probability theory and stochastic processes, reinforcement learning, agent based modelling, difference equations, ordinary and partial differential equations as well as principal component analysis and the kmeans clustering algorithm.

2.1 Probability theory

The definitions below are based on the works by Allen *et al.* (L. J. Allen 2010), Kotz *et al.* (Kotz and Nadarajah 2000) and Pinsky *et al.* (Pinsky and Karlin 2010).

The sample space is the set S which contains all possible outcomes of a random experiment. A simple experiment is tossing a coin. In this case the sample space is S =$\{$H,T$\}$ where H and T stand for heads and tails. Next, A is the collection of subsets of S called events, and probability is a function $\mathbb{P} : A \to [0, 1]$ defined on A: for any $a \in A$, $\mathbb{P}(a) \in [0, 1]$.

If B_1 and B_2 are two events which are elements of A, then the *conditional probability* of the event B_1 given event B_2, is denoted as $P(B_1|B_2)$ and defined as

$$\mathbb{P}(B_1|B_2) = \frac{\mathbb{P}(B_1 \cap B_2)}{\mathbb{P}(B_2)}$$

given that $\mathbb{P}(B_2) > 0$. Similarly,

$$\mathbb{P}(B_2|B_1) = \frac{\mathbb{P}(B_1 \cap B_2)}{\mathbb{P}(B_1)},$$

given that $\mathbb{P}(B_1) > 0$.

2.1.1 Random variables

A random variable X is a real-valued function defined on the sample space $S, X : S \to \mathbb{R} = -\infty, \infty$. We define $\mathbb{P}$ as an associated probability on S. Further, let R be the range of X, $R_x = \{x|X(s) = x, s \in S\}$. We write $\mathbb{P}(X \leq x)$ as an abbreviation for $\mathbb{P}(\omega \in S : X(\omega) \leq x)$. The *cumulative distribution function* (cdf) of a random variable X, denoted by $F_X(x)$, is

$$F_X(x) = \mathbb{P}(X \leq x), \quad -\infty < x < +\infty.$$

There are two important types of random variable: discrete and continuous.

A *discrete random variable* has a countable range (commonly a subset of the integers) whereas a *continuous random variable* has an uncountable range (often all positive real numbers). For a discrete random variable X, the *probability mass function* (pmf) describes the probability that X takes a specific value in its range R_x:

$$f_X(x) = \mathbb{P}(X = x),$$

and the expectation $\mathbb{E}[X]$ is a sum:

$$\mathbb{E}[X] = \sum_{x \in R_x} x f_X(x).$$

For a *continuous random variable* Y with the cdf $F_Y(y)$, the probability density function (pdf) is a non-negative, integrable function $f_Y : \mathbb{R} \to [0, \infty)$ such that

$$F_Y(y) = \int_{-\infty}^{x} f_Y(y) \mathrm{d}y;$$

the expectation value is an integral:

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y. \tag{2.1}$$

The *variance* $\mathrm{Var}(X)$ of a random variable X, of either type, is defined as

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}[X]^2).$$

The variance quantifies how much the values deviate from the expected value. The standard deviation of X, $\sigma = \sqrt{\mathrm{Var}(X)}$, is easier to interpret because it has the same units as the random variable itself.

Exponential distribution

The exponential distribution (see Figure 2.1) is used in this thesis due to its memoryless property. This means that the time until the next event is independent from how much time has already passed. An addditional benefits of the exponential distributions is that it allows for simplified calculations and mathematical tractability, related to its memoryless properties. A non-negative continuous random variable Y follows the exponential distribution with parameter $\lambda > 0$ if its probability density function is

$$f_Y(y) = \begin{cases} \lambda \mathrm{e}^{-\lambda y} & y \geq 0, \\ 0 & y < 0. \end{cases} \tag{2.2}$$

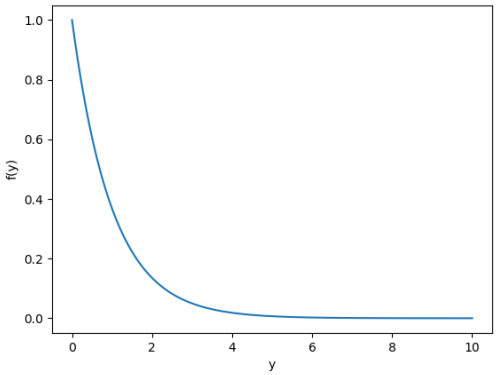


Figure 2.1: The probability density function of the exponential distribution, (see equation 2.2), with $\lambda = 1$.

Uniform distribution

The uniform distribution (see Figure 2.2) is applied when all values within a specific range are equally likely to occur. If a continous random variable follows the *uniform distribution* with parameters $-\infty < a < b < +\infty$, its probability density function is

$$f_Y(y) = \begin{cases} \dfrac{1}{b-a} & y \in [a,b], \\ \\ 0 & \text{otherwise.} \end{cases}$$

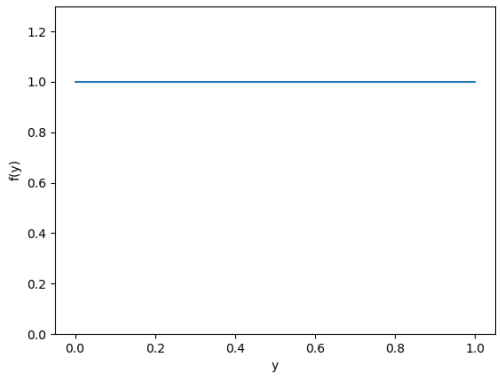


Figure 2.2: The probability density function of the uniform distribution with $a = 0$ and $b = 1$.

Gumbel distribution

The Gumbel distribution (see Figure 2.3) is applied to simulate the occurrence of extreme values (minimum or maximum). The cumulative distribution function is

$$F_Y(y, a, b) = e^{-e^{-(y-a)/b}} \quad y \in \mathbb{R},$$

with $-\infty < a < +\infty$ and $b > 0$. For the standard Gumbel distribution, $a = 0$ and $b = 1$. This gives us the cumulative distribution function

$$F_Y(y, 0, 1) = \mathrm{e}^{-\mathrm{e}^{-y}},$$

and the probability density function

$$f_Y(y) = \mathrm{e}^{-(y+\mathrm{e}^{-y})}.$$

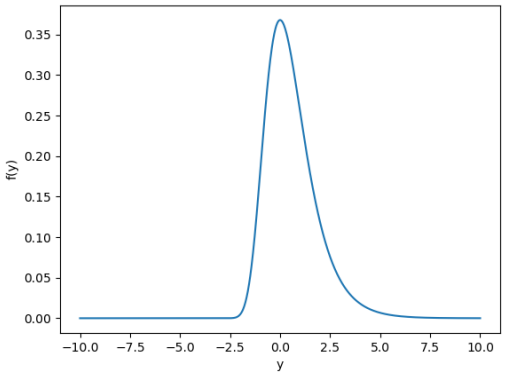


Figure 2.3: The probability density function of the standard Gumbel distribution.

2.2 Stochastic processes

A stochastic process is used to describe the evolution of a system or a phenomenon over time in a probabilistic or random manner. It is a collection of random variables indexed by time, where each random variable represents the state of the system at a specific time point. To explain this concept further we use definitions from Allen (L. J. Allen 2010).

A stochastic process is a collection of random variables $\{X(t) : t \in T\}$. T is an index set and S is the common sample space of the random variables. For each fixed $t, X(t)$ denotes a single random variable defined on S. For each fixed $s \in S, X(t)$ corresponds to a function defined on T which is called a realisation of the process. In this thesis, the index set T, used to denote time, which is typically continuous so that $T = [0, +\infty)$. The state space S_x is a set of possible values for the process. Even though the index set is continuous, many of the stochastic processes described in this thesis have a discrete state space, S_x.

A stochastic process $\{X(t) : t \in [0, \infty)\}$, defined on the state space S_x is a continuous time Markov chain (CTMC) if it satisfies the following condition:

For any sequence of real numbers satisfying $0 \leq t_0 \leq t_1 \leq ... \leq t_n \leq t_{n+1}$,

$$\begin{aligned}\mathbb{P}(X(t_{n+1}) =& x_{n+1} | X(t_0) = x_0, X(t_1) = x_1, ..., X(t_n) = x_n) \\ =& \mathbb{P}(X(t_{n+1}) = x_{n+1} | X(t_n) = x_n),\end{aligned}$$

for any $x_0, x_1, ..., x_{n+1} \in S_x$. This is known as the Markov or 'memoryless' property. The next state x_{n+1} which the Markov process transitions to depends only on the current state x_n at the most recent time t_n of the process and not on previous states or history of the

process.

2.2.1 Transition probabilities

A CTMC $\mathcal{X} = \{X(t) : t \in [0, \infty)\}$ has a probability $p_i(t)$ associated with the random variable $X(t)$ being in a certain state i.

$$p_i(t) = \mathbb{P}(X(t) = i), \quad i \in S_x.$$

The transition probability defines the relation between the random variables $X(t)$ and $X(s)$,

$$p_{ji}(s,t) = \mathbb{P}(X(t) = j | X(s) = i), \quad s < t,$$

for $i, j \in S_x$. If the transition probabilities only depend on the time interval $t - s$ and not explicitly on the times s and t, then the resulting CTMC is said to be time-homogeneous with homogeneous probabilities.

$$p_{ji}(t-s) = \mathbb{P}(X(t) = j | X(s) = i) = \mathbb{P}(X(t-s) = j | X(0) = i),$$

for $s < t$. The transition probabilities may be arranged in a square matrix,

$$\mathbb{P}(t) = (p_{ji}(t))_{j,i \in S_X}.$$

The entries in each row sum up to 1 as in the time interval $[0, t]$ the process must travel from state i to another state $j \in S_x$ or remain in the same state.

The forward Kolmogorov differential equations are a system of equations that describes the rate of change of the transition probabilities $p_{ji}(t)$.

$$\frac{dp_{ji}(t)}{dt} = \sum_{k \in S_x} q_{jk} p_{ki}(t) \quad \forall i, j \in S_x,$$

where q_{jk} are the transition rates.

2.2.2 Interevent times

To calculate sample paths of a continuous Markov chain it is necessary to know the distribution of the time between successive events or the "interevent time". The random variable $T_n = W_{n+1} - W_n \geq 0$ describes the time between changes of state (interevent

time) where W_n is the time at which there is a nth jump with state space S_x. Let $N(t)$ be the population size at time t. We assume that at jump n there is state i such that $N(W_n) = i$ where $i \in S_x$. Then probability to leave state i for the small period of Δt is

$$\sum_{j \neq i} p_{ji}(\Delta t) = \alpha(i)\Delta t + \mathcal{O}(\Delta t),$$

where $\alpha(i)$ is the total propensity calculated from the sum of the rates q_{ji} for all j to leave the state i. The probability to remain in the same state is

$$p_{ii}(\Delta t) = 1 - \alpha(i)\Delta t + \mathcal{O}(\Delta t).$$

The probability $G_n(t)$ that $N(W_n)$ remains in state i for some time t in terms of the interevent time is

$$G_n(t) = \mathbb{P}\left(t + W_n < W_{n+1}\right) = \mathbb{P}\left(t < T_n\right).$$

There is a positive probability of moving to another state if the state i is not an absorbing state. Then at t=0, $G_n(0) = \mathbb{P}\left(T_n > 0\right) = 1$. For Δt sufficiently small,

$$G_n(t + \Delta t) = G_n(t)p_{ii}(\Delta t) = G_n(t)(1 - \alpha(i)\Delta t + \mathcal{O}(\Delta t)). \tag{2.3}$$

Subtraction of $G_n(t)$ and division by Δt and taking the limit as $\Delta t \to 0$ leads to the following ordinary differential equation:

$$\frac{dG_n(t)}{dt} = -\alpha(i)G_n(t)$$

which is a first-order and homogeneous differential equation with initial condition $G_n(0) = 1$. The solution is

$$G_n(t) = \mathbb{P}\left(T_n > t\right) = e^{-\alpha(i)t}. \tag{2.4}$$

This means that

$$\mathbb{P}\left(T_n \leq t\right) = 1 - G_n(t) = 1 - e^{-\alpha(i)t} = F_n(t), \quad t \geq 0.$$

The function $F_n(t)$ is the cumulative distribution for the time T_n that corresponds to an exponential random variable with parameter $\alpha(i)$. The waiting time between one event $N(W_n) = i$ until another event $N(W_{n+1} = i + 1)$ is an exponential random variable with

the parameter $\alpha(i)$ and the mean of

$$\mathbb{E}(T_n) = \frac{1}{\alpha(i)}$$

with the variance of

$$Var(T_n) = \frac{1}{[\alpha(i)]^2}.$$

Equation (2.3) can be rewritten as

$$\mathbb{P}\left(T_n \geq t + \Delta t\right) = \mathbb{P}\left(T_n \geq t\right)\mathbb{P}\left(T_n \geq \Delta t\right),$$

which is equivalent to

$$\begin{aligned}
\mathbb{P}\left(T_n \geq t + \Delta t \mid T_n \geq \Delta t\right) &= \frac{\mathbb{P}\left(T_n \geq t + \Delta t\right)}{\mathbb{P}\left(T_n \geq \Delta t\right)} \\
&= \frac{1 - \left(1 - \exp\left(\alpha(i)\left(t + \Delta t\right)\right)\right)}{1 - \left(1 - \exp\left(\alpha(i)\left(\Delta t\right)\right)\right)} \\
&= \exp\left(-\alpha(i)t\right) \\
&= \mathbb{P}\left(T_n \geq t\right).
\end{aligned}$$

Using the conditional probability and the exponential distribution, the latter shows that the change of state only depends on the current state and not on the previous state. This is known as memory-less property of exponential distributions.

Now let U be a uniform random variable defined on $[0, 1]$ and let T be a continuous random variable on $[0, \infty)$. Then

$$T = F^{-1}(U),$$

where F is the cumulative distribution of the random variable T (L. J. Allen 2010, Theorem 5.5).

For an exponential random variable $T \sim Exp(\alpha)$, the function $F^{-1}(U)$ is found by solving $F(T) = 1 - \exp(-\alpha T) = U$ for U:

$$T = F^{-1}(U) = -\frac{\log(1 - U)}{\alpha}.$$

Since U is a uniform random variable on [0,1], so is also $1 - U$. Now the inter-event time

can be expressed in terms of the uniform random variable U as follows:

$$T = -\frac{\log(U)}{\alpha}.$$

Now given a particular state of the process $N(W_n) = i$ the inter-event time T_n is exponentially distributed with rate $\alpha(i)$. Thus,

$$T_n = -\frac{\log(U)}{\alpha(i)}.$$

Hence the inter-event time can be drawn from a uniform distribution for Gillespie simulations.

2.2.3 Pure death process

The random variable $N(t)$ denotes the population size at time t which decays with cell death in a pure death process. This process is being visualised in a transition diagram in Figure 2.4. To simulate this process with a continuous Markov chain (CTMC), we define $\{N(t) : t \in [0, \infty)]\}$. The initial population size is $N(0) = N_0$, the state space is finite $S_x = \{0, 1, 2...N_0\}$.

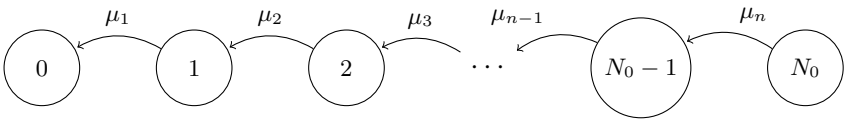


Figure 2.4: Transition diagram of the pure death process simulated as a continuous time Markov chain

Transition probabilities

For a sufficient small Δt, the population size either decreases by 1 or stays at the same size. The transition probabilities satisfy

$$\mathbb{P}\left(N(t+\Delta t) = j | N(t) = i\right) = \begin{cases} \mu_i \Delta t + \mathcal{O}(\Delta t), & \text{if } j = i-1 \\ 1 - \mu_i \Delta t + \mathcal{O}(\Delta t), & \text{if } j = i \\ \mathcal{O}(\Delta t), & \text{otherwise.} \end{cases} \tag{2.5}$$

Equation (2.5) can be used to derive the forward Kolmogorov differential equations for p_{ji}.

$$\begin{aligned} p_{ji}(t+\Delta t) &= p_{j+1,i}(t)[\mu_{j+1}\Delta t + \mathcal{O}(\Delta t)] + p_{ji}(t)[1-\mu_j\Delta t + \mathcal{O}(\Delta t)] \\ &\quad + \sum_{k\neq -1,0,1} p_{j+k,i}(t)\mathcal{O}(\Delta t) \\ &= p_{j+1,i}(t)\mu_{j+1}\Delta t + p_{ji}[1-\mu_j\Delta t] + \mathcal{O}(\Delta t). \end{aligned} \tag{2.6}$$

Subtracting $p_{ji}(t)$ from equation (2.6), dividing by Δt and taking the limit as $\Delta t \to 0$, yields the forward Kolmogorow differential equation for the death process:

$$\frac{dp_{ji}(t)}{dt} = \mu_{j+1}p_{j+1,i}(t).$$

2.2.4 Birth and death process

The random variable $N(t)$ represents the population size at time t which grows with division of cells and decays with cell death. This process is simulated with a continuous time Markov chain (CTMC) $\{N(t) : t \in [0,\infty)]\}$. The state space is infinite $S_x = \{0, 1, 2...\}$.

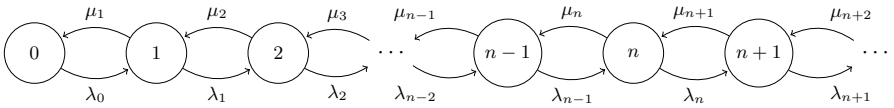


Figure 2.5: Transition diagram of the birth and death process simulated as a continuous time Markov chain

The possible transitions in this stochastic process are depicted in Figure 2.5 where λ_i and μ_i represent the birth and death rate of the population at size $N(t) = i$. If one considers $\lambda_0 = 0$ then 0 is an absorbing state so that the cell population does not re-grow after extinction.

Transition probabilities

The transition probabilities are defined as

$$p_{ji}(\Delta t) = \mathbb{P}\left(N(t+\Delta t) = j | N(t) = i\right) = \begin{cases} \lambda_i \Delta t + \mathcal{O}(\Delta t), & \text{if } j = i+1 \\ \mu_i \Delta t + \mathcal{O}(\Delta t), & \text{if } j = i-1 \\ 1 - (\lambda_i + \mu_i)\Delta t + \mathcal{O}(\Delta t), & \text{if } j = i \\ \mathcal{O}(\Delta t), & \text{otherwise,} \end{cases} \tag{2.7}$$

where $\mathcal{O}(\Delta t)$ is the Landau order symbol.

The forward Kolmogorov differential equations for p_{ji} can be derived directly from the assumption in equation (2.7). For Δt sufficiently small, we consider the transition probability $p_{ji}(t + \Delta t)$:

$$\begin{aligned} p_{ji}(t+\Delta t) &= p_{j-1,i}(t)[\lambda_{j-1}\Delta t + \mathcal{O}(\Delta t)] \\ &\quad + p_{j+1,i}(t)[\mu_{j+1}\Delta t + \mathcal{O}(\Delta t)] \\ &\quad + p_{ji}(t)[1 - (\lambda_j + \mu_j)\Delta t + \mathcal{O}(\Delta t)] \\ &\quad + \sum_{k \neq -1,0,1} p_{j+k,i}(t)\mathcal{O}(\Delta t) \\ &= p_{j-1,i}(t)\lambda_{j-1}\Delta t \\ &\quad + p_{j+1,i}(t)\mu_{j+1}\Delta t \\ &\quad + p_{ji}(t)[1 - (\lambda_j + \mu_j)\Delta t] \\ &\quad + \mathcal{O}(\Delta t), \end{aligned} \tag{2.8}$$

which holds for all i and j in the state space with the exception of $j = 0$. If $j = 0$ then

$$p_{0i}(t+\Delta t) = p_{1i}(t)\mu_1 \Delta t + p_{0i}(t)[1 - \lambda_0 \Delta t] + \mathcal{O}(\Delta t). \tag{2.9}$$

Subtracting $p_{ji}(t), p_{0i}(t)$ from Equations (2.8) and (2.9), dividing by Δt and taking the limit as $\Delta t \to 0$, yields the forward Kolmogorow differential equation for the birth and

death process of the cell population. Assuming that $\lambda_0 = 0$ we obtain:

$$\frac{dp_{ji}(t)}{dt} = \lambda_{j-1}p_{j-1,i}(t) - (\lambda_j + \mu_j)p_{ji}(t) + \mu_{j+1}p_{j+1,i}(t),$$
$$\frac{dp_{0i}(t)}{dt} = \mu_1 p_{1i}(t).$$

2.2.5 Generator matrix

We use the transition probabilities p_{ji} to derive the transition rates q_{ji} which form the infinitesimal generator matrix Q. This matrix Q defines the relationship between the rate of change of the transition probabilities. We assume that the transition probabilities $p_{ji}(t)$ are continuous and differentiable for $t \geq 0$ and satisfy

$$p_{ji}(0) = 0, \quad j \neq i, \quad \text{and} \quad p_{ii}(0) = 1.$$

We define

$$q_{ji} = \lim_{\Delta t \to 0+} \frac{p_{ji}(\Delta t) - p_{ji}(0)}{\Delta t} = \lim_{\Delta t \to 0+} \frac{p_{ji}(\Delta t)}{\Delta t}, \quad i \neq j.$$

We note that $q_{ji} \geq 0$ because $p_{ji}(\Delta t) \geq 0$. Further, we define

$$q_{ii} = \lim_{\Delta t \to 0+} \frac{p_{ii}(\Delta t) - p_{ii}(0)}{\Delta t} = \lim_{\Delta t \to 0+} \frac{p_{ii}(\Delta t) - 1}{\Delta t}.$$

Since in most cases, the transition probabilities have the following properties:

$$p_{ji}(t) \geq 0 \quad \text{and} \quad \sum_{j=0}^{\infty} p_{ji}(t) = 1, \quad t \geq 0,$$

it follows that

$$1 - p_{ii}(\Delta t) = \sum_{j=0, j\neq i}^{\infty} p_{ji}(\Delta t) = \sum_{j=0, j\neq i}^{\infty} [q_{ji}\Delta t + o(\Delta t)].$$

Since the sum of each row of the transition probability matrix is one, we can show that

$$q_{ii} = \lim_{\Delta t \to 0+} \frac{-\sum_{j=0, j\neq i}^{\infty} [q_{ji}\Delta t + o(\Delta t)]}{\Delta t} = -\sum_{j=0, j\neq i}^{\infty} q_{ji}.$$

The transition rates can be represented by the infinitesimal generator matrix $Q = (q_{ji})$

$$Q = \begin{pmatrix} q_{00} & q_{01} & q_{02} & \cdots \\ q_{10} & q_{11} & q_{12} & \cdots \\ q_{20} & q_{21} & q_{22} & \cdots \\ q_{30} & q_{31} & q_{32} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$= \begin{pmatrix} -\sum_{i=1}^{\infty} q_{i0} & q_{01} & q_{02} & \cdots \\ q_{10} & -\sum_{i=0, i\neq 1}^{\infty} q_{i1} & q_{12} & \cdots \\ q_{20} & q_{21} & -\sum_{i=0, i\neq 2}^{\infty} q_{i2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

The matrix Q has the property that each column sum is zero and the ith diagonal element is the negative of the sum of the off-diagonal elements in that column.

The infinitesimal generator matrix Q for the birth and death process has the following structure

$$Q = \begin{pmatrix} 0 & \mu_1 & 0 & 0 & \dots \\ 0 & -\lambda_1 - \mu_1 & \mu_2 & 0 & \dots \\ 0 & \lambda_1 & -\lambda_2 - \mu_2 & \mu_3 & \dots \\ 0 & 0 & \lambda_2 & -\lambda_3 - \mu_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

2.2.6 Gillespie algorithm

The Gillespie algorithm generates numerical realisations of the birth and death process. In this direct method two uniform random variables are needed per iteration. The first one is used to simulate the time to the next event and the second uniform random variable chooses the event. The direct method works well on small population sizes (population size in the order of thousands). Depending on computational power and the rates, an alternative method is required for larger populations because the time between events becomes so small that each sample path takes a long time to compute. For large samples a more efficient but approximate method is the tau-leaping (Fu et al. 2013).

The Gillespie algorithm is executed as follows:

Initialise the system with the initial conditions $t = 0$, maximum time T, and $i = N_0$. Then, while $t < T_{max}$:

1. Sample two random numbers u_1, u_2 from the uniform distribution on $[0, 1)$.

2. Calculate the total propensity $\alpha(i)$:

$$\alpha(i) = \mu_i + \lambda_i.$$

3. Compute the time when the next event takes place: $t + dt$ where dt is given by

$$dt = \frac{-\log(u_1)}{\alpha(i)}.$$

4. Using the second random number u_2 to choose a reaction that occurs at time $t + dt$. Let

$$\mathbb{P}\left(\text{Division}\right) = \frac{\lambda_i}{\lambda_i + \mu_i} \tag{2.10}$$

and $\mathbb{P}\left(\text{Death}\right) = 1 - \mathbb{P}\left(\text{Division}\right)$. Choose the division reaction if $u_2 < \mathbb{P}\left(\text{Divison}\right)$, $i \to i + 1$ otherwise choose death reaction, $i \to i - 1$. In other words, the division range is from 0 to $\mathbb{P}\left(\text{Division}\right)$ and the death range is from $\mathbb{P}\left(\text{Division}\right)$ to 1.

2.2.7 Time to extinction

Here, the time until the process reaches a certain state is studied. We consider a population with size a that reaches size b in a certain time which is referred as the first passage time, $T_{b,a}$. We define $T_{i+1,i}$ to be the random variable for the time that it takes to go from state i to $i+1$. From the derivation for the inter-event time (see Equation (2.4)), we obtain that the pdf of the inter-event time has an exponential distribution with parameter $\lambda_i + \mu_i$:

$$f_i(t) = (\lambda_i + \mu_i)e^{-(\lambda_i+\mu_i)t}.$$

Due to the characteristics of the exponential distribution, the average transition time from state i to $i+1$ or $i-1$, is $1/(\lambda_i+\mu_i)$ which is the mean of the exponential distribution. With the birth probability $\lambda_i/(\lambda_i + \mu_i)$ the process jumps from state i to state $i + 1$ and to state $i - 1$ with the death probability $\mu_i/(\lambda_i + \mu_i)$.

If $a < b$, then the time needed to jump from state a to b is

$$T_{b,a} = T_{a+1,a} + T_{a+2,a+1} + ... + T_{b,b-1}.$$

The expected time to go from state a to b is

$$E(T_{b,a}) = E(T_{a+1,a}) + E(T_{a+2,a+1}) + ... + E(T_{b,b-1}).$$

This is called mean first passage time. Let's consider the time it takes for the process to reach 0, i.e. for a population with size m to become extinct. We need to consider that the process can jump to $i+1$ and $i-1$:

$$\begin{aligned} E(T_{i-1,i}) = & \frac{\mu_i}{\lambda_i + \mu_i}\left(\frac{1}{\lambda_i + \mu_i}\right) \\ & + \frac{\lambda_i}{\lambda_i + \mu_i}\left(\frac{1}{\lambda_i + \mu_i} + E(T_{i,i+1}) + E(T_{i-1,i})\right) \\ & = \frac{1}{\mu_i} + \frac{\lambda_i}{\mu_i}E(T_{i,i+1}). \end{aligned}$$

Hence, time to extinction is calculated as follows:

$$\begin{aligned} E(T_{0,m}) = & \left[\frac{1}{\mu_1} + \frac{\lambda_1}{\mu_1\mu_2} + \frac{\lambda_1\lambda_2}{\mu_1\mu_2\mu_3} + ...\right] + \left[\frac{1}{\mu_2} + \frac{\lambda_2}{\mu_2\mu_3} + \frac{\lambda_2\lambda_3}{\mu_2\mu_3\mu_4} + ...\right] + \\ & \left[\frac{1}{\mu_3} + \frac{\lambda_3}{\mu_3\mu_4} + \frac{\lambda_3\lambda_4}{\mu_3\mu_4\mu_5} + ...\right] + \left[\frac{1}{\mu_m} + \frac{\lambda_m}{\mu_m\mu_{m+1}} + \frac{\lambda_m\lambda_{m+1}}{\mu_m\mu_{m+1}\mu_{m+2}} + ...\right] \end{aligned} \tag{2.11}$$

For the pure death process, this simplifies as follows

$$E(T_{0,m}) = \frac{1}{\mu_1} + \frac{1}{\mu_2} + \frac{1}{\mu_3} + \frac{1}{\mu_4} + ... + \frac{1}{\mu_m}.$$

If $\mu_i = \mu i$, we obtain

$$E(T_{0,m}) = \sum_{r=1}^{m} \frac{1}{\mu}\frac{1}{r}.$$

2.3 Reinforcement learning

This section introduces the reinforcement learning algorithm and its mathematical and computational framework. The definitions given are from Sutton et al (Sutton and Barto 2018).

Reinforcement learning is an algorithm that maps situations to actions with the goal

to maximise a numerical reward signal. The learning algorithm is not being provided instructions what to do; instead it needs to discover which actions yield the largest reward by trial and error. Delayed reward is a characteristic: The aim is not to maximise the immediate reward but to choose actions in a way that all subsequent rewards are considered. Therefore, the algorithm considers the whole problem with the aim to optimise the long term reward.

2.3.1 Finite Markov decision process

Markov decision processes (MDPs) are an extension of Markov chains. Actions and rewards are added to allow choice and encouragement. If only one action exists for a state and all rewards are the same, then a Markov decision process reduces to a Markov chain. Reinforcement learning uses finite MDPs where the transition probabilities or rewards are unknown. This process consists of a set of states, actions and rewards $(\mathcal{S}, \mathcal{A}, \mathcal{R})$ with finite elements. In the finite MDP, the learner interacts with the environment in a sequence of discrete time steps $t = 0, 1, 2, 3....$ At each time step t, the learner perceives the state of the environment $S_t \in \mathcal{S}$ and selects an action $A_t \in \mathcal{A}(S_t)$. As the consequence, the learner receives a reward $R_{t+1} \in \mathcal{R} \subset \mathbb{R}$ in the next state S_{t+1}.

In a finite MDP, the random variables R_t and S_t have discrete probability distributions depending only on the previous state and action. The probability that those random variables take values $s' \in \mathcal{S}$ and $r \in \mathcal{R}$, respectively, at time t, given the preceding state and action, can be defined as the conditional probability:

$$p(s', r|s, a) = P\{S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a\}$$
$$\text{for all} \quad s', s \in \mathcal{S}, r \in \mathcal{R}, a \in \mathcal{A}(s).$$

The function $p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \rightarrow$[0,1) is an ordinary deterministic function of four arguments. p specifies a probability distribution for each choice of s and a:

$$\sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{R}} p(s', r|s, a) = 1 \quad \text{for all} \quad s \in \mathcal{S}, a \in \mathcal{A}(s).$$

From this we can define other functions.

The state-transition probabilities, a three argument function $p : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow$ [0,1] is

defined as

$$p(s'|s,a) = P\{S_t = s'|S_{t-1} = s, A_{t-1} = a\} = \sum_{r\in\mathcal{R}} p(s', r|s, a).$$

The expected rewards for the state-action pairs is a two argument function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and is defined as

$$r(s,a) = \mathbb{E}[R_t|S_{t-1} = s, A_{t-1} = a] = \sum_{s'\in\mathcal{S}} r \sum_{r\in\mathcal{R}} p(s', r|s, a).$$

The time t is omitted from the notation of $r(s,a)$ as after a large number of iterations, the obtained reward for a state-action pair $r(s,a)$ is constant. The expected rewards for the state-action-next-state triples are a three argument function $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$. It is defined as

$$r(s,a,s') = \mathbb{E}[R_t|S_{t-1} = s, A_{t-1} = a, S_t = s'] = \sum_{r\in\mathcal{R}} r\frac{p(s', r|s, a)}{p(s'|s, a)}.$$

2.3.2 Return

The rewards that are received after each time step are denoted as

$$R_{t+1}, R_{t+2}, R_{t+3}, R_{t+4}, \ldots$$

The return may simply be defined as the sum of the rewards G_t. However, to assign the rewards different importance, the discount factor $0 \leq \gamma \leq 1$ is used:

$$\begin{aligned} G_t &= R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \\ &= \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \end{aligned}$$

The discount factor represents the importance of future rewards compared to immediate rewards. If it is close to 0, the agent is myopic and tries to maximise immediate rewards. On the other hand, with a large γ the return objective takes future rewards into account more strongly and may be called farsighted.

2.3.3 Policy

A policy maps the states to probabilities of selecting each action. If the learning algorithm follows the policy π at time t, then $\pi(a|s)$ describes the probability that $A_t = a$ if $S_t = s$ for $a \in \mathcal{A}(s)$ and $s \in \mathcal{S}$.

2.3.4 Value functions

The value of the state s, denoted as $v_\pi(s)$, is the expected return in state s following the policy π. It is defined as:

$$v_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}\middle|S_t = s\right], \quad \text{for all} \quad s \in \mathcal{S}.$$

$\mathbb{E}_\pi[\cdot]$ is the expected value of a random variable if the learning algorithm follows policy π, and t is any time step. This function is the state-value function for policy π.

The value of taking an action a in the state s under policy π, denoted as $q_\pi(s, a)$, is the expected return in state s taking the action a following the policy π. It is defined as:

$$q_\pi(s,a) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}\middle|S_t = s, A_t = a\right],$$
$$\text{for all} \quad s \in \mathcal{S}, a \in A(s).$$

This function is called the action-value function for policy π.

2.3.5 Bellmann equation

From this, the Bellmann equation can be derived which shows the relationship between the value of a state and its successor states.

$$\begin{aligned} v_\pi(s) &= \mathbb{E}_\pi[G_t|S_t = s] \\ &= \mathbb{E}_\pi[R_{t+1} + \gamma G_{t+1}|S_t = s] \\ &= \sum_a \pi(a|s) \sum_{s'}\sum_r p(s', r|s, a)\left[r + \gamma\mathbb{E}_\pi[G_{t+1}|S_{t+1} = s']\right] \\ &= \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma v_\pi(s')] \quad \text{for all} \quad s \in \mathcal{S}. \end{aligned}$$

2.3.6 Temporal difference method

The temporal difference method updates the value of a state $V(s_t)$ by using the immediately observed reward R_{t+1} when the algorithm moves to state s_{t+1},

$$V(s_t) \leftarrow V(s_t) + \alpha[R_{t+1} + \gamma V(s_{t+1}) - V(s_t)],$$

where α is the learning rate.

2.3.7 Dynamic programming

To search for the optimal policy dynamic programming is employed. Through iterative evaluation, the value of a state $V(s)$ is found.

Algorithm 1 Value iteration

Initialise array $V(s) = 0$, for all $s \in S$
Repeat
  For each $s \in S$:
    $v \leftarrow V(s)$
    $V(s) = \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma v_\pi(s')]$
    $\Delta \leftarrow |v - V(s)|$
until $\Delta < \theta$ (a small positive number)
Output $V \approx v_\pi$

Firstly, arbitrary values are being assigned to the value function of a state $V(s)$. Then a loop is started where the previous value of a state $V(s)$ is saved as v. The reinforcement learning algorithm acts in the environment and collects a new reward r which is used to calculate a new value for $V(s)$ according to the Bellmann equation $V(s) = \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma v_\pi(s')]$. Now Δ is assigned to be the difference of the new and old value for a state $V(s)$, see Algorithm 1. This algorithm is repeated until the value for a certain state is stable and the difference of the old and updated value ($\Delta \leftarrow |v - V(s)|$) is smaller than a small threshold θ.

2.3.8 Epsilon greedy algorithm

The balance between exploration and exploitation is important in reinforcement learning. While exploring the environment the learning algorithm collects rewards for certain actions. To maximise the collected rewards, the learning algorithm must prefer actions that were effective producing rewards in the past. But to gain further rewards and learn about more favourable actions, the environment has to be explored at the cost of obtaining a low reward. Neither exploration or exploitation alone will lead to a successful learning. The algorithm must try various actions and progressively favor those which lead to a more favourable state and long term reward. In addition on a stochastic task, each action must be executed many times to obtain a reliable estimate of its expected reward.

To balance the choice of either exploitation or exploration, the epsilon greedy algorithm can be employed (Sutton and Barto 2018). The epsilon value states the likelihood of taking a random action or exploit the learned information about the environment. Depending on the environment and the gained knowledge the value of epsilon can be

adjusted.

2.4 Agent based modelling

Agent based models consist of one or a set of different agents that have attributes and act autonomously in an environment according to rules that they follow during the simulation. To apply this concept to PKPD modelling and simulation of cancer growth, individual cells or a group of cells can be represented as agents (Metzcar et al. 2019). They interact in an environment which can consist of different conditions i.e. different spatial location, concentration of drugs, nutrient, receptors, or other agents. Interactions with the environment and neighbouring agents are based on internal decision-making rules. An agent undergoes growth, quiescence, apoptosis or necrosis as a response to surrounding environmental conditions or interaction with other agents (Zhihui Wang et al. 2015). Additionally, agents can have memory, where they store information about past interactions or environmental conditions. Overall, this modelling technique is a bottom-up method without central coordination and simple activities at the microscopic scale lead to complex behavior on the macroscopic scale (Seel 2011; Bonabeau 2002; Cosgrove et al. 2015).

Agents

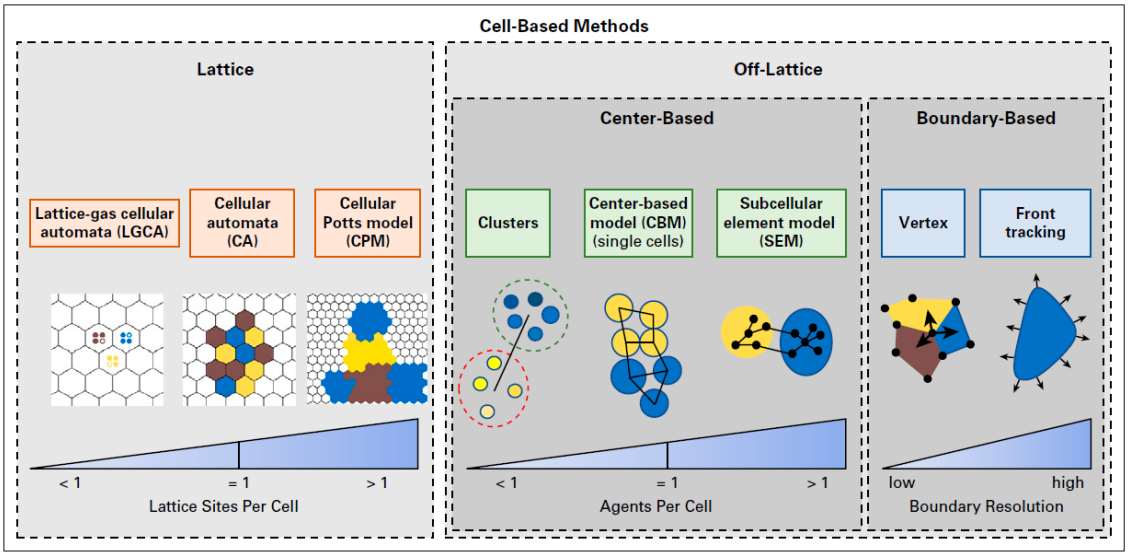


Figure 2.6: Overview of cell-based modelling methods (Figure taken from Metzcar et al. (Metzcar et al. 2019) under the Creative Commons Attribution License (CC BY) permission)

To model the cells as agents, there are two main paradigms: lattice-based models that contain cells along a rigid grid, and off-lattice models without this restriction (Metzcar

et al. 2019). To create a lattice-based model, regular structured grids (e.g. Cartesian [2D or 3D], dodecahedral [3D]) or unstructured grids are used (Rejniak and A. R. Anderson 2011; Metzcar et al. 2019). Lattice-based methods can be further classified by their spatial resolution (Metzcar et al. 2019). Cellular automaton models contain a single cell in one lattice site. Discrete lattice-based rules are used to update a single cell at every time step. To reduce grid artifacts, lattice sites are updated in random order. A cell can remain, move to a unoccupied adjacent lattice site (randomly, or following a directional stimulus such as chemo- or haptotaxis), die and vacate a lattice site, or divide and allocate a progenitor cell in a neighbouring site. The number of neighbour cells depends on the shape of a grid. For example, in a square lattice, cells have either four or eight neighbours (von Neumann or Moore neighbourhood) whereas cells can be surrounded by six symmetrically located cells in a hexagonal grid (Rejniak and A. R. Anderson 2011; Metzcar et al. 2019). Lattice gas cellular automaton models contain multiple cells in one lattice. Instead of tracking every single cell, this model traces a group of cells that moves through channels from individual lattice sites (Metzcar et al. 2019). Cellular Potts models utilise multiple lattice sites to simulate one cell. An individual cell can occupy multiple pixels (2D) or voxels (3D). This is an advantage when the resolution of individual cell morphologies is required. Changes in cell shape and direct cell-cell interaction are directed by the concept of Monte Carlo simulations and energy minimalization. A disadvantage of this method is the computational intensity and the challenge of calibration from Monte Carlo steps to physical time (Rejniak and A. R. Anderson 2011; Metzcar et al. 2019). Off-lattice methods may be considered more realistic because cells can take arbitrary relative positions and freedom to move in any direction instead of being ordered on a grid. However, this comes with the disadvantage of higher computational cost because special algorithms are necessary to efficiently handle cell-cell neighbourhoods. During movement cell collisions need to be avoided which can be challenging in densely packed areas or populations. After division, the placement of daughter cells need to be determined to avoid overlapping with other cells or mutually exclusive cell areas. In addition, chemical values of the environment are usually computed on regular grids and interpolation techniques need to be applied to transfer values between the cellular off-lattice individuals and the chemical fields (Rejniak and A. R. Anderson 2011). Off-lattice models can be categorised into 'centre based' models that focus on cell volumes (or masses) and boundary-based methods. Further classification is by the level of morphological detail (Metzcar et al. 2019). 'Centre based'

methods can be divided into three approaches: 'centre based' models, sub-cellular element model, and clusters. 'centre based' models simulate each cell's centre of mass or volume by using one agent per cell. Cell positions are updated after exchanges of adhesive, repulsive, locomotive, and drag-like forces between cell centres (Metzcar et al. 2019). Cells can be represented with different shapes. Spheres simulate cells as single points with a predefined minimum distance. Deformable ellipsoids have two axes with different lengths forming an elliptical shape. This allows for a definition of cell orientation and polarization (Rejniak and A. R. Anderson 2011). Another 'centre based' method is the sub-cellular element model. In this model, cell morphology is displayed in greater detail by having multiple agents for sub-cellular elements of each cell. Agents interact with adhesive and repulsive forces. A benefit of this model is a better approximation of cell bio-mechanics but this comes with higher computational cost. A different option is clusters. An agent simulates a cluster of cells or functional units (e.g. breast glands or colon crypts). Agents interact by mechanical forces or alternative rule-based motion. This technique can incorporate heterogeneous details into cell clusters with less computational cost. Boundary-tracking methods include vertex-based models or immersed boundary methods (a front tracking method). They are computationally intensive however, they are useful for describing detailed cell mechanics to fluid and solid tissue mechanics. In vertex-based models cells are polygons (2D) or polyhedra (3D). Forces on the vertices are computed which is useful in modelling confluent tissues. Front-tracking methods such as immersed boundary method can be used for greater spatial resolution (Metzcar et al. 2019).

Environment

To model the environment and the interaction of agents in the environment agent based models are embedded in multi-scale models. Multi-scale modelling aims to include various spatio-temporal scales from atomic to molecular, cellular, multi-cellular, organ, and up to multi-organ systems. On the molecular scale, interactions (i.e. receptor-ligand interactions, consumption and production of oxygen, nutrient, and cell-cell signalling molecule concentration) are described with ordinary differential equations. Levels of those molecular entities are part of the environment and can cause changes in cell behaviour. For example signalling molecules such as growth factors can initiate cascades that lead to proliferation. Interaction between cells occurs on the microscopic scale. Cells are simulated with agent based modelling. Biological rules imposed by the modeller determine cell interaction with

the environment or other cells. Agents can interact with each other, influence the micro environment and enter quiescence or undergo apoptosis or necrosis. Local conditions and environmental changes are described using partial differential equations. For example, the availability of resources such as oxygen, nutrient, and hormonal distribution are modeled through diffusion from molecularly rich regions (i.e. blood vessels, tumour edge). Macroscopic scale modelling covers the entire tumour tissue and comprises many agents. This scale includes the general pattern of growth, total cell number, the extent of metastasis, tumour morphology and vascularization. Tumour tissue pH, nutrient availability, diffusion of oxygen and signalling molecules are modeled across the entire tumour by using partial differential equations.

A hybrid multi-scale approach can be used for simulation of cancer development. Cells on the micro-scale are modeled as agents while effects that can be approximated using a continuum description are studied with PDEs or ODEs and solved for the entire tumour instead for each agent individually. Part of the model is discretised to study the effect of mutation-induced, more aggressive phenotypes on metastasis or cell and extracellular matrix interaction. Discrete modelling is needed to simulate the multi-step process of changes at the cellular level that leads to tumour growth and metastasis. The hybridization of a discrete model in continuum environment provides a more complete description of the tumour morphology, higher accuracy of model predictions, lower computational costs (Zhihui Wang et al. 2015).

Programming of agent based models

Agent based models can be implemented in object oriented programming languages such as Python, Java, C, or C++. Differences between those languages are user friendliness, speed and available libraries. Python is beginner friendly and has a rich environment of standard libraries whereas Java, C and C++ are computationally faster (Macal and North 2010; J. A. Baggio and R. Baggio 2020). An advantage of coding the model is that the researcher has control over all aspects of the model, additional features can be easily implemented or changed. Disadvantages are time-consuming programming and the need for programming skills (Heppenstall et al. 2011; Gilbert 2019). Since many models use the same or similar building blocks with small variations, libraries for different programming languages are available which provides conceptual frameworks and templates that allow the user to design a customised model. Examples are MASON for Java, SWARM for Java,

Objected C, Repast for Java, C++, and MESA for Python. A benefit is the reduced model construction time. Disadvantages are the time needed to learn the programming language, and the complexity of the toolkit, which means that it can take some time before one can take full advantage of the wide range of features. In addition, the desired function might not be available (Heppenstall et al. 2011; Gilbert 2019). More suitable for beginners are software packages like Netlogo, which provides complete systems in which models can be created and executed, and the results visualised. Such packages are easier to learn and useful for rapid development of basic or prototype models. However, the researcher has less flexibility, is restricted to the design framework supported by the corresponding software, and extension or integration of additional tools might not be possible. In addition, the speed of execution can be slower (Heppenstall et al. 2011; Gilbert 2019).

2.5 Difference and differential equations

Biological processes are often expressed with difference or differential equations, as in those processes changes happen depending on time, space or stage of development. Difference equations are applied to model discrete time steps while differential equations describe a change over continuous time. If only temporal changes are modelled, then ordinary difference or differential equations are used. If in addition space or age is of interest, then partial difference or differential equations are needed. The definitions in this section are from Allen (L. Allen 2007).

2.5.1 Ordinary difference equations

A difference equation of order k has the form

$$f(x_{t+k}, x_{t+k-1}, ...x_{t+1}, x_t, t) = 0, \quad t = 0, 1, ...,$$

where f is a real-valued function of the real variables $x_t, x_{t+1}, ..., x_{t+k-1}, x_{t+k}$ and t. f must depend on x_t and x_{t+k}, so that the order is k. An autonomous difference equation does not depend explicitly on t while a non-autonomous one does.

A common form of a difference equation is

$$x_{t+k} + a_1 x_{t+k-1} + ... + a_{k-1} x_{t+1} + a_k x_t = b_t, \quad t = 0, 1, ... \quad .$$

If $a_k \neq 0$, then the order of the difference equation is k. The coefficients are assumed to be real and the function to be real valued. The coefficients $a_j, j = 1, ..., k$, can be functions of t and x_i for $i = t, ..., t + k - 1$. On the right hand side of the equation, the function b_t can depend on t but not on the state variables.

The difference equation is said to be linear if the coefficients $a_j, j = 1, ...k$ are constant or depend on t but not on the state variables, otherwise it is said to be non-linear. The difference equation is homogeneous if the equation is linear and $b_t = 0$ for all t, otherwise it is non-homogeneous.

A system of k first-order equations can be expressed as

$$x_i(t+1) = f_i(x_1(t), x_2(t), ...x_k(t), t), \quad i = 1, 2, ..., k$$

and in a matrix notation as

$$X(t+1) = A(t)X(t) + B(t),$$

where vector $X(t) = (x_1(t), x_2(t), ...x_k(t))^T$, A is a square matrix with dimension k $A(t) = (a_{ij}(t))_{i,j-1}^k$ and vector $B(t) = (b_1(t), ..., b_k(t))^T$

A solution to a difference equation is a function x_t, $t = 0, 1, 2, ..$ such that when substituted into the equation makes it a true statement. For a system of difference equations the solution is a set of functions $x_i(t), i = 1, 2, .., k$, that when substituted into the systems of equations makes each of them a true statement.

2.5.2 Ordinary differential equations

A differential equation of order n has the form

$$f(x, dx/dt, d^2x/dt^2, ..., d^nx/dt^n, t) = 0.$$

An autonomous differential equation does not depend explicitly on t, otherwise it is non-autonomous. Typically, a differential equation of order n has the form

$$\frac{d^nx}{dt^n} + a_1(t)\frac{d^{n-1}x}{dt^{n-1}} + ... + a_{n-1}(t)\frac{dx}{dt} + a_n(t)x = g(t).$$

A differential equation is linear if the coefficients $a_i, i = 1, ..., n$, and g are either constant or functions of t but not functions of x or any of its derivatives. Otherwise, the differential equation is said to be non-linear. The linear differential equation is homogeneous if $g(t) = 0$ and otherwise non-homogeneous.

A first-order system of differential equations can be expressed as

$$\frac{dX}{dt} = F(X(t), t),$$

where the vector $X(t) = (x_1(t), x_2(t), ...x_n(t))^T$, $F = (f_1, f_2, ..., f_n)^T$ and $f_i = f_i(x_1(t), x_2(t), ...x_n(t), t)$. The system of differential equations is autonomous if the right side does not depend explicitly on t, otherwise it is non-autonomous.

A first-order system is linear if it can be expressed as

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{ij}(t)x_j + g_i(t),$$

$i = 1, ..., n$. Otherwise it is said to be non-linear. The system is homogeneous if it is linear and $g(t) = 0$, otherwise it is non-homogeneous.

A solution of a differential equation or system of differential equation is a scalar function $x(t)$ or a vector function $X(t)$ which makes the differential equation or system an identity if substituted into it.

2.5.3 Partial differential equations

When change of entities over time and space needs to be considered, partial differential equations are applied. Here, as an example, we derive the diffusion equation from first principles. The diffusion equation is written as follows

$$\frac{\partial U}{\partial t} = D\frac{\partial^2 U}{\partial x^2} \tag{2.12}$$

where $t \in [0, \infty)$ and $x \in \mathbb{R}$. To obtain the derivative of Equation (2.12), the spatial domain $\mathbb{R}$ is divided into intervals of length Δx and the time domain into intervals of length Δt. $U(t, x)$ is a concentration at time t and position x. Let λ_r be the probability to move to the right a distance of Δx and λ_l be the probability to move to the left a distance of Δx during the time period Δt, $0 \leq \lambda_l + \lambda_r \leq 1$. As a result, the concentration in a

time period Δt at time t and position x is

$$U(t,x) = (1-\lambda_r-\lambda_l)U(t-\Delta t,x) + \lambda_l U(t-\Delta t, x+\Delta x) + \lambda_r U(t-\Delta t, x-\Delta x).$$

The first term on the right side describes the proportion of the concentration that remains in position x, the second term refers to the proportion that moves to the left (from $x+\Delta x$ to x) and the third term is the proportion of the concentration that moves to the right (from $x-\Delta x$ to x). If it is equally likely to move to both directions, then $\lambda_r = 1/2 = \lambda_l$ and

$$U(t,x) = \frac{1}{2}U(t-\Delta t, x+\Delta x) + \frac{1}{2}U(t-\Delta t, x-\Delta x).$$

By subtracting $U(t-\Delta t,x)$ from both sides and dividing by Δt, we obtain

$$\frac{U(t,x)-U(t-\Delta t,x)}{\Delta t} = \frac{(\Delta x)^2}{2\Delta t}\left[\frac{U(t-\Delta t,x+\Delta x)-2U(t-\Delta t,x)+U(t-\Delta t,x-\Delta x)}{(\Delta x)^2}\right]$$

The term on the left side is a difference equation approximation to $\partial U/\partial t$, and the expression on the right side in the square brackets is the difference equation approximation to $\partial^2 U/\partial x^2$. Hence, if $\Delta t \to 0, \Delta x \to 0$, and $(\Delta x)^2/(2\Delta t) \to D = constant$, then the diffusion Equation (2.12) is derived. D is the diffusion constant with the dimensions $(distance)^2/time$.

The diffusion equation with three spatial dimensions, $U(t,x,y,z)$, is as follows:

$$\frac{\partial^2 U}{\partial t^2} = D\left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}\right) = D\Delta U,$$

where $\Delta = \nabla^2$ is the Laplace operator. Moreover, this can be rewritten as

$$\partial_t U = \nabla^2 U = \left(\partial_{xx} + \partial_{yy} + \partial_{zz}\right) U.$$

This can be discretised using a forward Euler in time and a second-order central finite difference, such that

$$\frac{U^{n+1}_{i,j,k} - U^n_{i,j,k}}{\Delta t} = \frac{U^n_{i-1,j,k} - 2U^n_{i,j,k} + U^n_{i+1,j,k}}{(\Delta x)^2} + \frac{U^n_{i,j-1,k} - 2U^n_{i,j,k} + U^n_{i,j+1,k}}{(\Delta y)^2} + \frac{U^n_{i,j,k-1} - 2U^n_{i,j,k} + U^n_{i,j,k+1}}{(\Delta z)^2},$$

which implies that

$$U_{i,j,k}^{n+1} = U_{i,j,k}^{n} + \Delta t\Big(\frac{U_{i-1,j,k}^{n} - 2U_{i,j,k}^{n} + U_{i+1,j,k}^{n}}{(\Delta x)^2} + \frac{U_{i,j-1,k}^{n} - 2U_{i,j,k}^{n} + U_{i,j+1,k}^{n}}{(\Delta y)^2} + \frac{U_{i,j,k-1}^{n} - 2U_{i,j,k}^{n} + U_{i,j,k+1}^{n}}{(\Delta z)^2}\Big).$$

2.6 Data analysis

In this section, the principal component analysis (PCA) technique and the k-means clustering algorithm are explained based on Chatfield et al (Chatfield 2018), Sinaga et al (Sinaga and M.-S. Yang 2020), and Syakur et al (Syakur et al. 2018). We are using those techniques to analyse the data produced by the agent based model in order to understand the behaviour of the different kinds of agents. Those data sets are multi-dimensional and by using PCA, the first two principal components can be identified. In comparison to the initial data set those principal components can describe the variation in the data set with fewer dimensions making the data set more interpretable and easier to visualise. After applying PCA, we plot the data in two dimensions and use the k-means clustering algorithm to identify clusters of data points that are closely related, to further classify and understand the data set.

2.6.1 Principal component analysis

This technique is applied to summarise, visualise and reduce the dimensions of a of a multi-dimensional, highly correlated data set. Those are data sets that consist of a large number of dimensions or features for each observation. The goal is to group a p-dimensional data set into p principal components (PCs). This is accomplished by linear transformation of the data into a new coordinate system. By applying this method, it can be found that only a few PCs hold more variability than any single original variable. As a result, those PCs become the focus to further analyse the data.

The linear transformation of the data set is as follows: Let $X = [X_1, .., X_p]$ be a p-dimensional data set with mean μ and covariance matrix α. To transform the data set into uncorrelated p principal components $Y_1, ..., Y_p$ with decreasing variance, each Y_j needs to be a linear combination of X's.

$$\begin{aligned} Y_j &= a_{1j}X_1 + a_{2j}X_2 + ... + a_{pj}X_p \\ &= a_j^T X, \end{aligned}$$

with $a_j^T = [a_{1j}, ...a_{pj}]$ being a vector of constants. For the first PC, the constants a_1^T in $Y_1 = a_1^T X$ must be found so that $Var(a_1^T X)$ is maximal. To ensure that the variance doesn't become unbound, the constrain $a_j^T a_j = 1$ is used. By applying the method of Lagrange multipliers, we obtain that a_1 is the eigenvector of the sample covariance matrix α, corresponding to the largest eigenvalue. Similarly for the other PCs, for the kth PC $Y_k = a_k^T X$, a_k^T is the eigenvector of the sample covariance matrix corresponding to the kth largest eigenvalue.

2.6.2 k-means clustering algorithm

k-means clustering is an unsupervised learning algorithm where data points are assigned to a cluster based on their distance to the centre of that cluster.

The algorithm starts with randomly determining cluster centres to the data set. Using the Euclidean distance each data point is then assigned to the nearest cluster centre. This process is repeated in order to optimise the k-means objective function, until the stopping criterion is reached, for example ending at the maximum number of iterations.

In detail, the algorithm is as follows: Let $X = \{x_1, .., x_n\}$ be a data set in a d-dimensional Euclidian space $\mathbf{R}^d$. We define $A = \{a_1, .., a_c\}$ as the c cluster centres and $z = [z_{ij}]_{n \times c}$ where z_{ij} is a binary variable (i.e. $z_{ij} \in \{0, 1\}$) indicating whether the data point x_i is in the jth cluster, $j = 1, ..., c$. The k-means objective function is

$$J(z, A) = \sum_{i=1}^{n} \sum_{j=1}^{c} z_{ij} ||x_i - a_j||^2.$$

The k-means algorithm is repetitively applied to minimise the k-means objective function $J(z, A)$ and updates the cluster centres and membership function a_j and z_{ij}, as follows:

$$a_j = \frac{\sum_{i=1}^{n} z_{ij} x_i}{\sum_{i=1}^{n} z_{ij}}$$

$$z_{ij} = \begin{cases} 1, & \text{if} \quad ||x_i - a_j||^2 = min_{1 \leq j \leq c} ||x_i - a_j||^2 \\ 0, & \text{otherwise.} \end{cases}$$

where $||x_i - a_j||$ is the Euclidian distance between the data point x_i and the cluster centre a_j.

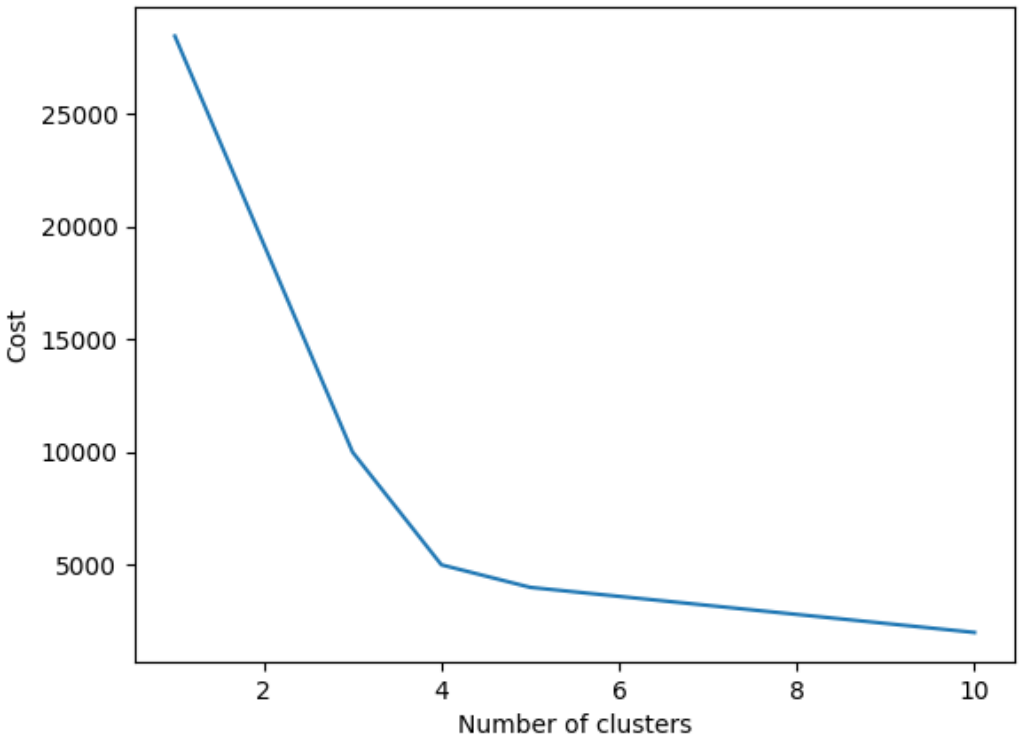


Figure 2.7: Using equation 2.13 the cost for each cluster number is calculated. The cost is plotted against the number of cluster which creates an elbow shape. The elbow point is used to estimate the optimal number of clusters.

Elbow method

Different approaches exist to determine the number of clusters needed for a given data set. The Elbow method calculates the cost function for each value of K as the WCSS (Within-Cluster Sum of Squares) which is the sum of the squared distance between each point and the centroid in a cluster:

$$WCSS = \sum_{k=1}^{c} \sum_{x_i \in S_k} ||x_i - a_k||^2, \tag{2.13}$$

where x_i is a point and a_k is the centre of cluster S_k. The plot of K versus the cost function (Equation 2.13) looks like an elbow (see Figure 2.7). With an increased number of clusters, the cost function will decrease as the points are closer to their centroid. At $K = 1$ the cost function is the largest. Figure 2.7 shows that the graph rapidly changes at a point and therefore creates an elbow shape. After that point, the graph is almost parallel to the horizontal axis. The corresponding K value at that point is the optimal number of clusters.

Chapter 3

Step-by-step comparison of ordinary differential equation and agent based approaches to pharmacokinetic-pharmacodynamic models

3.1 Introduction

Molecular targeted therapy is a valuable approach in the treatment of tumours, offering benefits such as selective binding to both extracellular and intracellular targets due to its small size, superior performance over cytotoxic chemotherapy, and minimal side effects. A majority of these therapies focus on inhibiting kinases responsible for regulating cell proliferation, metabolism, and immune modulation (Y. Song et al. 2023). The cellular pathway RAS–RAF–MEK–ERK is a chain of proteins that transmits extracellular signals to specific intracellular targets. ERK regulates a wide array of substrates related to cell proliferation, differentiation, and metabolism. RAS mutations are present in approximately 33% of human cancers, with B-RAF mutations observed in 8%. The abnormal activation of this signaling pathway contributes significantly to tumour development. Consequently, various inhibitors have been developed to target this pathway, significantly altering cancer treatment strategies (Y. Song et al. 2023).

The MAPK signaling pathway (shown in Figure 3.1) is initiated by binding of growth factors, cytokines, and extracellular mitogens which lead to dimerization, activation, and transphosphorylation of receptor tyrosine kinases (RTKs), such as epidermal growth factor receptor (EGFR), HER kinase, mesenchymal to epithelial transition factor (MET), and fibroblast-growth factor receptor. Then the activated growth factor receptors interact with a series of adaptor proteins, such as growth factor receptor-bound protein 2 (GRB2). As a consequence, guanine nucleotide exchange factors (GEFs) are recruited to the plasma membrane. GEFs activate membrane-bound RAS small guanosine triphosphate (GTP)ases (H-RAS, N-RAS, and K-RAS). These small GTPases are intrinsically stagnant and have the role of a guanosine diphosphate (GDP)/GTP-regulated switcher by catalyzation of the conversion of inactive GDP-bound RAS to active GTP-bound RAS. Next, the active GTP-bound RAS recruits RAF serine/threonine kinases to the plasma membrane, triggering a complex series of events involving phosphorylation, dimerization, and protein-protein interactions that activate the RAF serine/threonine kinases. The RAF serine/threonine kinases are RAF family members (A-RAF, B-RAF, and C-RAF) and MAP kinase kinase kinases (MAPKKK). They use RAS proteins as common upstream activators and activate MEK1 and MEK2, which are kinase effectors MAP kinase kinases (MAPKKs), through phosphorylation. As the next step, the tyrosine and serine/threonine dual-specificity kinases MEK1 and MEK2 catalyze activation via phosphorylation effector MAP kinases ERK1 and ERK2 as their substrates.

Even though they are considered to be functionally equivalent, MEK1/2 are regulated differentially and non-interchangeably during a variety of cellular events, including epidermal hyperplasia and tumourigenesis. Whereas RAF and MEK are very substrate-specific, the ERK1/2 phosphorylate a multitude of nuclear and cytoplasmic targets (more than 600) including transcription factors, kinases, phosphatases, and cytoskeletal proteins. This leads to diverse cellular responses such as cell proliferation, survival, differentiation, motility, metabolism, programmed cell death, embryogenesis, and angiogenesis (Y. Li et al. 2019).

In physiological conditions, MAPK signaling is evolutionarily conserved and tightly controlled by feedback loops at multiple levels. However, the RAS-RAF-MEK-ERK signaling cascade is one of the most potent signaling pathways that dominate carcinogenesis. Although mutations in MEK and ERK-encoding genes are rare gain-of-function mutations in RAS-encoding genes and alteration in gens of the RAF family members are frequently

observed in human cancer. As a consequence of the overactivation, an autocrine/paracrine loop is established that supplies proliferative signals and stimulates cell growth. Additionally, the expression of cell cycle regulators is changed. This leads to premature cell cycle arrest and inhibits progression. Pro-apoptotic proteins are repressed, and anti apoptotic proteins are activated. Up-regulation of telomerase stimulates senescence evasion. Epithelial-to-mesenchymal transition (EMT) is increased. Cell invasiveness and motility are elevated. Interactions between cancerous and stromal cells are impaired. This affects angiogenesis and detection of cancer cells by the immune system (Y. Li et al. 2019).

These findings motivate the development of small-molecule inhibitors targeting the kinases of the MAPK pathway. The narrow substrate specificity, low mutation rate, and the unique structural characteristics of MEK1/2 make it an ideal target for therapeutic development. Contrary to other kinase inhibitors, MEK inhibitors provide high specificity by binding to a hydrophobic pocket adjacent to but not overlapping with the adenosine triphosphate (ATP)-binding site. This binding freezes the conformation of MEK and hinders interaction with ERK activation loops. Therefore, MEK inhibition is directly proportional to the degree of activation of the MAPK pathway in RAS, RAF or EGFR-mutant tumours (Y. Li et al. 2019).

One example for a MEK inhibitor is cobimetinib (also known as GDC-0973). It is a potent, selective MEK1 inhibitor with a half maximal inhibitory concentration (IC50) estimate of 4.2 nmol/L in an in vitro biochemical assay against purified MEK1 enzyme. This value is a measurement of potency and describes how much drug is needed to inhibit a biological process by half (Aykul and Martinez-Hackert 2016). Cobimetinib is more than 100-fold selective for MEK1 over MEK2 and showed no significant inhibition when tested against a panel of more than 100 of serine-threonine and tyrosine kinases (Wong et al. 2012). By inhibiting the MEK enzyme (see Figure 3.1), it decreases phosphorylated ERK (pERK) and as a consequence decrease cell proliferation and cell survival. Drug resistance significantly undermines the effectiveness of various treatments for human cancers, including targeted therapy. The approved MEK inhibitors offer advantages such as high selectivity, initial effectiveness, and low side effects. However, clinical responses to these inhibitors vary greatly among patients, and resistance tends to develop within a year without additional intervention (Y. Song et al. 2023). The mechanisms behind this resistance include the reactivation of the MAPK pathway, leading to subsequent reactivation of ERK. This reactivation is triggered by alterations or mutations in molecules

upstream of ERK within the MAPK pathway, such as RAS, RAF, NF1, or MEK. MEK mutations can result in either overactivation of MEK or render inhibitors ineffective in binding to MEK. Another significant consequence of MEK inhibitors is the reactivation of multiple receptor tyrosine kinases upstream of the MAPK pathway, which initiates a signaling cascade that drives cellular growth and proliferation, contributing to adaptive resistance. Receptor tyrosine kinases have diverse pathways through which they promote cell signaling, and the reactivation of receptor tyrosine kinases following MEK inhibition stimulates cellular growth via these multiple pathways. In addition to the MAPK pathway, there are parallel pathways that promote cell proliferation and growth, such as the PI3K, STAT, and Hippo signaling pathways. When the MAPK pathway is blocked, cancer cells may resort to these alternative pathways to acquire the necessary signals for growth, leading to resistance against MEK inhibitors. Another possible mechanism of resistance involves tumour cells altering their phenotypes and rewiring metabolic pathways, leading to a different feedback mechanism and subsequent MAPK activation after MEK inhibitor treatment (Kun et al. 2021).

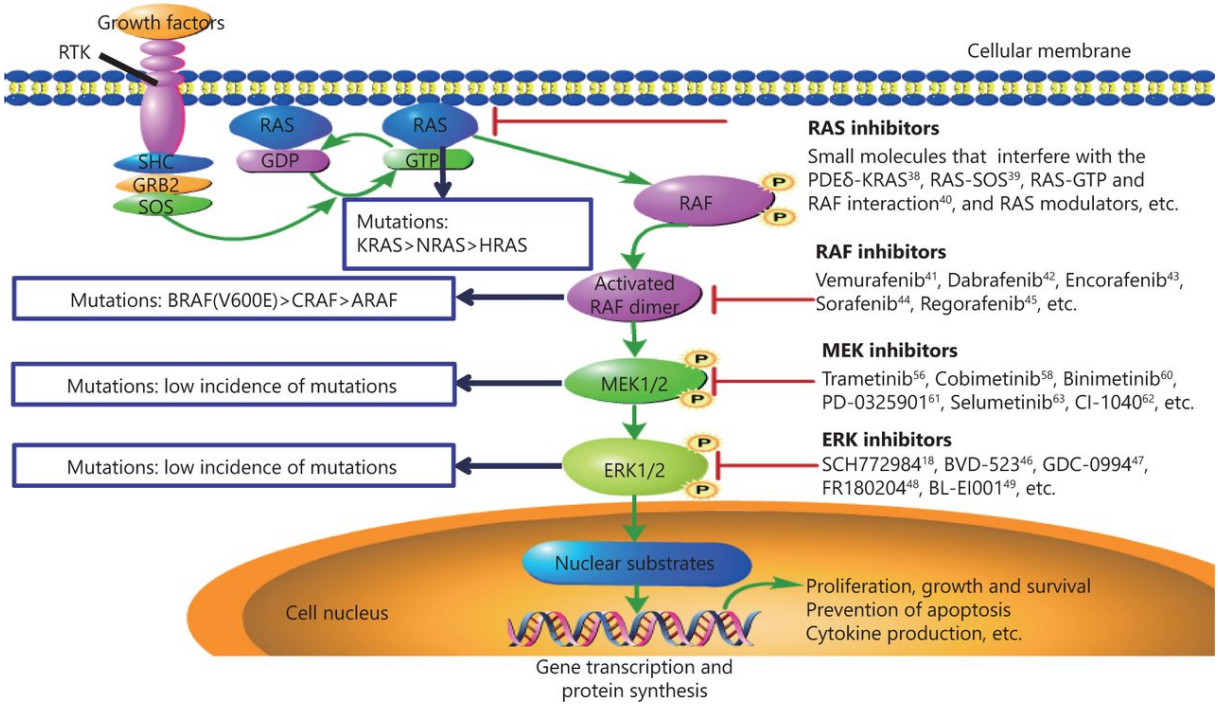


Figure 3.1: The simplified schematic of the intercellular RAS-RAF-MEK-ERK signaling pathway shows the involved enzymes, occurance of mutations and drug target sites. Reprinted with permission from Cancer Biology & Medicine. Copyright (2023) by Cancer Biology & Medicine

(Y. Li et al. 2019)

To improve therapy it is crucial to understand the behaviour of the cell population and consequences of treatment. Therefore, in this chapter we introduce models to simulate

a cancer cell population undergoing MEK inhibitor treatment. Another objective of this chapter is to introduce agent based modelling (ABM), a widely employed methodology in both social sciences and systems biology (Bianchi and Squazzoni 2015; Gary An et al. 2009) and compare it to the widely used models based on ODEs. The chapter provides a simple example, utilizing a PKPD-ABM for studying the effects of the MEK inhibitor cobimetinib in anti cancer treatment. In this ABM, tumour cells are simulated as individual agents during the treatment with cobimetinib, with the drug's time course and impact serving as the environmental factors. In order to maintain simplicity, the model does not incorporate cell location. Cell death and division take place at the microscopic scale, and a summary of cell behavior is used to determine the total number of tumour cells at the macroscopic level. The model's implementation is carried out using the Python programming language, and the code is available at https://github.com/VanThuyTruong/Tutorial . The cobimetinib concentration time-course and ERK phosphorylation are simulated using a PKPD model previously published by Wong et al (Wong et al. 2012). This model serves as an illustrative example and is implemented in accordance with the original reference, for the sole purpose of numerical simulation. To accommodate pharmacokinetics, a hybrid approach is adopted, integrating PK through an ODE system, coupled with a tumour ABM to analyze cell interactions under varying drug concentrations. A comparison is made between the behavior of the hybrid PKPD-ABM tumour model and a PKPD-ODE tumour model. This comparative analysis aims to highlight the distinct advantages and limitations inherent in each method.

3.2 Example: PKPD-ABM and PKPD-ODE model for simulating anti cancer treatment with cobimetinib

3.2.1 PK model

The PK of the drug was characterised by an ODE model that links dose taken orally to a plasma compartment and a tumour micro environment compartment (see part A in Figure 3.2). The PK in the tumour compartment induces a reduction of phosphorylation of the ERK pathway that is used as the driving force to implement the PKPD model, either as a full ODE model (part C1 in Figure 3.2) or as an hybrid ODE-ABM (part C2 in Figure 3.2).

We use the model by Wong et al. (Wong et al. 2012) to describe kinetics of drug

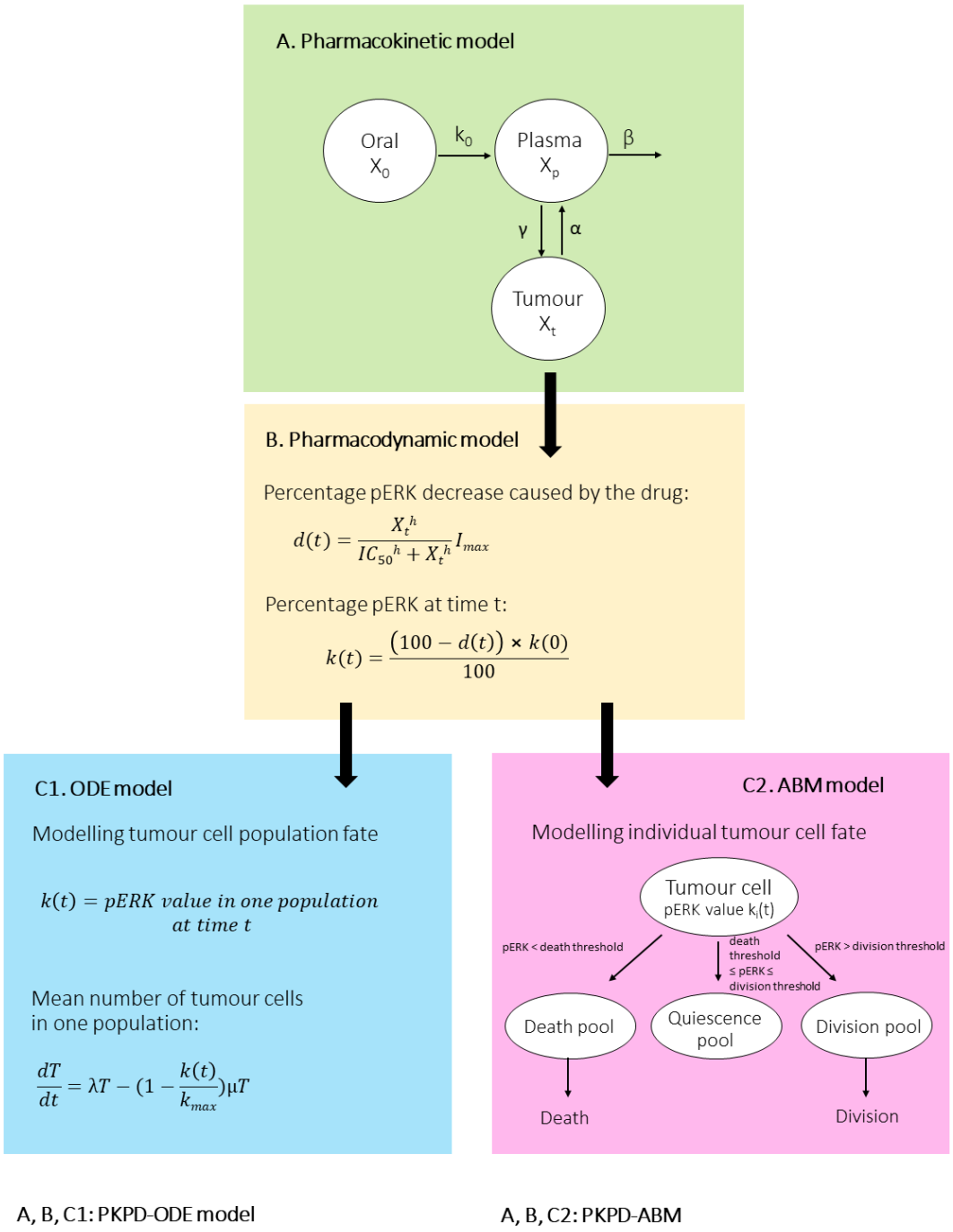


Figure 3.2: Model structure. A) The PK model describes tumour disposition of cobimetinib. B) The PKPD model links the percentage pERK decrease, $d(t)$, to the amount of cobimetinib inside the tumour compartment. Phosphorylated ERK plays an important role in cell division. Therefore, pERK could be seen as a biomarker of tumour growth and a decrease of pERK causes a decrease in cell division. C1) The ODE model simulates the effect of the pERK decrease on the number of tumour cells in the population. $k(t)$ is the pERK value inside one population. In the simulation cases in this chapter, we model different populations with a different pERK value. The mean number of cells increases with the birth rate λ and decreases with death rate μ. A high pERK value favors birth, a low value favors decay over growth. C2) An ABM, where tumour cell death or division is driven by the percentage pERK decrease caused by the amount of cobimetinib inside the tumour compartment, $d(t)$, and the individual pERK value of each tumour cell, $k_i(t)$. A,B,C1) together result in a PKPD-ODE and A,B,C2) together result in a PKPD-ABM.

concentration in oral (X_0), plasma (X_p) and tumour (X_t) compartments after a weight-based oral dose of cobimetinib. The ODEs are linear, with absorption rate k_0 from the oral to the plasma compartment, elimination rate β from the plasma and inter-compartmental rate γ from the plasma to the tumour compartment. Transfer from the tumour to the plasmas is included, with the inter-compartmental rate α. Because the drug is a small molecule, it is assumed that drug excess is high and loss due to target binding, catabolism, or elimination in the tumour compartment is negligible. Therefore, no elimination rate is introduced in the tumour compartment. The time courses follow the following ODEs:

$$\frac{\mathrm{d}X_0}{\mathrm{d}t} = -k_0 X_0, \tag{3.1}$$

$$\frac{\mathrm{d}X_p}{\mathrm{d}t} = k_0 X_0 + \alpha X_t - \gamma X_p - \beta X_p, \tag{3.2}$$

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \gamma X_p - \alpha X_t. \tag{3.3}$$

The initial condition $X_0(0)$ is the given dose (1 or 3 mg/kg of body weight) converted to μmoles/kg where the molecular weight of cobimetinib is 531.3 g/mol. Both the initial drug amount in the plasma compartment $X_p(0)$ and the drug amount in the tumour compartment $X_t(0)$ are equal to zero. Our parameter values are based on those of Wong et al. (Wong et al. 2012).

3.2.2 PD model

The PD model (N. H. Holford and Sheiner 1982) describes the effect of the anti cancer drug by the following relationship between $X_t(t)$ and the percentage pERK decrease, $d(t)$:

$$d(t) = \frac{X_t(t)^h}{\mathrm{IC}_{50}^h + X_t(t)^h} I_{\max}, \tag{3.4}$$

where IA_{50} is the cobimetinib amount in the tumour compartment at which the percentage pERK decrease is half of I_{max}, the maximum percentage pERK decrease, and h is the Hill coefficient (Wong et al. 2012).

Depending on the initial pERK value, $k(0)$, and the percentage pERK decrease caused by the drug $d(t)$ at time t, the pERK value is:

$$k(t) = \frac{100 - d(t)}{100} k(0). \tag{3.5}$$

In Figure 3.2, we illustrate the two types of tumour model, based on Equation (3.5), that we will consider: agent based (C2) and population based (C1). In the agent based model, we assign a different initial value $k_i(0)$ to each tumour cell, drawn uniformly in the range $(0, k_{\max})$. Values of $k_i(0)$ greater than 100 represent cells that have activating mutations in B-RAF resulting in higher activation of MEK. Individual values of $k_i(t)$ evolve according to Equation (3.5). In the population based model, a single value $k(t)$ represents the average pERK status of one population of tumour cells. In the simulation cases below we model different populations of tumour cells with the ODE model, which have different average pERK values (different starting $k(0)$). That is, cell-to-cell variability is introduced in the agent based model; the total number of cells is an integer that depends on time because individual cells may die or divide. The population-based model, on the other hand, is a set of ODEs where one value of $k(t)$ represents the influence of the drug on the population of tumour cells. The population size is also governed by an ODE, which we now introduce.

3.2.3 PKPD-ODE

The population-based model is the PKPD-ODE model shown as C1 in Figure 3.2. A single pERK value $k(t) \in (0, k_{\max})$ characterises the tumour cell population and is updated according to Equation (3.5). The size of the tumour cell population at time t, $T(t)$, obeys:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \lambda T - \left(1 - \frac{k(t)}{k_{\max}}\right) \mu T. \tag{3.6}$$

The tumour population grows with the exponential term λT and shrinks depending on the $k(t)$ and $kmax$ value. The constant $\lambda > 0$ represents the division rate. Death is modelled as a decreasing function of $k(t)$. In this chapter, we make the choice $\mu = 2\lambda$, which means that division and death balance when $k(t) = \frac{1}{2}k_{\max}$. There is more than one possible pair of μ and λ. We use the value $\lambda = 0.0828$ day^{-1} of Wong *et al.* (Wong et al. 2012) for comparability. As we will see, this choice also facilitates comparison with the ABM.

3.2.4 PKPD-ABM

In the ABM, each tumour cell is impacted by the pERK decrease individually according to Equation 3.5. One agent is a tumour cell i with a value $k_i(t)$ as an attribute, representing its degree of intra-cellular phosphorylation of ERK, determined by the amount of drug in

the tumour compartment $X_t(t)$ according to Equation (3.3) from the PK model. As the drug concentration lowers, the cell's pERK value recovers back to its initial value $k_i(0)$, unless the cell dies according to the decision-making algorithm.

Decision-making algorithm

The fate of a cell is governed by a set of rules depending on its ERK phosphorylation level (see Figure 3.2 part C2). pERK thresholds determine the behaviour of the cells. A cell whose pERK value is above the division threshold is said to be in the division pool (in this chapter the division threshold is set as $\frac{1}{2}$, see Equation 3.7); a pERK value below the death threshold consigns the cell to the death pool (in this chapter the death threshold is set as $\frac{1}{4}$, see Equation 3.8). Inside the division or death pool, a division or death event happens with a certain probability per unit time. In between these thresholds, a cell remains quiescent.

The ABM is a Markov process that can be summarised as follows. At time t, the number of tumour cells is $N(t)$, each cell with its pERK value. The probability that cell i divides in the interval $(t, t + \Delta t)$ is $\lambda_i(t)\Delta t$ where

$$\lambda_i(t) = \lambda \max\left\{k_i(t) - \frac{1}{2}k_{\max},\, 0\right\}. \tag{3.7}$$

As seen in Equation 3.7, a cell has a certain division rate depending on its current pERK value ($k_i(t)$) when its current pERK value ($k_i(t)$) is above the division threshold of $\frac{1}{2}$. Otherwise, the division rate is zero.

Similarly, the probability that cell i dies in the interval $(t, t + \Delta t)$ is $\mu_i(t)\Delta t$ where

$$\mu_i(t) = \mu \max\left\{\frac{1}{4}k_{\max} - k_i(t),\, 0\right\}. \tag{3.8}$$

As Equation 3.8 shows, a cell as a certain death rate depending on its current pERK value ($k_i(t)$) when its current pERK value ($k_i(t)$) is below the death threshold of $\frac{1}{4}$. Otherwise, the death rate is zero.

At the cellular level there are birth and death events. The probability that event j is the first event is

$$\mathbb{P}\left(j \text{ first }\right) = \frac{\text{rate of event } j}{\text{sum of all rates}}.$$

The probability that the event is a division event is

$$\mathbb{P}\left(\text{division}\right) = \frac{\sum_{a=1}^{N(t)} \lambda_a(t)}{\sum_{a=1}^{N(t)} \mu_a(t) + \sum_{a=1}^{N(t)} \lambda_a(t)},$$

where $N(t)$ is the number of cells at timepoint t. And the probability for the death event is

$$\mathbb{P}\left(\text{death}\right) = \frac{\sum_{a=1}^{N(t)} \mu_a(t)}{\sum_{a=1}^{N(t)} \mu_a(t) + \sum_{a=1}^{N(t)} \lambda_a(t)}.$$

If the event is a birth event, then the probability that, of all the $N(t)$ possible cells, it is cell a that divides, is

$$\mathbb{P}\left(\text{cell } a \text{ divides} \mid \text{event is a division}\right) = \frac{\lambda_a(t)}{\sum_{a=1}^{N(t)} \lambda_a(t)}. \tag{3.9}$$

Similarly, if the event is a death event, then the probability that, of all the $N(t)$ possible cells, it is cell a that dies, is

$$\mathbb{P}\left(\text{cell } a \text{ dies} \mid \text{event is a death}\right) = \frac{\mu_a(t)}{\sum_{a=1}^{N(t)} \mu_a(t)}. \tag{3.10}$$

A cell is more likely to be chosen to proliferate and produce one daughter cell when the pERK value is high. The initial pERK value of the parent cell will be inherited to the offspring. A low pERK value will make a tumour cell more likely to die.

Markov process

The macroscopic scale of the agent based model shows the tumour population with the overall cell number. The random variable $N(t)$ represents the tumour population size at time t, which grows with division of tumour cells and decays with cell death. This process is simulated with a continuous time Markov chain (CTMC) $\{N(t) : t \in [0, \infty)\}$ with infinite state space and therefore the tumour growth is unlimited. To limit the tumour growth

a finite state space with and absorbing state with the maximum number of tumour cells can be applied. Here, we use a infinite state space as for our modelling purposes, the cancer population will shrink due to the drug treatment hence no limits are required.f The Markov process can be used because the change of state in our simulation only depends on the current state and not the history of the tumour cells.

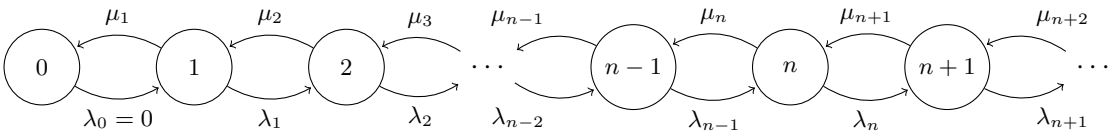


Figure 3.3: The transition diagram of the Markov process shows the population scale with the number of cells and birth and death rates

The possible transitions in this stochastic process are depicted in Figure 3.3 where λ_t and μ_t represent the birth and death rate of the population at size $N(t) = k$. The rates are defined as:

$$\lambda_t = \sum_{a=1}^{k} \lambda_a(t), \tag{3.11}$$

$$\mu_t = \sum_{a=1}^{k} \mu_a(t). \tag{3.12}$$

State 0 with $\lambda_0 = 0$ is an absorbing state because the tumour cell population does not re-grow after extinction.

Gillespie algorithm

We use the Gillespie algorithm, which is a well-established method for designating which of these $2N(t)$ possible events is the first, after t, to occur. We construct the following sums:

$$\Lambda(t) = \sum_{a=1}^{N(t)} \lambda_a(t) \qquad \text{and} \qquad M(t) = \sum_{a=1}^{N(t)} \mu_a(t). \tag{3.13}$$

We decide which type of event occurs, then decide which cell it happens to. At each update, the current values of $\mu_i(t)$ and $\lambda_i(t)$ are used to construct the sums $\Lambda(t)$ and $M(t)$. With probability $\Lambda(t)/(M(t)+\Lambda(t))$, the first event after time t is a division event. In that case, the probability that, of the $N(t)$ possible cells, it is cell i that divides, is $\lambda_i(t)/\Lambda(t)$. With probability $M(t)/(M(t)+\Lambda(t))$, the first event after time t is a death event. In that case, the probability that cell i dies is $\mu_i(t)/M(t)$. The initial pERK value of the parent cell at the time of division is inherited by the daughter cells. The time increment corresponding to each update is also a random variable, exponentially-distributed with mean equal to

$(M(t)+\Lambda(t))^{-1}$. To complete one step of the algorithm, each cell's pERK value is updated based on the current drug amount $X_t(t)$.

3.2.5 Example of the decision making algorithm

A simple example, starting with two cells, is shown in Figure 3.4. A treatment with a single dose of 3mg/kg of cobimetinib is given. In this simulation case, the death threshold is set at 50% and the division threshold at 100 %. More parameters are displayed in Table 3.1. One cell has an initial pERK value equal to 11%. That is, it is in the death pool. The blue cross at bottom left indicates the death of this cell. The second cell, with an initial pERK value equal to 195% of the pERK division threshold, is in the division pool. The early-time effect of the drug is to decrease its pERK value. Division events, the first of which occurs shortly after 50 hours, are indicated by crosses. A division event is the creation of a new cell, whose initial pERK value is the pERK value of the mother cell at the time of division. Figure 3.4 shows that the original cell, displayed with a turquoise line, gave birth to a daughter cell, displayed with an orange line. The drug affects the daughter cell, hence its pERK value decreases from its inherited value of 125% right after division. At times greater than 250 hours, the drug is largely eliminated from the tumour tissue and each tumour cell regains its initial pERK value.

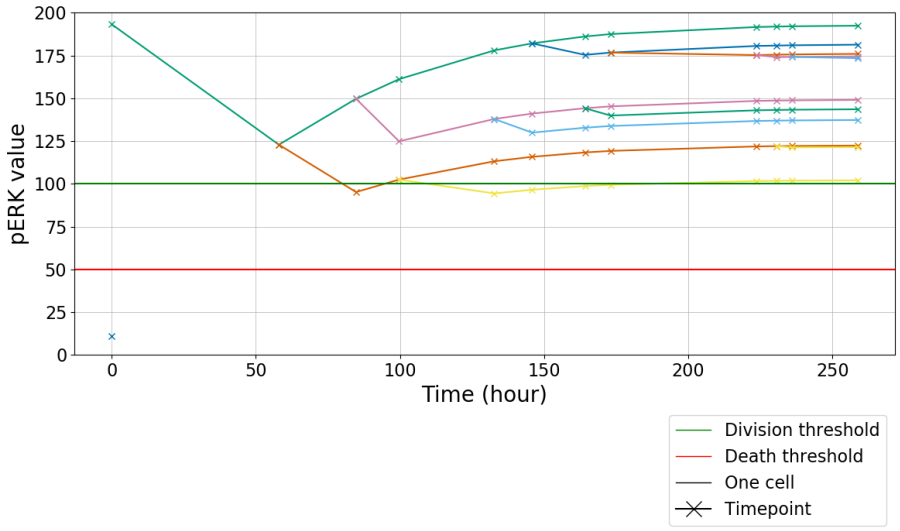


Figure 3.4: pERK values and behaviour of a population starting with two cells with a pERK value of approximately 11% and 195% pERK

Parameter	Value
Pharmacokinetic model	
Initial dose	1 mg/kg or 3 mg/kg
k_0	1.08 h^{-1} (Wong et al. 2012)
β	3.13 h^{-1} (Wong et al. 2012)
γ	1.74 h^{-1} (Wong et al. 2012)
α	0.04 h^{-1} (Wong et al. 2012)
Pharmacodynamic model	
IA_{50}	0.78 $\mu mol/kg$ (Wong et al. 2012)
$I_{max}(\%)$	97 (Wong et al. 2012)
h	1 (Wong et al. 2012)
ODE	
Initial pERK value	varied, depending on simulation scenario
Initial cell number	100 Cells [assumed]
k_{max}	200% [assumed]
λ	0.0828 $days^{-1}$ (Wong et al. 2012)
μ	0.1656 $days^{-1}$ [assumed]
ABM	
Initial pERK value	varied, depending on simulation scenario
Initial cell number	100 Cells [assumed]
Death threshold	varied, depending on simulation scenario
Division threshold	varied, depending on simulation scenario
λ	0.0828 $days^{-1}$ (Wong et al. 2012)
μ	0.0828 $days^{-1}$ [assumed]

Table 3.1: Parameters for the simulations

3.3 Simulation outcomes

Different simulation scenarios were explored with both models. The drug dose and the distribution of initial pERK were varied for each simulation. The ODE model shows the behaviour at the population level with the overall cell number (see for example Figure 3.5d where each blue line represents one tumour population) and the homogeneous pERK value of each population (see for example Figure 3.5b where each purple line describes the pERK value inside one population). The ABM gives further insight into each cell of the population with the pERK value of each cell over time (see for example Figure 3.5c where each colour represents one cell) and the number of cells in the division, quiescence, and death pool (see for example Figure 3.6b). It is possible to track the pERK value of each cell inside the population over time as in Figure 3.5c, where each coloured line represents one cell which can die and cause the line to end or give birth to an offspring which creates a new line starting from the time-point of birth (marked with a cross). The crosses describe the discrete time-points for events caused by the Gillespie algorithm. The scatter plots show each cell inside the population with its pERK value at specific time-points (see Figures 3.6c-3.6f) and videos at https://github.com/VanThuyTruong/Tutorial/tree/main/videos show the behaviour of the cells in the scatter-plots with cell birth and death over the simulation period.

Single dose treatment with 3 mg/kg

First, a treatment with a single dose was simulated with 3 mg/kg of drug per body weight (see Figure 3.5 and 3.6). The initial pERK values for the ODE model and ABM were sampled from a uniform distribution in the range $0 - 200$. 100 is assumed to be the average pERK value, while cells with a pERK value higher than 100 are assumed to have acquired a mutation which allows them to divide faster by having a higher pERK value. The ODE model simulates 100 tumour cell populations with different initial pERK (see Figure 3.5b) while the ABM follows the fate of 100 tumour cells for a given patient (see Figure 3.5c). In both models exposure-driven pERK reduction induces tumour shrinkage followed by regrowth as the drug level falls (see Figures 3.5b-3.5e). A difference is that the ABM is able to track cell death and division of each cell (Figure 3.5c).

Simulating 100 times with the ABM (Figure 3.5e) yields 100 different population histories because the initial pERK distribution in each trial is different, and due to the

stochasticity of the Gillespie algorithm. We observe that the total number of cells in the population may be higher in comparison to the ODE model, because the cells with a higher pERK values are more likely to survive and divide (see Figures 3.5d and 3.5e).

Figures 3.5c and 3.6a-3.6f provide a closer look into one population. We see that the pERK value of a majority of cells falls under the death threshold quickly after treatment start at 0 hours (see Figure 3.6d), those cells go into the death pool (see Figure 3.6b) and therefore can die, which decreases the number of cells (see Figure 3.6a). When the drug level falls, the cells recover and the pERK value increases (see Figure 3.6e). After approximately 70 h, the drug concentration is low enough for the cells to gain a pERK value above the division threshold and enter the division pool, those cells cause a population growth (see Figure 3.6a, 3.6b, 3.6f). This simulation shows that cells with a low pERK level which are assumed not being mutated are most affected by the drug while drug treatment leads to an evolution of a population consisting of mutated cells with a high pERK value.

Bimodal distribution of pERK values

The next example aims to recapitulate a tumour where cells have a distribution of pERK levels based on two given phenotypes. We construct bimodal distributions from two normal distributions where 50 cells in the ABM or 50 populations in the ODE model have a pERK value with a mean of 60% while 50 cells in the ABM or 50 populations in the ODE model have a pERK value with a mean of 190%. The standard deviations are 10% in both distributions. A single dose of 3 mg/kg was chosen, hence the PKPD model of the drug treatment of the examples in Figures 3.5 till 3.8 are identical. The ODE model shows distinct behaviour for populations with different pERK values (see Figure 3.7c). While the ABM does not display the bi-modality on the population level (Figure 3.7d). In the ABM the heterogeneity is shown on the microscopic level (Figure 3.7b) and the cell number is summed up for the macroscopic level (population level). Therefore, the bimodality is only seen at the microscopic level. Taking a closer look into the example, we see that for the ABM, the number of tumour cells initially falls (Figure 3.8a). This is caused by the cells of the population with the mean pERK of 60% which are either in the death or quiescence pool (see Figures 3.8b, 3.8c). They die out or remain in the quiescence pool and therefore are not able to contribute to the population growth (see Figures 3.7b, 3.8c-3.8f). The cells with high pERK value around 190 % are resilient to the drug treatment, because

the decrease of active pERK caused by the drug is not enough to bring those cells under the death threshold. Then, the drug amount in the tumour compartment decreases and these tumour cells leave the quiescence pool and enter the division pool where they can divide again and cause an increase of the total cell number (see Figures 3.7b, 3.8c-3.8f). The tumour cells start dividing and pass their pERK value at the division time-point to the daughter cells. Looking at Figure 3.7b, we see that the offsprings have initially lower pERK values than the parent cell, because the pERK decrease caused by the drug immediately affects them. During the time course of the simulation, their pERK value increases until they regain their inherited pERK values (Figures 3.7b, 3.8f). Because of the heritage of the pERK value, a typical pattern evolves and in the time course of the simulation there will be a tumour population with a high metabolic phosphorylation pool (compare Figures 3.8c and 3.8f). The overall population growth is driven by the cells with a high pERK value. Therefore, comparing the overall cell number of 100 simulations with a uniform pERK distribution (Figure 3.5e) and the bimodal distribution (Figure 3.7d) we see similar behaviour with a constant population growth after a slight decrease until approximately 50h. This simulation emphasises that tumour growth is driven by selection of cells with a high pERK value. Due to the heterogeneity of pERK values, the tumour cell population survives.

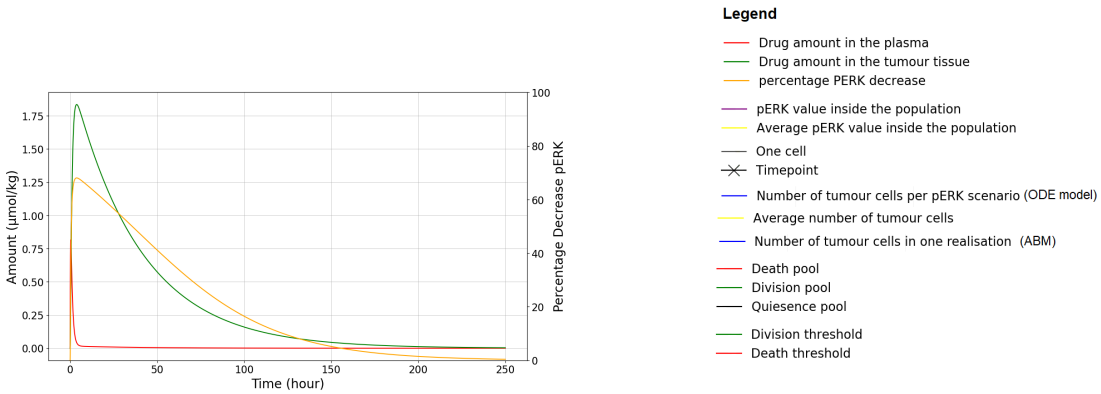
Higher division threshold

The effect of changing division and death thresholds was explored in Figure 3.9. The parameters are the same as in Figure 3.5 except for the division threshold of 150% pERK, instead of 100% pERK. This was compared to an ODE model (Figures 3.5b, 3.5d) and an ABM with the division threshold of 100% pERK (see Figures 3.5c, 3.5e). A treatment of a single dose of 3mg/kg was given.

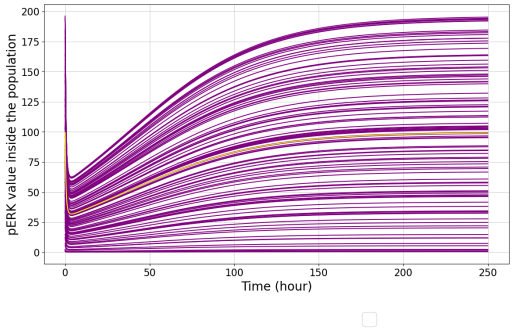
The higher division and death thresholds lead to a longer residence of the tumour cells in the quiescence pool (see Figures 3.9d-3.9h). As the Figures show, the drug suppressed quickly the majority of cells into the death pool (Figure 3.9f) where cells die. Although the drug amount in the tumour compartment decreases at 93 hours, the majority of cells are not able to reach the division pool (Figure 3.9g). Division starts at a later time than in the example with a lower division threshold (compare Figures 3.9h and 3.6f). This leads to a lower growth of the tumour population (compare Figures 3.9b, 3.5e).

Multiple treatment cycles

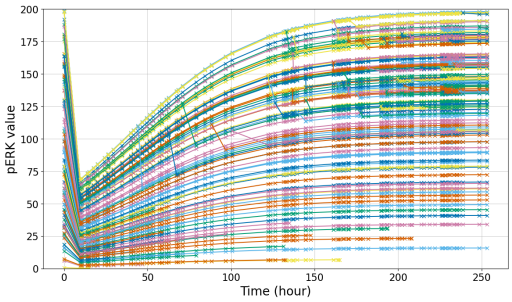
A treatment of 1mg/kg administered every 24 hours with 10 cycles was simulated in Figure 3.10. The pERK value in the PKPD-ODE model and the PKPD-ABM varies according to the treatment cycle (Figures 3.10b, 3.10c). The pERK-time course in the ABM has less regular onset-offset due to the time steps of the Gillespie algorithm (Figure 3.10c). The time points are sparse because the time steps in the Gillespie algorithm depend on the number of cells in the death and division pools which are lower because most of the cells are trapped in the quiescence pool (see Figure 3.11b). Overall, the ODE and ABM show similar behaviour. Due to the constant presence of drug, Figure 3.11b shows almost no division of cells, while cells oscillate between division and quiescence pools (Figures 3.11c-3.11f). This results in a decrease of cell number. There is a peak every 24 hours in the division pool, followed by a peak in the death pool (see time points every 24 hours in Figure 3.11b). While the treatment cycles continue, the drug accumulates, which causes the pERK decrease to be high enough to keep the cells outside of the division pool, so that the total cell number slowly decreases (Figures 3.11b, 3.11a). After approximately 200 hours, the cells in the death pool go extinct while some cells with initially high pERK values remain dormant in the quiescence pool (see Figure 3.11b). This resembles drug resistance of mutated cells with a high pERK values.



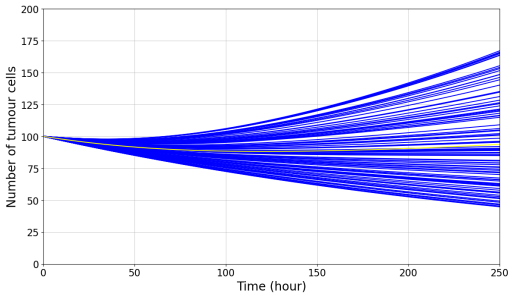
(a) PKPD model: Drug amounts in the plasma and tumour compartment (PK, left-hand scale) and percentage pERK decrease (PD, right-hand scale)



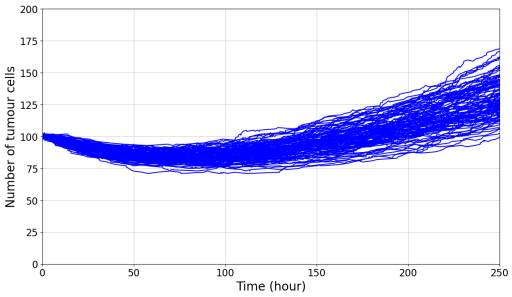
(b) ODE: Individual pERK value of every cell population



(c) ABM: Each line is the pERK value of one cell in one realisation

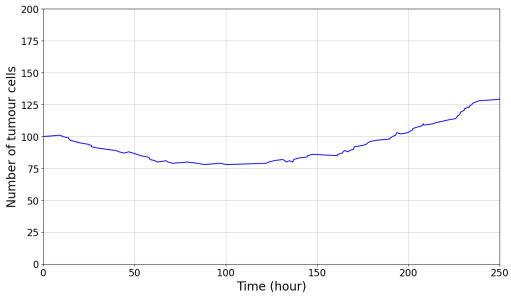


(d) ODE: Overall cell number of 100 tumour cell populations

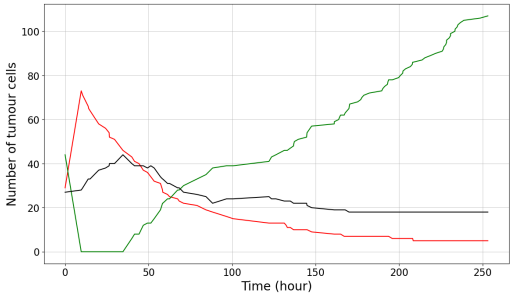


(e) ABM model: Overall tumour cell numbers in 100 realisations

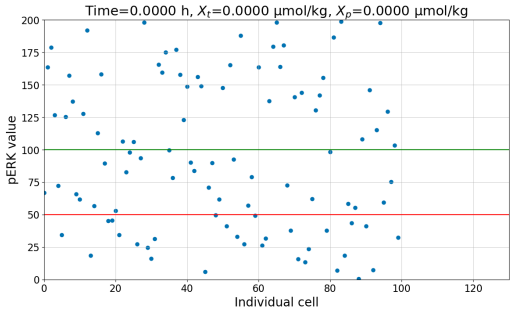
Figure 3.5: Simulation of the ODE model and ABM with a single oral dose of 3 mg/kg. pERK values are initially uniformly distributed. The parameter values are given in Table 3.1. Figure 3.5a shows the PKPD model (model A,B in Figure 3.2), common to the ODE model and ABM. Figures 3.5b and 3.5d show multiple trajectories of the PKPD-ODE model (model C1 in Figure 3.2), each with the same initial cell population size but a different initial pERK value. Figure 3.5c shows individual pERK values in one realisation of the ABM (model C2 in Figure 3.2). Here, each cell has a different initial pERK value, chosen from the uniform distribution in $(0, 200)$. Figure 3.5e shows the total cell numbers in 100 such realisations of the ABM.



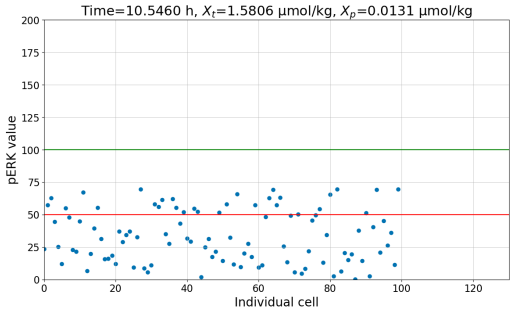
(a) ABM: Overall number of tumour cells in one realisation



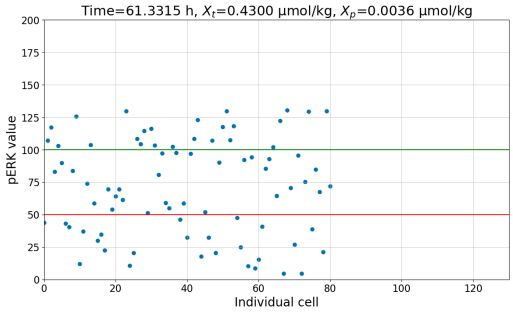
(b) ABM: Number of cells inside the division, death and quiescence pools over time in one population



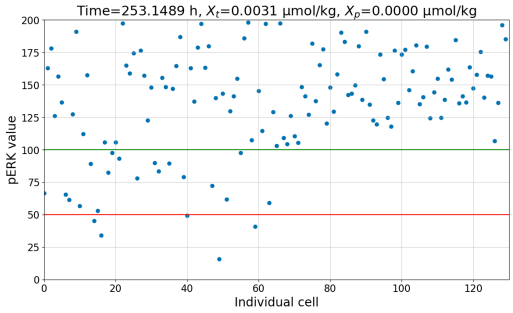
(c) ABM: pERK value of each cell in one population at 0 hours



(d) ABM: pERK value of each cell at 10.5 hours in one population

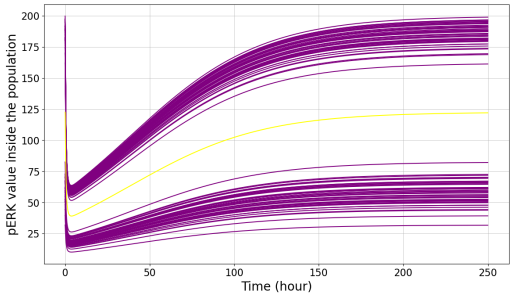


(e) ABM: pERK value of each cell at 61.3 hours in one population

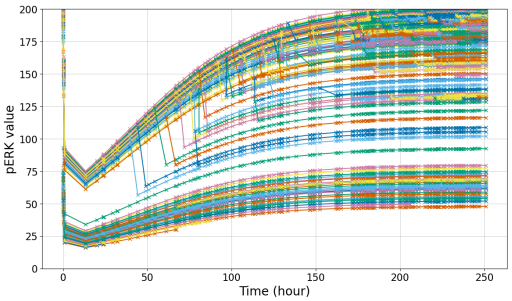


(f) ABM: pERK value of each cell at 253.1 hours in one population

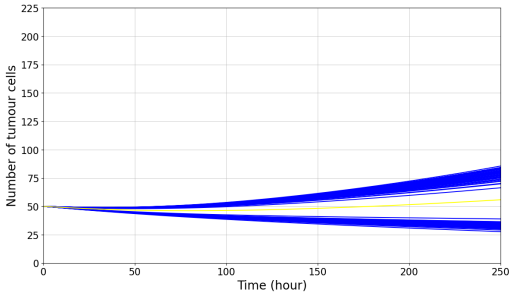
Figure 3.6: One realisation of the ABM. In Figures 3.6c – 3.6f, one dot represents the pERK value of one cell at one timepoint. The ABM may provide a more realistic model because it captures heterogeneity, different scales and emergent behaviour. On the other hand, ODE based models are suitable for modelling well-mixed compartments with mass transfer and simple interactions at one scale level. A video of the scatter plots can be found at https://github.com/VanThuyTruong/Tutorial/blob/main/videos/3mgkg%20single%20dose%20uniform%20distribution.mp4



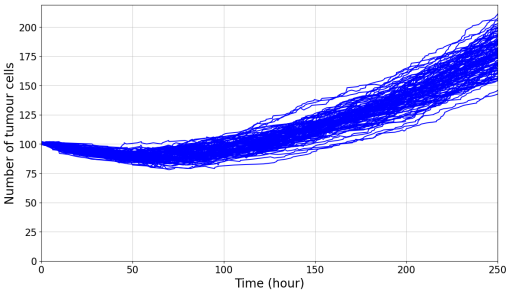
(a) ODE: pERK values. Each line corresponds to a different initial condition.



(b) ABM: pERK values of individual cells in one realisation

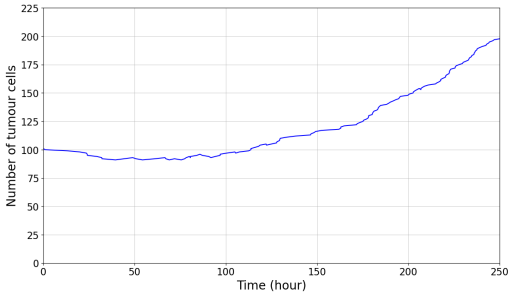


(c) ODE: tumour cell numbers in 100 tumour cell populations.

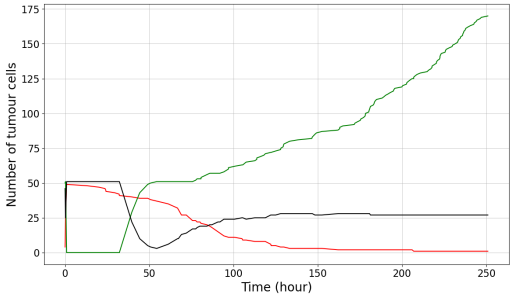


(d) ABM model: Overall tumour cell number in 100 realisations

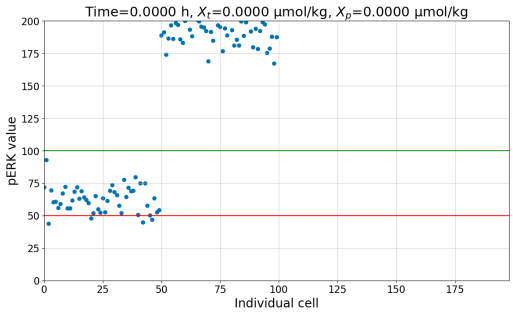
Figure 3.7: Simulation of a single dose of 3 mg/kg. pERK values are initially bimodally distributed. In the PKPD-ODE model the bimodal distribution is on the population level. 50 populations have a high pERK starting value and 50 populations have a low pERK starting value. In the PKPD-ABM the bimodal distribution is in the cellular level. Inside one population 50 cells have a high pERK starting value and 50 cells have a low pERK starting value. Figures 3.7a, 3.7c show the PKPD-ODE (model C1 in Figure 3.2), Figures 3.7b and 3.7d show the PKPD-ABM (model C2 in Figure 3.2) as comparison. A video of the scatter plots can be found at https://github.com/VanThuyTruong/Tutorial/blob/main/videos/bimodal%20pERK.mp4



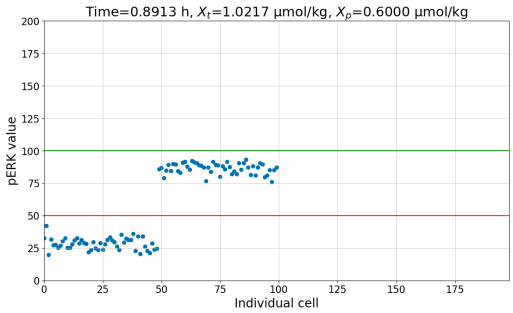
(a) ABM: Overall number of cells in one realisation



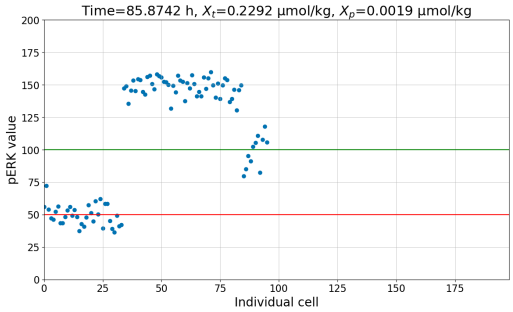
(b) ABM: Number of cells inside the division, death and quiescence pool over time in one population



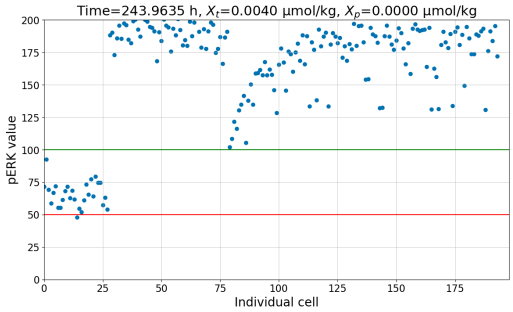
(c) ABM: pERK value of each cell in one population at 0 hours



(d) ABM: pERK value of each cell at 0.9 hours in one population

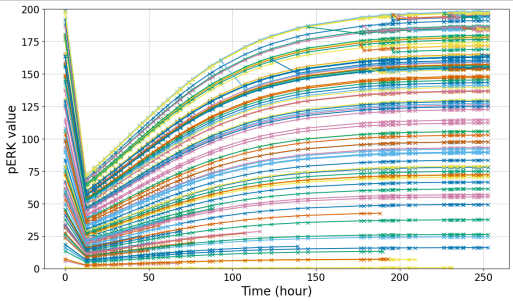


(e) ABM: pERK value of each cell at 85.9 hours in one population

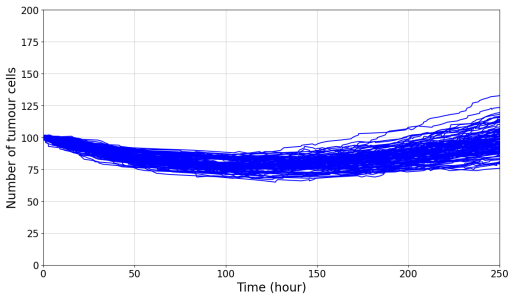


(f) ABM: pERK value of each cell at 243.9 hours in one population

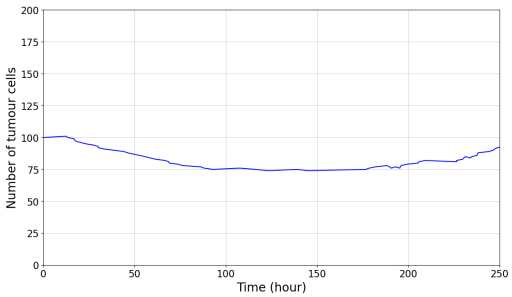
Figure 3.8: Figures 3.8a – 3.8f show the behaviour of one population in the PKPD-ABM (model C2 in Figure 3.2). A video of the scatter plots can be found at https://github.com/VanThuyTruong/Tutorial/blob/main/videos/bimodal%20pERK.mp4



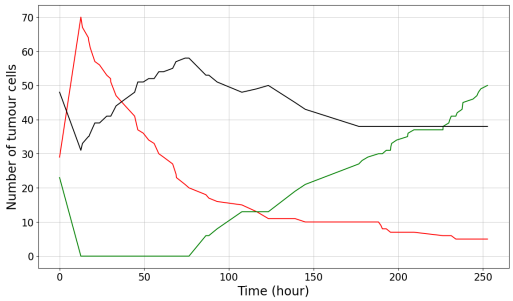
(a) ABM: pERK value of each cell in one population with with a division threshold of 150% and a death threshold of 100%



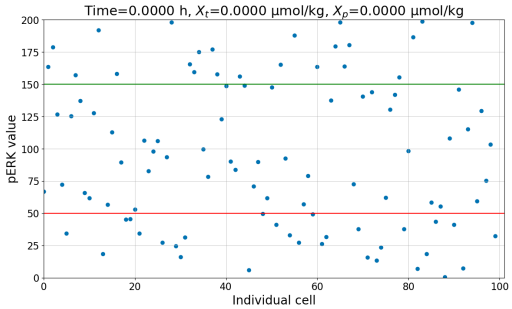
(b) ABM model: Overall tumour cell number of 100 realisations with a division threshold of 150% and a death threshold of 100%



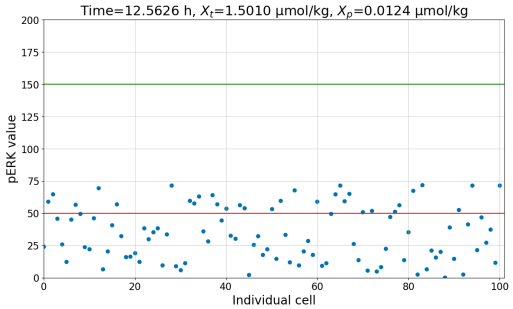
(c) ABM: Overall number of tumour cells within one realisation



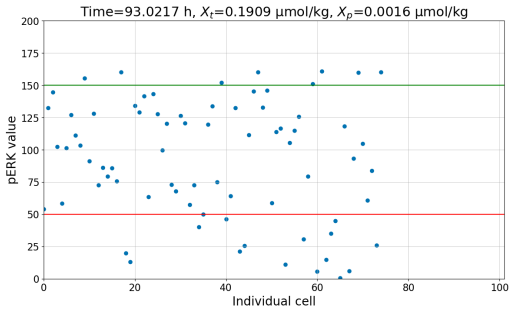
(d) ABM: Number of cells inside the division, death and quiescence pool over time in one population



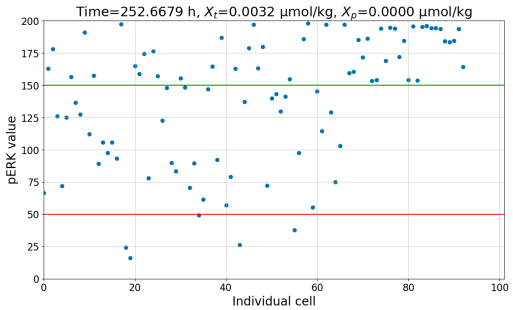
(e) ABM: pERK value of each cell in one population at 0 hours



(f) ABM: pERK value of each cell at 12.6 hours in one population

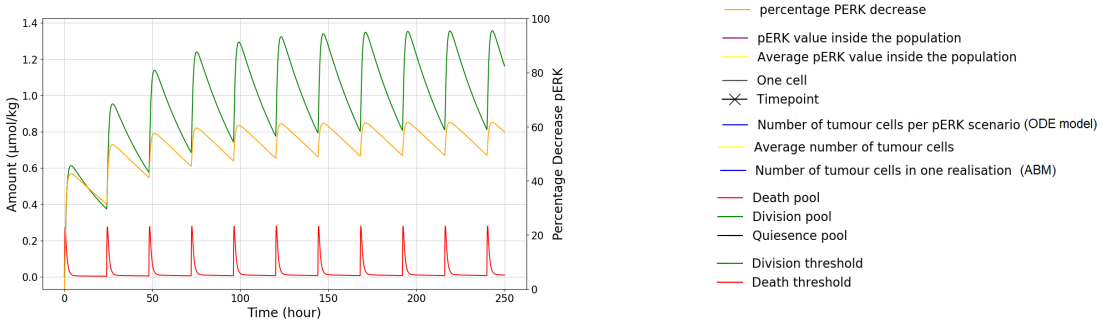


(g) ABM: pERK value of each cell at 93.0 hours in one population

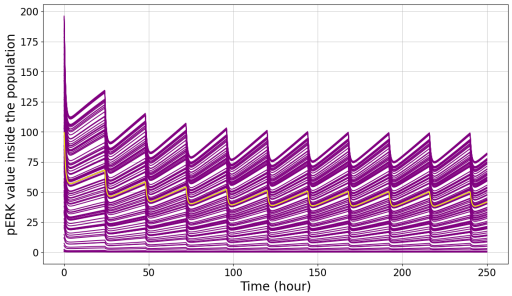


(h) ABM: pERK value of each cell at 253.7 hours in one population

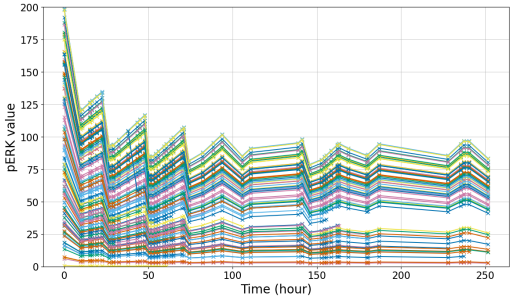
Figure 3.9: Simulation of a single dose of 3 mg/kg. pERK values are initially uniformly distributed. Figures 3.9a and 3.9b show the PKPD-ABM with a division threshold of 150% pERK, (model C2 in Figure 3.2) as comparison. Figures 3.9c – 3.9d show the behaviour of one population in the PKPD-ABM (model C2 in Figure 3.2). A video of the scatter plots can be found at https://github.com/VanThuyTruong/Tutorial/blob/main/videos/150%20division%20threshold.mp4



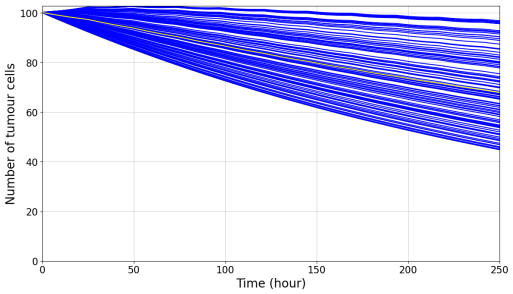
(a) PKPD model: Drug amount in the plasma and tumour compartment (PD, left-hand scale) and percentage pERK decrease (PD, right-hand scale) with a multiple dose treatment of 1 mg/kg daily



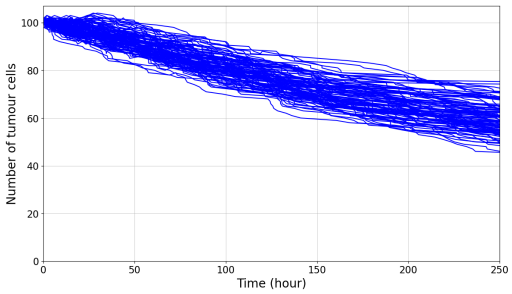
(b) ODE: Individual pERK value of every cell population



(c) ABM: Each line is the pERK value of one cell in one realisation

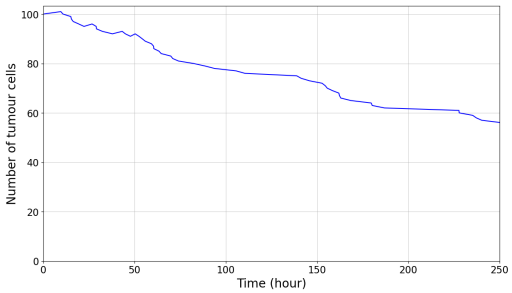


(d) ODE: Overall cell number of 100 tumour cell populations

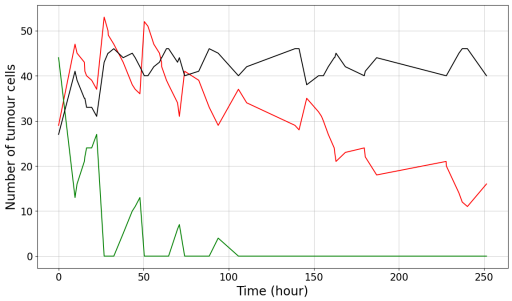


(e) ABM model: Overall tumour cell number of 100 realisations

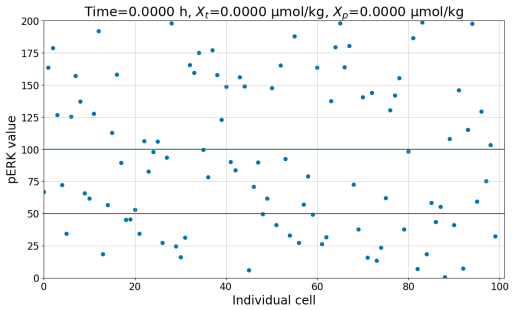
Figure 3.10: Simulation of a multiple treatment cycles of 1 mg/kg every 24 hours. pERK values are initially uniformly distributed. Figure 3.10a shows the PKPD model (model A,B in Figure 3.2), Figures 3.10b, 3.10d show the PKPD-ODE model (model C1 in Figure 3.2), Figures 3.10c and 3.10e show the PKPD-ABM (model C2 in Figure 3.2) as comparison. A video of the scatter plots can be found at https://github.com/VanThuyTruong/Tutorial/blob/main/videos/multiple%20cycles%201mgkg.mp4



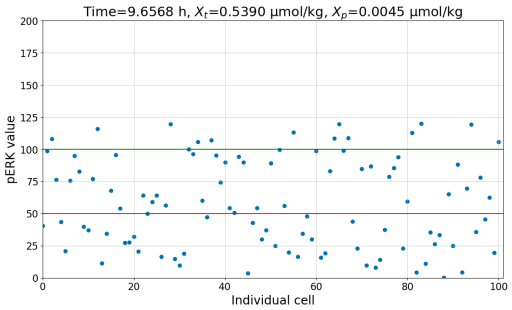
(a) ABM: Overall number of tumour cells within one realisation



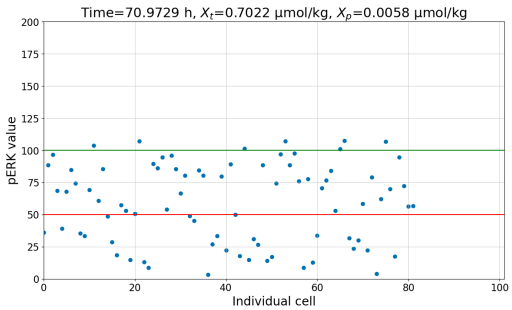
(b) ABM: Number of cells inside the division, death and quiescence pool over time in one population



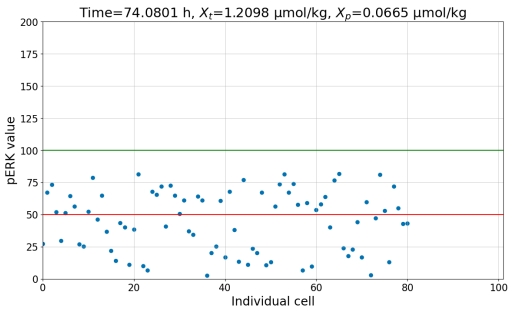
(c) ABM: pERK value of each cell in one population at 0 hours



(d) ABM: pERK value of each cell at 9.7 hours in one population



(e) ABM: pERK value of each cell at 70.9 hours in one population



(f) ABM: pERK value of each cell at 74.1 hours in one population

Figure 3.11: Figures 3.11a – 3.11f show the behaviour of one population in the PKPD-ABM (model C2 in Figure 3.2). A video of the scatter plots can be found at https://github.com/VanThuyTruong/Tutorial/blob/main/videos/multiple%20cycles%201mgkg.mp4

3.4 Discussion and conclusions

In this chapter, we present four simulation examples based on two types of models. Table 3.2 shows a head-to-head comparison of ABM and ODE based models based on their key properties (scale, dynamics of interactions, population, space, memory, stochasticity), strengths and limitations, and implementation general principles (model building and qualification, communication and applicability).

3.4.1 Model properties

A key feature of ABMs is their ability to incorporate heterogeneity and autonomy at the cellular/microscopic level. Agents can have different attributes, which leads to a heterogeneous population. There is no central coordination that determines the agent's behaviour. An behaviour of the biological system is simulated depending on individual attributes and rules that collectively result in an emergent behaviour. Local phenomena driven by discrete decision and interaction between individual agents on the microscopic scale lead to emergence of patterns on the macroscopic scale. Adaptation of the agents due to environmental changes is possible with rules set by the modeller, enabling memory to be stored to modify future behaviours depending on past stages. The incorporation of spatial component in ABMs enables to locate single agents geographically in the biological system on top of chronological information on their past and future fate (Milling and Schieritz 2003; Bauer et al. 2009; Solovyev et al. 2013; Figueredo et al. 2014). In our anti cancer treatment example, the agents are tumour cells with different levels of pERK and, depending on this attribute, show different behaviour (death, quiescence, or division). The tumour cells with high pERK values divide and pass their pERK values on to their offspring, while the cells with a lower pERK value die or stay quiescent. The pERK value decreases with high drug level but after lowering of drug amount in the tumour compartment the cells regain their inherited pERK value. Consequently, a population with high pERK values over time arises which could be a biologically plausible way of modelling treatment resistance. The variability is defined as an implicit feature of each agent and is governed by stochastic processes (Gillespie algorithm in this model). There are other examples in the literature which took advantage of the ABM properties. Cockrell and Axelrod (C. Cockrell and Axelrod 2019) use the ABM technique for simulating the heterogeneous cell types and their proliferation kinetics inside human colon crypts where quiescent stem cells are at the bottom of the crypt, proliferating cells are close to the bot-

tom third, and differentiated cells are placed in the top two thirds. Proliferating cells are killed by cytotoxic drugs while quiescent stem cells are resistant to cytotoxic drugs due to their low probability of dividing (C. Cockrell and Axelrod 2019). Due to their properties, ABMs are suitable to simulate the highly diverse tumour microenvironment where interaction of different cell types plays a critical role in cancer development, progression, and control. Kather et al created a two dimensional on-lattice ABM containing tumour cells, active and exhausted lymphocytes, stroma, and necrosis. This ABM was used to predict survival in an independent patient cohort, and guide new strategies for immunotherapy in colorectal cancer (Kather, Poleszczuk et al. 2017). Based on that model, Kather et al further developed a 3D on-lattice ABM of human solid tumour tissue including tumour cells, fibroblasts, and myeloid and lymphoid immune cells which reproduces key features of the tissue architecture of human colorectal cancer. Effects of chemotherapy, immunotherapies, and cell migration inhibitors alone and in combination were simulated (Kather, Charoentong et al. 2018).

In comparison, ODEs are typically used to represent a subset of patients' outcomes following therapeutic interventions at the macro-level (in this chapter, tumour burden for the PD measurement). Patients belong to an underlying homogeneous population and individuals inside a population behave independently (Solovyev et al. 2013; Figueredo et al. 2014). Individuals and interaction between individuals are not explicitly considered. Instead, macro-level outputs are driven by mass transfer dictated by stoichiometric equilibrium and compartmentalization of the system (e.g, blood and tumour compartments). The stochastic model is implemented by assuming parameter probability distributions at the individual level (between-subject variability) and in the measurement noise (residual variability).

Additionally, adaptation of the behaviour to changing circumstances such as environmental factors is limited unless pre-specified in the parameter settings and initial conditions. In most circumstances, ODEs do not have inherent memory features (Milling and Schieritz 2003; Figueredo et al. 2014). In the ODE model, all the cells in the population have the same pERK and death of one cell or division of one cell cannot be tracked individually. Space is not typically implemented and PDEs are required to model alterations in time and space. However, the use of PDEs can be complicated and computationally challenging (Bauer et al. 2009).

3.4.2 Model building and model qualification

Model construction in ABMs can happen in two ways based on hypothesis-testing and data calibration. First, ABMs can use model assumptions from empirically observed phenomena, and simulate previously unknown collective behaviours (micro-known to produce macro-unknown). In this way, predictions are made using validated theories of agent behaviours. Second, when data for a phenomena are not observable, an ABM can be built by using hypothetical model assumptions to reproduce empirically observed collective behaviour by simulation (micro-unknown to reproduce macro-known) (Sayama 2015). Model calibration is challenging when ABMs require a large number of parameters from single cell data (in vivo/in vitro) (Bauer et al. 2009) and small increases in agent based model complexity can lead to large increases in required calibration data (Srikrishnan and Keller 2021). In addition to the initial pERK distribution, death and birth rate in the ODE example, the ABM required death and birth thresholds for the decision-making algorithm. Consequently, there is the risk of overfitting and non-identifiability. When parameter values determine behaviour that is directly observable in experiments, those values can be calibrated from data. An example is the lattice-based multiscalar cellular automaton model of Delgado-SanMartin et al. where a modelling platform is used to obtain model parameters from multiple in-vitro assays (Delgado-SanMartin et al. 2017). Another example is the tumour micro-environment ABM from Kather et al which can be calibrated in a patient-specific way by using immuno-histochemical data (Kather, Charoentong et al. 2018). Analytical methods are required when parameters do not represent directly observable quantities. In that case, statistical estimation techniques such as maximum likelihood estimation or the method of moments are applied to a given dataset to select appropriate values. If analytical methods are not suitable for a given ABM, methods involving the generation of simulated data need to be considered. Those methods can be classified in frequentist approaches and Bayesian approaches. Frequentist approaches are distance-based or likelihood-based (e.g. the simulated minimum distance method or the methods of simulated moments) (Platt 2020). One example for calibration of ABMs using machine learning would be the work from Lamperti et al (Lamperti et al. 2018).

In contrast, the ODE model provides a robust statistical framework with model selection based on data-driven decision metrics. Algorithms for numerical solutions of ODEs are widely available and understood, as are methods of data fitting and calibration against observed data (Tornøe et al. 2004; Dartois et al. 2007). Data are needed from the macro

or population level. This method is often used for data fitting and it is calibrated against observed data with goodness-of-fit graphical analysis of residuals and predictions, distribution of the residual errors and random-effects parameters and simulation-based diagnostic tools such as visual predictive check (Tornøe et al. 2004). ODE models can have a descriptive purpose or predictive and descriptive purposes. ODEs have the advantage of being simple and, often, requiring fewer parameters from experiments than ABMs, but this comes with the risk of oversimplification. ODE models provide an approximation of biological processes that rely on stochasticity not being important at a whole organism level. Resulting from this limitation, there is a risk that this method leads to inadequate representation and spurious results when not used appropriately (Bauer et al. 2009). In principle, ODE models are based on population-level assumptions and data, whereas ABMs are constructed from understanding at the level of individual agents. For example, in vitro assays (Delgado-SanMartin et al. 2017) or immunohistochemical data (Kather, Charoentong et al. 2018) may inform a model. When data for macroscopic phenomena are not available, an ABM can be built as a hypothesis-testing device, where assumptions lead to observable collective behavior via simulation (Sayama 2015).

3.4.3 Strengths and limitations

One major strength of ABMs is the establishment of emergent behaviour not directly imposed by the modeller. As such, ABMs can simulate previously unknown collective behaviours (micro-known to produce macro-unknown) (Sayama 2015). This level of flexibility provides valuable information to provide biologically plausible hypothesis to explain macro-level phenomenon such as treatment resistance. Our example showed that the single dose drug treatment of 3 mg/kg is not enough to supress all cells into the death pool. Instead only cells with a low pERK value are erradicated by the drug treatment while the mutated cells with a high pERK value survive. The surviving cells will divide and inherit their high pERK value to the offspring. This leads to a more aggressive mutated tumour population with high pERK values. The simulation with bimodal distributed pERK values emphasised this hypothesis. Another hypothesis drawn from the multiple dose simulation is that cells with a high pERK value remain dormant in the quiescent pool during drug treatment. After the drug therapy is finished or paused they will return to the division pool, divide, and pass their high pERK value to their offsprings which causes a more aggressive mutated tumour population. Conversely, it is difficult to discriminate between

competing ABMs. Additionally, there is the risk of overparameterization. Due to their theoretical considerations and to the limited empirical data available for calibration, validation of agent based simulations can be difficult. In contrast, ODE based models have a more robust framework to evaluate their major assumptions with likelihood ratio tests and data-driven approach to discriminate between hierarchical model candidates. The model qualification is also well-established with simulation-based diagnostics (visual predictive checks) and goodness-of-fit criteria relating to the empirical data available. One limitation of ODE based models is the oversimplification of the biological system unless numerous differential equations are implemented (see Milberg et al (Milberg et al. 2019)) and its rigid framework given its inability to generate emerging behaviours in sub-scale systems.

3.4.4 Computational resources

ABMs can be computationally challenging, depending on their complexity and the chosen model paradigm (Metzcar et al. 2019). In the cellular potts models, changes in cell shape and direct cell-cell interaction are directed by Monte Carlo simulations and energy minimalization. Off-lattice methods bring the benefit of a more realistic simulation because cells can have various positions with respect to each other and freedom to move in any direction instead of being ordered on a grid. However, this comes with the disadvantage of higher computational cost, because special algorithms are necessary to efficiently handle cell-cell neighbourhoods. During movement, cell collisions need to be avoided, which can be challenging in densely packed areas or populations. After division, the placement of daughter cells needs to be determined to ensure non-overlapping with other cells or mutually exclusive cell areas. In addition, chemical values of the environment are usually computed on regular grids and interpolation techniques need to be applied to transfer values between the cellular off-lattice individuals and the chemical fields. Sub cellular element models have the benefit of a better approximation of cell bio-mechanics, but this comes with higher computational cost. Boundary-tracking methods are useful for describing detailed cell mechanics to fluid and solid tissue mechanics but are as well highly computational intensive (Metzcar et al. 2019; Rejniak and A. R. Anderson 2011). ODE based models, in contrast, require less computational resources due to their simple structure.

3.4.5 Model comparison and communication

When it comes to model communication and comparison, a strength of ABMs is the use of biological rules which makes communication of the model assumptions, and outputs, easier and more intuitive for a non-modelling audience. Components of two ABMs can be combined in a modular fashion to create meta-models. This modelling approach is currently less popular in the PKPD modelling community, but it has a lot of potential for simulating biological processes. One risk and important weakness of using ABMs lies in the context of sharing and comparing results. Each ABM can have different rules and outcomes vary in each run. This makes comparison between different models challenging.

On the other hand, advantages of ODE-based models are the simplicity of implementation, the direct interpretability of model parameters, their ability to combine different sub-models into one meta-model, and its applicability to fit experimental data in a statistically robust manner with clear decision rules for model selection and well-established tools for model qualification. Model comparison is less complicated due to similar model structure. Communication of the model with non-modelling audiences can be challenging because mass transfer and binding kinetics are less intuitive to communicate than a set of biological processes and rules relating to biological phenomena. All that being said, this modelling approach is commonly used in the PKPD community and numerous examples exist.

3.4.6 Applicability

The comparison between the two modelling approaches shows that ABMs are an object-oriented, rule-based, discrete-event computational models with heterogeneous agents in which the behaviour of individual agents, and pattern formation is crucial (Solovyev et al. 2013). The simulations are more intuitive and easier to interpret since they recapitulate processes closer to the biology. Granularity is high, for example direct information of cell behaviour dynamics in the tumour micro-environment is provided on the microscopic scale which can be summarised to the macroscopic scale to provide a tissue-specific overview. Therefore, ABMs are suitable to simulate complex biological systems with sub-scale components (molecular, cellular, tissue, organism) and inherent emerging behaviour. These model features are very relevant for developing quantitative solutions in research questions and clinical problems relating to system biology and quantitative clinical pharmacology.

On the other hand, systems of ODEs are well-suited for simulating processes that can be approximated as homogeneous, well-mixed systems with central coordination and would be best suited for traditional pharmacometrics analyses with sufficient data (population PK and PD models, PBPK models), or for simplistic theoretical PKPD models with limited binding interactions between model components. For quantitative clinical pharmacology models, ODE could also be implemented to recapitulate complex biological systems but would rely on extensive model assumptions, including parameter distributions (Milberg et al. 2019). Simulations are limited to one scale, and entities that are entered in the model do not have sub-scale inner ability. ODE-based models only account for variability at the observable level but not at the sub-scale level like ABM. Hence, ODEs are appropriate to simulate well-mixed compartments with mass transfer and simple interactions at one scale level.

3.4.7 Hybrid multi-scale models

To take advantage of both methods, hybrid multi-scale models can be employed. Multi-scale modelling aims to include various spatio-temporal scales from atomic to molecular, cellular, multi-cellular, organ, and up to multi-organ systems. On the cellular level, part of the model can be discretised using an ABM to study phenomena that are heterogeneous with emergent behaviour. Those phenomena can be for example effects of mutation-induced, more aggressive phenotypes on metastasis or cell and extracellular matrix interaction. Discrete modelling simulates the multi-step emergent process of changes of cell agents that leads to tumour growth and metastasis. Macroscopic scale modelling covers the entire tumour tissue and comprises many agents. This scale includes the general pattern of growth, total cell number, the extent of metastasis, tumour morphology and vascularization. When explicit representation of individuals is not needed, such as for homogenous entities, a continuous description with PDEs or ODEs can be used. On the molecular scale, molecular interactions (e.g. receptor-ligand interactions, consumption and production of oxygen, nutrient, and cell-cell signalling molecule concentration) can be described with ODEs. To model local conditions and environmental changes such as availability of oxygen, nutrient, and hormonal distribution through diffusion from molecularly rich regions (e.g. blood vessels, tumour edge), PDEs can be used. Solutions are provided for the entire scale i.e. the tumour tissue instead for each agent individually. Using ODEs or PDEs provides an easy way to implement the simulation with lower computational

cost in comparison to ABMs. Levels of molecular entities solved by ODEs and PDEs are part of the environment and can cause changes in cell behaviour of the corresponding ABM. For example, signalling molecules such as growth factors can initiate cascades that lead to proliferation. The hybridization of a discrete model in a continuum environment provides a more complete description of the tumour morphology, higher accuracy of model predictions, and lower computational costs (Zhihui Wang et al. 2015).

Comparing hybrid multi-scale ABMs to a multi-scale ODE model as in the work of Milberg et al (Milberg et al. 2019), we can see additional differences between those methods. Their physiology-based quantitative pharmacology model predicts how the interaction of the immune system and the tumour micro-environment in a patient affects the efficacies of checkpoint blockade therapies administered as mono-, combo- and sequential therapies. The model accounts for tumour draining lymph node priming and activation of naïve T cells into effector cells in the presence of mature antigen presenting cells and regulatory T cells, the subsequent migration of the effector cells to the tumour through systemic circulation, interaction in the tumour micro-environment which includes regulatory T cells and myeloid-derived suppressor cells, and cancer cells, the release of tumour-associated or tumour-specific antigens by the cancer cells by natural death and killing, the uptake of the tumour-associated or tumour-specific antigens by antigen-presenting cells, their maturation and subsequent migration back to the draining lymph node compartment for presentation to naive effector T cells for their priming and activation, which starts the cycle again. In contrast to the ABM technique, here space for local interactions are accounted for by separate compartments such as lymph node, blood, tumour, lung, GI tract, spleen, and liver, and the peripheral compartment. Instead of rules for interaction, the model consists of 282 ODEs and 218 algebraic equations to describe for example binding reaction, interaction between cells, and trafficking of cells with rates. Heterogeneity is represented in the percent expression of each immune checkpoint in the cancer cells as an input into the model (Milberg et al. 2019). In contrast, an ABM-based implementation of this model would require only six agents (with attributes in parentheses): Antigen, antigen presenting cells (resident, mature, and CTLA4 expression), CD8+ T-cells (naïve, primed, activated, CTLA4, and PD-1), Tregs (CTLA4 and PD1), myeloid derived suppressive cells, and tumour cells (proliferating and PD-L1). The behavior of these agents is then controlled by a total of 16 interaction rules, with each cell type making use of a subset of these. In principle, this could be much more efficiently implemented and easily

adapted should further interactions or cell types be required.

There are various of examples for hybrid multi-scale ABMs in the literature. Oduola and Li modeled cancer growth during drug therapy with lapatinib using a multi-scale model with a stochastic hybrid system (Oduola and X. Li 2018). Similar to our example, this model also uses a combination of ODE and ABM. Concentration of proteins and gene expression levels are represented with ODEs at the molecular level as homogenous entities. A PKPD model describes the relationship between drug concentration and the effect on protein expression, which is associated with cell proliferation. The cellular level contains a cellular automata model on a 2D grid. Each cell is modelled individually. A Markov chain determines cell fate (proliferation, decay, or quiescence). Transition probabilities are influenced by the downstream gene expression as levels modulated by the drug therapy and the conditions of the tumour micro-environment. At the multi-cellular level, the behaviour of the cell population leads the structure of the tumour tissue. Drug treatment can cause a progressive degree of cellular decay from the surface to the tumour core. (Oduola and X. Li 2018)

A multi-scale ABM of tumour angiogenesis is provided by Olsen et al. (Olsen and Siegelmann 2013). This model includes the molecular level (VEGF, diffusion), cellular level (genetic control, space), and tissue level (cells, blood vessels, angiogenesis). Single tissue and cancer cells are modeled with an agent in a three-dimensional grid. Tumour cell proliferates according to a spheroid cluster pattern, with growth limited by available oxygen. The agent's behaviour is determined by a 'life protocol' based on Hanahan and Weinberg's hallmarks of cancer (Hanahan and Weinberg 2000). This includes proliferation (including rate parameters, generation potential, and space restrictions), proliferation suppression mechanisms, self-testing at a checkpoint before the replication decision, repairing damage, and apoptosis (self-death) (Olsen and Siegelmann 2013).

Chaplain and Powathil developed two hybrid multi-scale models that study the effects of intra-cellular heterogeneity in cancer treatment (Chaplain and Powathil 2016). These models incorporate multiple interactions involved in the sub-cellular, cellular and micro environmental levels. The micro-environment contains the concentration of oxygen modeled with a PDE. At the subcellular level, the cell-cycle depends on concentration of Cdk-cyclin B complex, the APC-Cdh1 complex, the active form of the p55cdc-APC complex, the total p55cdc-APC complex, the active form of Plk1 protein and the mass of the

cell. These concentrations are described with ODEs. Cell-cycle, phase-specific chemotherapy drug concentration is simulated using PDEs. Lastly, cellular automaton and cellular potts model simulate the cellular level (Chaplain and Powathil 2016).

Another hybrid multi-scale model of Chaplain, Adamson, and Powathil studies the effect of the combination of radiation and chemotherapy (Powathil et al. 2013). In this model, cell cycle regulation and oxygenation status are incorporated into the model, since both critically affects radiation sensitivity. At the cellular level, a cellular automaton is modeled in a two-dimensional grid. The subcellular level contains cell cycle dynamics which are modeled by a system of ODEs. The micro-environment includes oxygen, which is supplied by randomly distributed blood vessels. The local oxygen concentration and the chemotherapeutic drug concentrations are modeled by a system of PDEs. A modified linear quadratic model studies the effects of radiation therapy. This includes radiation damage, effects of hypoxia, and cell-cycle in determining the cell-cycle phase-specific radiosensitivity. (Powathil et al. 2013)

The multi-scale compartment model by Gong et al. (Gong, Milberg et al. 2017) describes the biological processes involved in tumour development and anti tumour immune response. Cytotoxic T lymphocytes and cancer cells are modelled as agents in a three-dimensional space. Rules for division, migration, cytotoxic killing, and immune evasion are set. An ABM simulates the cellular-tissue scale (tumour micro-environment heterogeneity and immuno-architecture). PDEs describe the molecular scale (IL-2 secretion and transport).

Cess and Finley created a multi-scale agent based model of macrophages and T cells within the tumour micro-environment (Cess and Finley 2020). Tumour cells, M1 and M2 macrophages, and T cells are modelled as agents on a two dimensional lattice while diffusible factors such as IL-4 and IFN-γ are simulated with PDEs. To increase the biological detail, macrophages contain a mechanistic ODE model of intracellular signaling in response to IL-4 and IFN-γ which allows to predict the effects of specifically inhibiting a part of the intracellular signaling pathway. To improve computation time, neural networks are employed to reduce the mechanistic model into a simple input/output model. This modelling framework was used to study the cell-cell interaction in the micro environment such as the impact of immunosuppressive macrophages on T cell function, and how macrophage-based immunotherapies can reduce immunosuppression (Cess and Finley

2020).

3.4.8 Outlook

To sum up, ABMs can provide more detailed insights into complex biological systems and it can be complemented with ODEs in hybrid multi scale models. ABMs can be used as an implementation of mechanistic reasoning for hypothesis-testing and guide new experiments driven by the model. But to take this approach further, not only cell behaviour inside a tissue or organism could be simulated but virtual patient trials could be defined and calibrated with clinical data. One example is an ABM from Cockrell and An (R. C. Cockrell and Gary An 2018) which simulates the innate immune response and possible treatments for systemic inflammatory response syndrome. The ABM was used to simulate various treatment options with different dosing regimens and drug combination options to evaluate response of anti cytokine therapies (R. C. Cockrell and Gary An 2018). In anti cancer treatment modelling, ABMs could be calibrated with tumour size longitudinal data. This can then serve to optimise treatment protocol (e.g. by informing tumour biopsy sampling scheme, or pre-specify a patient enrichment strategy with inclusion/exclusion).

3.4.9 Conclusion

The comparison in this chapter shows that both methods have their strengths and weaknesses. Depending on modelling purpose, ABMs are suitable to simulate complex biological systems with sub-scale components (molecular, cellular, tissue, organism) and inherent emerging behaviour while ODEs are appropriate to simulate well-mixed compartments with mass transfer and simple interactions at one scale level. Both methods can complement each other in hybrid multi-scale models. With the advent of more single-cell experiments, spatial transcriptomics, and other technological advances in imaging, we anticipate that the use of ABMs in biology will increase with direct applications in PKPD and pharmacometrics to guide drug development of novel drugs in oncology and other therapeutic areas.

Comparator	**ODE based model**	**ABM**

Scale	• Macroscopic • Mean behavior at system level	• Microscopic • Individual behaviors at cellular level driving emergent behavior at the system level
Dynamics of interaction	Mass transfer dictated by stoichiometric equilibrium and compartmentalization of the system	Rule-based individual agent interaction and stochasticity
Individuals inside a population	Homogenous	Heterogenous
Space	Not typically implemented	Typically implemented
Memory	Not typically implemented	Typically implemented
Stochasticity	• Between subject variability and residual unexplained variability • Implemented at the population level by assuming parameter probabilistic distributions	• Implicit feature of single agent and governed by stochasticity • No distributional assumption
Model building	• Rigorous statistical framework for model selection • Data-driven	• Hypothesis generation and hypothesis-testing iterative learning • Simulation-based and/or data calibration depending on empirical evidence
Model qualification	• Simulation-based diagnostic (visual predictive check) • Data-based model qualification (goodness of fit)	• Model calibration based on single cell data (in vivo/ in vitro/ex vivo experiments) or any data source of relevance for the biological system of interest (micro-or macro-level)
Limitation	• Oversimplification • Structural rigidity (e.g., compartmentalization) • Scalability	• Overparameterization • Model discrimination • Uncertainty in outcome

Strengths	• Well established modeling framework • Simple implementation	• Emergent behavior to find plausible mechanisms for unforeseen outcomes (e.g., resistance or necrosis) • Easier to scale
Computational resources	Typically not a limitation unless large number of differential equations required	Complex ABMs demand high computational power and cause long running times
Model comparison	Straightforward due to similar model structure and model discrimination criteria	More complicated than ODE due to multiplicity of rules, choice of attributes and stochasticity driving emergent behaviours
Communication with audience	• Challenging due to abstract concept with mass transfer and binding kinetics (mathematical knowledge required) • Familiar concept with large example pool • Well established with regulatory agencies	• Easier to communicate with non-modeler audience as more biologically interpretable • Not extensively used by PKPD modelers and regulators • More challenging to defend as assumption-rich and less data-driven due to paucity of data at the subscale level
Applicability	• Modelling of well-mixed compartments with mass transfer and simple interactions at one scale level • PK models, PD models, and traditional pharmacometric models of exposure-response • Quantitative system pharmacology models with known or observable macro-level outcomes	• Simulation of complex biological systems with sub-scale components (molecular, cellular, tissue, organism) and inherent emerging behaviour • System biology models and quantitative system pharmacology models with limited empirical data and most relevant to elucidate unexpected behaviors (micro-known to produce macro-unknown)

Table 3.2: Comparison between ODE based model and ABM features and implementation principles for the model-based approach used to characterize PK and PKPD properties

Chapter 4

Stochastic pharmacodynamics: elimination of a heterogeneous population

4.1 Introduction

Cancer displays a dynamic nature, with genetic diversity driven by genomic instability contributing to the generation of tumour heterogeneity. Throughout the course of the disease, cancers tend to become more diverse. At a broader level, tumour heterogeneity can be categorised into intertumoural and intratumoural heterogeneity. Intertumoural heterogeneity pertains to variations between patients with the same type of tumour and is attributed to patient-specific factors like genetic variations, mutation profiles, and environmental influences. On the other hand, the understanding of intratumoural heterogeneity has grown with genomic medicine, referring to differences among tumour cells within a single patient. This can manifest as spatial heterogeneity, involving diverse subpopulations unevenly distributed across sites, or temporal heterogeneity, reflecting dynamic shifts in a tumour's genetic diversity over time (Dagogo-Jack and Shaw 2018). As cancers are exposed to treatments, they tend to become more diverse and complex. Resistance to treatment can develop due to the expansion of pre-existing subclonal populations or the emergence of drug-tolerant cells. Consequently, subsequent lines of therapy often yield less robust responses compared to initial treatments (Dagogo-Jack and Shaw 2018). In the context of intertumour heterogeneity, oncogenic mutations often guide treatment de-

cisions. Two scenarios arise: mutations can predict treatment ineffectiveness, as seen in metastatic colorectal cancer where anti-EGFR therapy works for KRAS wild-type but not KRAS-mutant tumours. Alternatively, certain tumours depend on oncogenic mutations and targeting these has proven beneficial, such as BRAF-mutant melanomas responding to a BRAF inhibitor (Melo et al. 2013). Leveraging intratumoural heterogeneity for therapeutic gain is a possibility. The idea of "evolutionary double-blind therapy" involves sequentially administering two treatments that prompt tumour cells to adapt in ways that make them vulnerable to the second treatment. While not yet applied clinically, this concept highlights the potential of understanding intratumoural heterogeneity in advancing medical oncology. (Fisher et al. 2013). Heterogeneity is essential not only in the field of oncology but also in various other domains. For instance, the expression of proteins differs among individual cells, and bacteria exhibit varying susceptibility to antibiotics (Gefen and Balaban 2009).

Agent based models naturally encompass this diversity and lend themselves well to computational investigations. However, converting these models into simple formulas is often a challenge (An et al. 2017; Truong et al. 2022). In this context, we present and analyze stochastic models involving a heterogeneous population of cancer cells subjected to drug influence. Our focus in this chapter is on a drug's impact designed to eliminate a group of tumour cells. Our models incorporate the concept of cell-to-cell variability, wherein each tumour cell possesses a unique attribute termed the 'regulator value', intricately linked with cell viability. Once the drug effectively reduces a cell's regulator value below a certain threshold, the cell enters the 'death pool', signifying its susceptibility to demise. Our model is inspired from an agent based model in which the regulator is phosphorolated ERK *in vivo*, but can also be applied to pharmacodynamics *in vitro*. We have in mind the reduction of phosphorolation of the ERK pathway by oral dose of the MEK inhibitor cobimetinib. The PD effect is modelled by a single variable that represents cellular pERK, as in the population-based PKPD model of Wong *et al.* (Wong et al. 2012) (used to fit tumour concentrations and pERK data) and the agent based model in chapter 3 and published in Truong *et al* (Truong et al. 2022). Our model resembles some of the simplest pharmacodynamic models in that a cell's regulator value decreases exponentially, with rate δ, when the drug is present. The key stochastic aspect of this system is the random death of cells whose regulator value is sufficiently small. An advantage of this stochastic approach is its inherent extinction point: the eradication of the last tumour

cell. The probability distribution of this 'extinction time' can be computed numerically and, in simpler models, even analytically. When utilizing ODEs, estimating the mean extinction time is only possible by tracking when a trajectory reaches a suitable lower bound. Thus, our question revolves around determining the time until the drug-induced elimination of all tumour cells.

4.2 Single sustained dose, no cell division

4.2.1 Single cell behaviour

To simulate a heterogenous tumour cell population, we define:

- The number of tumour cells is an integer that may reach zero in finite time.
- Each cell has a regulator value at any time t, scaled to the interval $(0, 1)$.
- We use the uniform distribution on the interval (0,1) as the initial condition for the regulator value of each cell.
- The action of the drug on an individual cell's regulator value is given by a deterministic relationship.
- $1/\delta$ describes the exponential decay rate of k caused by the drug.
- $1/\mu$ is the mean time for cells to die once their k value is below 0.25 and they arrive in the death pool
- A cell i has regulator value $k_i(t)$ at time t.
- The probability that this cell is alive at time t and dies before $t + \Delta t$ is $w(k_i(t))\Delta t$. Thus, $w(k_i(t))$ is its death rate.

The probability that a cell, that is alive at time 0 with regulator value k, is still alive at time t is found using the hazard-rate formula (N. Holford 2013)

$$s(t, k) = \mathbb{P}\left(\text{cell survives to time } t | k_i(0) = k\right) = \exp\left(-\int_0^t w(k_i(s))\mathrm{d}s\right). \tag{4.1}$$

Using (4.1), we calculate $S(t)$, the probability that a tumour cell chosen at random from the initial population, is still alive at time t.

We assume that the regulator value $k_i(t)$ of a cell i decreases exponentially if a constant

drug dose is administered.

$$k_i(t) = k_i(0)\exp(-\delta t), \tag{4.2}$$

and that cells with $k_i(t) < 0.25$ are in the death pool. In the simplest case, the death rate of any cell in the death pool is equal to a constant μ, so that $w(k_i(t)) = \mu$. Cells with $k_i(0) > 0.25$ are not initially in the death pool but enter the pool when their regulator value has decreased to 0.25. Figure 4.1 is a small-scale illustration of the resulting dynamics: the ten cells initially present die, one by one, as their regulator values decline under the influence of the drug, taking them into the death pool.

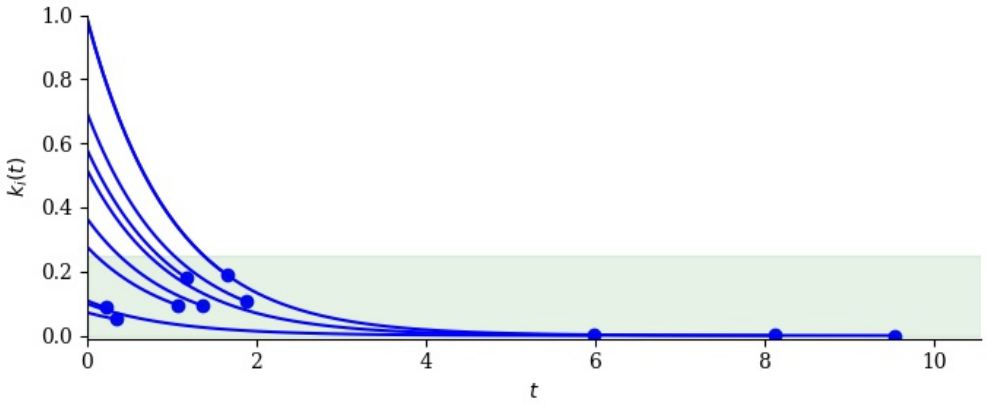


Figure 4.1: Illustrating the drug effect on a population of 10 cells. Each blue line is the regulator attribute of one cell as a function of time. Lines terminate in blue dots that indicate the death of a cell. The cells have different initial regulator values and enter the death pool (green shaded area) once the drug has decreased their regulator value below 0.25. The parameter values are $\delta = 0.2$ and $\mu = 1$.

To calculate the time a cell with the initial regulator value $k_i(0) = k$ enters the death pool t_k, we can use Equation (4.2). By setting $k(t_k) = 0.25$, one gets

$$t_k = \inf\{t \geq 0 :\ k_i(t) \leq 0.25 \mid k_i(0) = k\} \ = \begin{cases} 0 & 0 \leq k \leq \frac{1}{4} \\ \frac{1}{\delta}\log(4k) & \frac{1}{4} < k \leq 1. \end{cases} \tag{4.3}$$

This time is shown on the LHS in Figure 4.2. If the regulator value $k_i(t)$ is already equal or below 0.25, the t_k is zero, otherwise the time to enter the death pool increases logarithmically with k. On the RHS of Figure 4.2, two possible functions $w(k)$ are shown, a constant and non-constant k dependent death rate. For a constant death rate $w_0(k)$, we consider that the death rate is a constant value μ if the regulator value of a cell is 0.25 or

lower, meaning that the cell has entered the death pool. Otherwise, the death rate is set to be 0:

$$w_0(k) = \begin{cases} 0 & k > \frac{1}{4} \\ \mu & k < \frac{1}{4}. \end{cases} \tag{4.4}$$

For a non-constant death rate we use (4.1) with

$$w_{nc}(k) = \max\{\mu(1-4k), 0\}. \tag{4.5}$$

This is shown in the blue graph on the LHS of Figure 4.2. The death rate in the non-constant case is zero until the regulator value decreases below 0.25. Then the death rate increases linearly with the decrease of k and is the highest at $k = 0$.

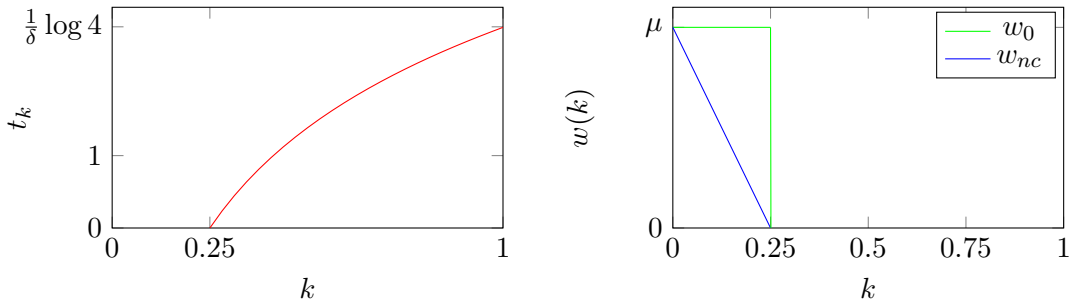


Figure 4.2: Left: The time, t_k, at which a cell enters the death pool is shown as a function of the cell's initial scaled regulator value k. Right: two functions: a constant death rate $w_0(k)$ and a k dependent death rate $w_{nc}(k)$.

Survival and death of individual cells with a constant death rate

Now we consider the case of a constant death rate. We consider a cell chosen at random from a population with initial regulator values uniformly distributed between 0 and 1. By integrating over $0-1$, we obtain the average survival probability $s(t,k)$ value of that population. Therefore, the probability that it survives to t is obtained by integrating (4.1):

$$S(t) = \mathbb{P}\left(\text{randomly-chosen cell survives to time } t\right) = \int_0^1 s(t,k)\mathrm{d}k. \tag{4.6}$$

To compute the probability that the cell i with initial regulator value $k_i(0) = k$ survives to time t, we need to consider the time until that cell arrives into the death pool.

To do so, we evaluate the integral in (4.1) using (4.4):

$$s(t,k) = \mathbb{P}\left(\text{cell survives to time } t | k_i(0) = k\right) = \begin{cases} 1 & t \leq t_k \\ \mathrm{e}^{-\mu(t-t_k)} & t > t_k. \end{cases} \tag{4.7}$$

Figure 4.3 shows $s(t,k)$ as a function of t using four values of k (upper) and as a function of k using four values of t (lower). It can been seen in the upper graph that for k values bigger than 0.25, the cell has initially a constant survival probability of 1 which decreases exponentially with $\mathrm{e}^{-\mu(t-t_k)}$ once the cell has entered the death pool after the time span t_k.

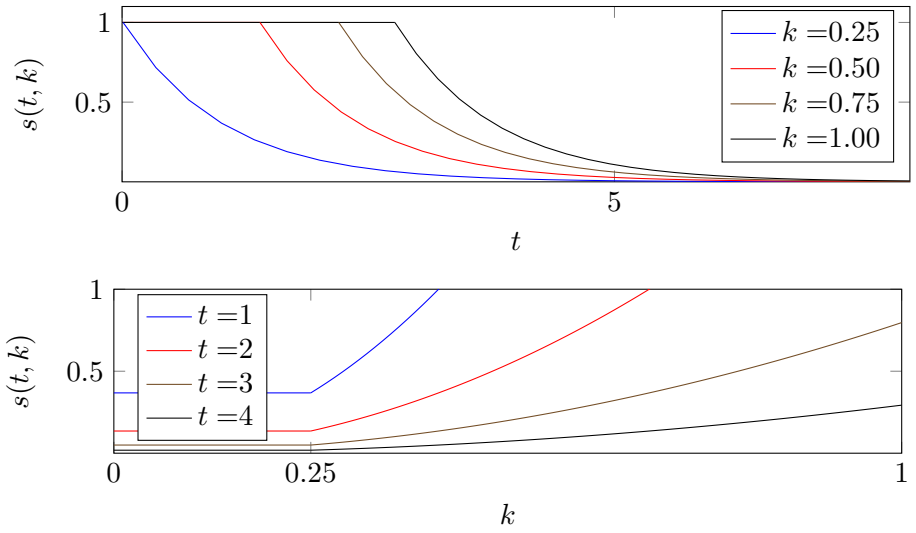


Figure 4.3: Upper: $s(t,k)$ is the probability that a cell, whose initial regulator value $k_i(0)$ is equal to k, is still alive at time t when $w(k) = w_0(k)$. If k is fixed then $s(t,k)$ is a decreasing function of t. Lower: The probability that a cell, whose initial regulator value is k, is still alive. If t is fixed then $s(t,k)$ is an increasing function of k. The formula used is (4.7), with $\mu = 1$ and $\delta = 0.5$.

In the lower panel of Figure 4.3 we can see the survival probability of a cell at different time points as a function of its regulator value k. We can see that for a k value in $(0, 0.25)$, the survival probability is constant. We assign that probability the function $S_1(t)$. Additionally, we can see that the survival probability depends on k for cells with a k value from 0.25 till $k = \frac{e^{t_k \delta}}{4}$. We assign that survival probability the function $S_2(t)$. When $t \leq t_k$ some cells are still outside the death pool and have the survival probability of 1 with their k value in $(\frac{e^{t_k \delta}}{4}, 1)$. We assign those cells the survival probability function $S_3(t)$. At times $t \geq t_k$, all cells have reached the death pool, but cells with a k value in $(0.25, 1)$ have a different survival probability than cells which had a k value below 0.25 at

$t = 0$. For those cells we assign the survival probability $S_4(t)$.

The cells which already have a $k \leq 0.25$ value at $t = 0$ and are in the death pool and remain there. For those cells we need to evaluate the integral in (4.6) as follows:

$$S_1(t) = \int_0^{\frac{1}{4}} s(t,k)\mathrm{d}k = \int_0^{\frac{1}{4}} \mathrm{e}^{-\mu(t-t_k)}\mathrm{d}k = \frac{1}{4}\mathrm{e}^{-\mu t}.$$

Since $t_k = 0$ for a $k \leq 0.25$, this leaves us with the term $\frac{1}{4}\mathrm{e}^{-\mu t}$.

Now we are looking into the case when $t < t_k$ and not every cell's regulator value has reached the death threshold if we chose a cell at random from a population with initial regulator values uniformly distributed between 0 and 1. This can be seen in the blue and red line in the lower panel in Figure 4.3. Apart from the cells with $k \leq 0.25$, we are evaluating the integral in two parts: one part contains the cells that have a survival probability of 1 and the other part has a decreasing survival probability as they have already reached the death pool at that given time point. The critical k value for entering the death pool at a given time can be deducted from Equation 4.3 which gives us

$$\begin{aligned} t_k &= \frac{1}{\delta}\log 4k \\ t_k\delta &= \log 4k \\ e^{t_k\delta} &= 4k \\ k &= \frac{e^{\delta t_k}}{4}. \end{aligned}$$

Therefore, we integrate from 0.25 till $\frac{e^{\delta t_k}}{4}$ and from $\frac{e^{\delta t_k}}{4}$ till 1.

$$S_2(t) = \int_{0.25}^{\frac{e^{\delta t_k}}{4}} \mathrm{e}^{-\mu(t-t_k)}\mathrm{d}k = \int_{0.25}^{\frac{e^{\delta t_k}}{4}} \mathrm{e}^{-\mu(t-\frac{1}{\delta}\log(4k))}\mathrm{d}k = \mathrm{e}^{-\mu t}\int_{0.25}^{\frac{e^{\delta t_k}}{4}} (4k)^{\delta/\mu}\mathrm{d}k$$

$$S_3(t) = \int_{\frac{e^{\delta t_k}}{4}}^{1} 1\mathrm{d}k = 1 - \frac{1}{4}\mathrm{e}^{\delta t}.$$

At times $t \geq t_k$, all cells have reached the death pool and we consider only the integral from 0.25 till 1 which gives us

$$S_4(t) = \int_{0.25}^{1} \mathrm{e}^{-\mu(t-t_k)}\mathrm{d}k = \int_{0.25}^{1} \mathrm{e}^{-\mu(t-\frac{1}{\delta}\log(4k))}\mathrm{d}k = \mathrm{e}^{-\mu t}\int_{\frac{1}{4}}^{1} (4k)^{\delta/\mu}\mathrm{d}k.$$

To summarise the partial integrals from above, the survival probability of a cell chosen at random from a population with initial regulator values uniformly distributed between 0 and 1 is as follows:

$$\begin{aligned} S(t) &= \frac{1}{4}\mathrm{e}^{-\mu t} + \int_{\frac{1}{4}}^{1} s(t,k)\mathrm{d}k \\ &= \frac{1}{4}\mathrm{e}^{-\mu t} + \begin{cases} S_2(t) + S_3(t), & \text{if} \quad \delta t < \log 4 \\ S_4(t), & \text{if} \quad \delta t \geq \log 4. \end{cases} \\ &= \frac{1}{4}\mathrm{e}^{-\mu t} + \begin{cases} \mathrm{e}^{-\mu t}\displaystyle\int_{\frac{1}{4}}^{\frac{1}{4}\mathrm{e}^{\delta t_k}} (4k)^{\mu/\delta}\mathrm{d}k + 1 - \frac{1}{4}\mathrm{e}^{\delta t}, & \text{if} \quad \delta t < \log 4 \\ \mathrm{e}^{-\mu t}\displaystyle\int_{\frac{1}{4}}^{1} (4k)^{\delta/\mu}\mathrm{d}k, & \text{if} \quad \delta t \geq \log 4. \end{cases} \end{aligned} \tag{4.8}$$

Note that, if $k > \frac{1}{4}$ then $\mathrm{e}^{\mu t_k} = (4k)^{\mu/\delta}$ as $t_k = 1/\delta \log(4k)$. On the RHS in (4.8), the term $\frac{1}{4}\mathrm{e}^{-\mu t}$ corresponds to cells that are in the death pool at $t = 0$ (and remain there). The term $1 - \frac{1}{4}\mathrm{e}^{\delta t}$ is the fraction of cells that are not in the death pool. Solving the integrals yields,

$$S(t) = \begin{cases} 1 - \dfrac{\mu}{4}\dfrac{\mathrm{e}^{\delta t} - \mathrm{e}^{-\mu t}}{\delta + \mu}, & \text{if} \quad \delta t \leq \log 4 \\ A\mathrm{e}^{-\mu t}, & \text{if} \quad \delta t \geq \log 4, \end{cases} \tag{4.9}$$

where

$$A = \int_0^1 \mathrm{e}^{\mu t_k}\mathrm{d}k = \frac{4^{\mu/\delta} + \frac{1}{4}\frac{\mu}{\delta}}{1 + \frac{\mu}{\delta}}. \tag{4.10}$$

The factor A describes the delay time for the drug to act and is an increasing function of the ratio μ/δ, with $A = 1$ when $\mu/\delta = 0$. In the limit $\mu/\delta \to 0$, the action of the drug is fast compared to the typical survival time of a cell in the death pool and there is no delay time for the drug to take effect.

The cumulative distribution function for the death times is $F(t) = 1 - S(t)$.

The probability density of single-cell death times, shown as the top panel in Figure 4.6,

is obtained from the derivative of $F(t)$:

$$f(t) = F'(t) = 1 - S'(t) = \begin{cases} \dfrac{\mu\delta \mathrm{e}^{\delta t} - \mu^2 \mathrm{e}^{-\mu t}}{4(\delta + \mu)}, & \text{if} \quad \delta t \leq \log 4 \\ \\ \mu A \mathrm{e}^{-\mu t}, & \text{if} \quad \delta t \geq \log 4. \end{cases} \tag{4.11}$$

The mean time to death of a randomly chosen single cell will be denoted $\mathbb{E}(\tau_1)$. It is equal to the sum of the mean time to arrive in the death pool plus the mean time to die once in the death pool:

$$\mathbb{E}(\tau_1) = \int_{\frac{1}{4}}^{1} t_k \mathrm{d}k + \frac{1}{\mu} = \frac{1}{\delta}(\log 4 - \frac{3}{4}) + \frac{1}{\mu}. \tag{4.12}$$

The mean time to arrive in the death pool is the average of all cells that are not in the death pool at $t = 0$, hence we integrate over 0.25 till 1. Once the cells are in the death pool, the time to die follows the exponential distribution with mean $\frac{1}{\mu}$.

The variance of τ_1 is equal to $\dfrac{1}{\mu^2}$. In other words, the variability of the mean time to death of a single cell is a function only of the time the cells take to die once they arrive in the death pool as in the death pool all the cells have the same probability to die.

Survival and death of individual cells with a non-constant death rate

For a non-constant death rate we use Equation (4.1) with

$$w_{nc}(k) = \max\{\mu(1 - 4k), 0\}. \tag{4.13}$$

The death rate is 0 if the cell's regulator value is 0.25 or above. Otherwise it will increase with a decreasing $k_i(t)$ value. Thus a cell's survival probability (as seen in Figure 4.4) is a function of its initial regulator value, given by

$$\begin{aligned} s_{nc}(t, k) &= \exp\left(-\mu \int_{t_k}^{t} (1 - 4k_0(s))\mathrm{d}s\right) \\ &= \exp\left(-\mu(t - t_k - \frac{4k_0}{\delta}(1 + \mathrm{e}^{-\delta(t - t_k)}))\right), \end{aligned} \tag{4.14}$$

with $t_k = 0$ if $k_i(0) \leq 0.25$.

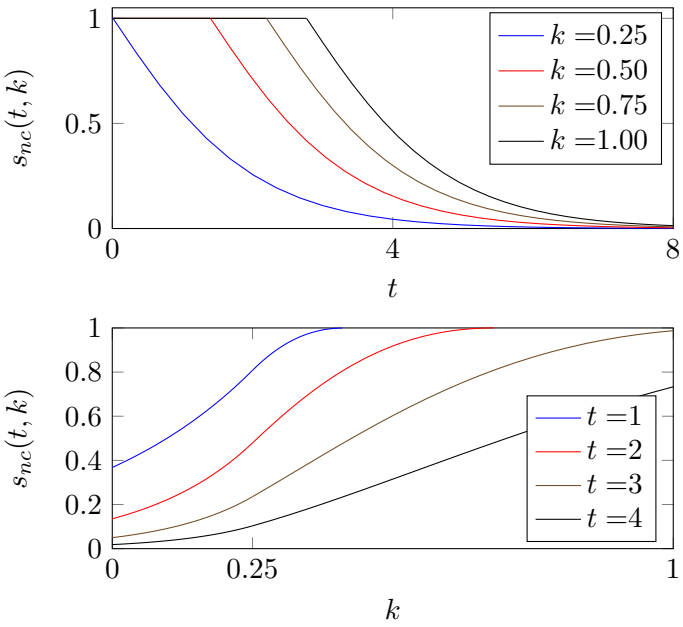


Figure 4.4: Upper: $s_{nc}(t,k)$ is the probability that a cell, whose initial regulator value $k_i(0)$ is equal to k, is still alive at time t. Lower: The probability that a cell, whose initial regulator value is k, is still alive at time t. The formula used is (4.14), with $\mu = 1$ and $\delta = 0.5$.

As $\delta t \to \infty$,

$$s_{nc}(t,k) \to \begin{cases} \mathrm{e}^{-\mu t}\mathrm{e}^{4k\mu/\delta}, & \text{if} \quad k \leq \frac{1}{4}, \\ \mathrm{e}^{-\mu t}(4k)^{\mu/\delta}\mathrm{e}^{\mu/\delta}, & \text{if} \quad k > \frac{1}{4}. \end{cases}$$

Now we only consider large times when $t > t_k$. This gives us

$$\begin{aligned} S_{nc}(t) =& \mathbb{P}\left(\text{randomly-chosen cell survives to time } t\right) \\ =& \int_0^1 s_1(t,k)\mathrm{d}k \to A_{nc}\mathrm{e}^{-\mu t} \qquad \text{as } \delta t \to \infty, \end{aligned} \tag{4.15}$$

where

$$A_{nc} = \mathrm{e}^{\mu/\delta}\frac{4^{\mu/\delta} - \frac{1}{4}}{1 + \frac{\mu}{\delta}} + \frac{\delta}{4\mu}(\mathrm{e}^{\mu/\delta} - 1). \tag{4.16}$$

4.2.2 Extinction of a cohort of n tumour cells

Suppose there are n tumour cells at $t = 0$, with regulator values uniformly distributed in $(0, 1)$. How long until all n cells die? An example realisation, with $n = 100$, calculated using the single sustained dose model described in Section 4.2, is shown in Figure 4.5. That is, the blue line is a number of cells surviving to time t when each, independently, is assigned an initial regulator value in $(0, 1)$ and, under the action of the drug, enters the death pool.

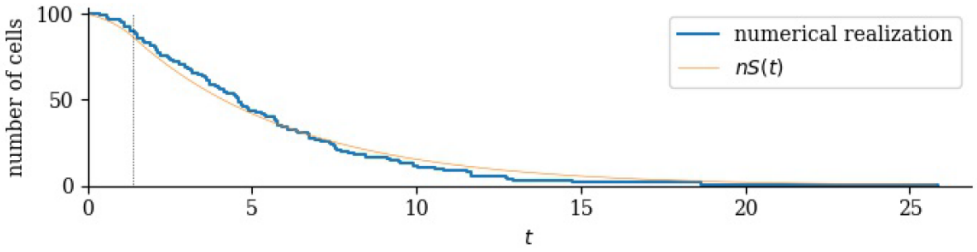


Figure 4.5: Blue: The number of surviving cells as a function of time in one realisation. Also shown is the smooth function obtained by averaging over many realisations, equal to the survival function $S(t)$ from Equation (4.9) multiplied by the initial number of cells. The vertical dotted line indicates $t = \log 4/\delta$. Here $n = 100$, $\delta = 1$ and $\mu = 0.2$.

We define the random variable N_t to be the number of cells alive at time t, with $N_0 = n$. Let τ_n be the first time that $N_t = 0$. Inspecting (4.9), we see that the single-cell survival probability has a simple exponential form as long as $\delta t > \log 4$. The form is taken if an individual lifetime is the sum of a fixed time of duration $\log A/\mu$ and an exponentially-distributed time with mean $1/\mu$. Each individual has a life time which is a constant plus a exponential distributed random variable. This simplifies to a pure death process.

The exponential random $T \sim Exp(\mu)$ variable has the following cumulative function:

$$\mathbb{P}(T \leq t) = 1 - e^{-\mu t}$$

Our modified variable considering one cell dies before time t is

$$\mathbb{P}(T \leq t) = \begin{cases} 0, & \text{if} \quad t < t_k, \\ 1 - e^{-\mu(t-t_k)}, & \text{if} \quad t \geq t_k, \end{cases}$$

which we derivative to get probability density function:

$$f(t) = \mu e^{\mu t_k} e^{-\mu t}.$$

This is for one fixed value of t_k, to average over all values of k, we use A:

$$A = \int_0^1 \mathrm{e}^{\mu t_k} \mathrm{d}k$$

The constant for the delay is A which describes the time the drug needs to take effect and suppress the cells into the death pool. The delay time for the drug effect is $\frac{\log A}{\mu}$ and the extinction time for the death process is $\frac{1}{\mu}\left(1 + ... + \frac{1}{n}\right)$. Using the harmonic series $\sum_{n=1}^{\infty} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ... + \frac{1}{n} \simeq \log n + \gamma$, where $\gamma = 0.577\ldots$, this can be simplified. Therefore, the time to extinction of n such individuals is given by (Renshaw 2011, Chapter 2.2.1)

$$\mathbb{E}(\tau_n) = \frac{1}{\mu}\left(\log A + 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) \simeq \frac{1}{\mu}\left(\log nA + \gamma\right). \tag{4.17}$$

For the variance of the time to extinction, the delay for the drug effect is not important. Hence, the variation can be described as $\frac{1}{(n\mu)^2} + \frac{1}{((n-1)\mu)^2} + ...$ For a large n, it can be simplified using the Basel problem which states that $\sum_{n=1}^{\infty} \frac{1}{n^2} \simeq \frac{\pi^6}{6}$. This gives us

$$\mathrm{var}(\tau_n) = \frac{1}{\mu^2}\left(1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2}\right) \simeq \frac{1}{\mu^2}\frac{\pi^2}{6}.$$

As can be seen in Figures 4.5 and 4.6, a considerable simplification arises because typical values of τ_n are large compared to $1/\delta$. When $\delta t > \log 4$, $\mathbb{P}\left(\text{randomly-chosen cell dies before time } t\right) = 1 - A\mathrm{e}^{-\mu t}$, and we are able to derive the probability density of τ_n explicitly.

Because each cell is independent, when t is larger than $\log 4/\delta$,

$$\mathbb{P}\left(\tau_n < t\right) = \mathbb{P}\left(n \text{ cells all die before } t\right) = (1 - A\mathrm{e}^{-\mu t})^n, \tag{4.18}$$

and to obtain the probability density of τ_n we derive Equation (4.18)

$$f_n(t) = \mu n A \mathrm{e}^{-\mu t}\left(1 - A\mathrm{e}^{-\mu t}\right)^{n-1}, \tag{4.19}$$

which attains its maximum value when $nA\mathrm{e}^{-\mu t} = 1$. Figure 4.6 shows the probability

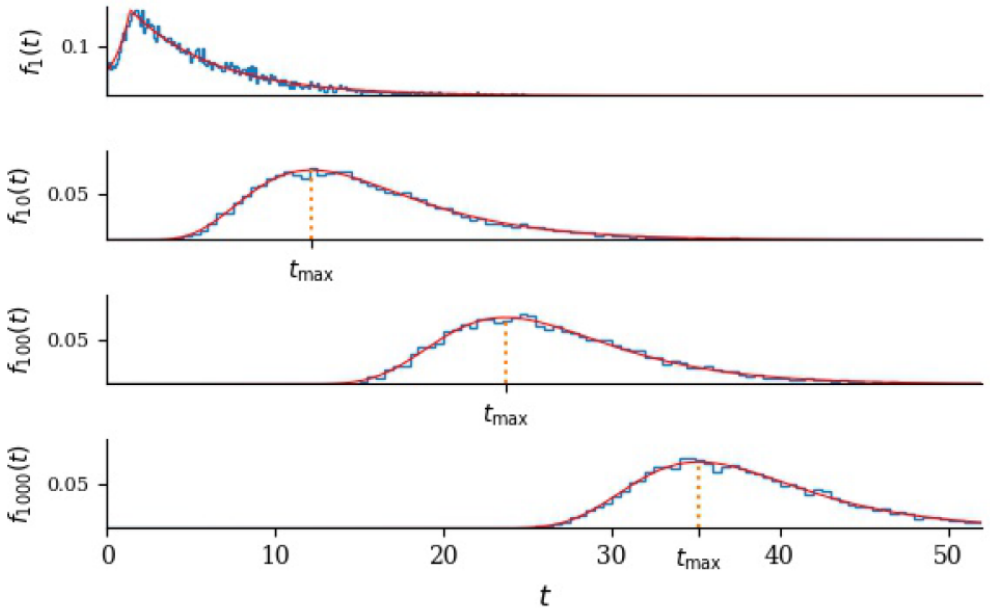


Figure 4.6: Probability density of extinction times, with $n = 1$, 10, 100 and 1000. The solid red lines are the exact results and the blue histograms are compiled from 10000 numerical realisations. The same horizontal scale is used in each case, with $\mu = 0.2$ and $\delta = 1$. Analytical solutions are shown in red, histograms from numerical simulation in blue. Top: $n = 1$. The maximum is at $t = \frac{1}{\delta}\log 4$, after which all cells are in the death pool. In each of the lower three panels, the vertical dotted line is $t_{\max} = \frac{1}{\mu}\log(nA)$. The ratio μ/δ determines the prefactor A.

density function, with different choices of n. The maximum of the density is at $t_{\max} = \frac{1}{\mu}\log nA$. It is striking, in Figure 4.6, that increasing n while keeping μ constant shifts the distribution to the right, maintaining its shape. To understand this, consider (4.18) when t is close to $t_{\max}$ and n is large so that $A\mathrm{e}^{-\mu t} \ll 1$ and $\log(1-A\mathrm{e}^{-\mu t})^n = n\log(1-A\mathrm{e}^{-\mu t}) \simeq -nA\mathrm{e}^{-\mu t}$. Since $\log(1) = 0$ and we take the exponential, Equation (4.18) simplifies to

$$\mathbb{P}\left(\tau_n < t\right) = \exp(-nA\mathrm{e}^{-\mu t}) \qquad \text{as } n \to \infty. \tag{4.20}$$

Now we define

$$T = \mu(t - t_{max}) = \mu\left(t - \frac{1}{\mu}\log(nA)\right) = \mu t - \log(nA)$$

$$\log(nA) = \mu t - T$$

$$nA = e^{\mu t - T}. \tag{4.21}$$

Inserting Equation (4.21) into Equation (4.20) gives us

$$\mathbb{P}\left(\tau_n < t\right) = \exp(-nA\mathrm{e}^{-\mu t}) = \exp(-e^{\mu t - T}\mathrm{e}^{-\mu t}) = \exp(-\mathrm{e}^{-T}).$$

The derivation gives us

$$f_n(t) = \mu \exp\left(-T - \mathrm{e}^{-T}\right) \qquad \text{as} \qquad n \to \infty \tag{4.22}$$

which is the probability density function of the Gumbel distribution. We can use this result in various ways. Firstly, it means that the random variable $T_n = \mu(\tau_n - t_{\max})$ has the Gumbel distribution (Gumbel 1935) as $n \to \infty$.

$$\begin{aligned}
T_n &= \mu(\tau_n - t_{\max}) \\
\frac{T_n}{\mu} &= \tau_n - t_{\max} \\
\tau_n &= t_{\max} + \frac{T_n}{\mu} = \frac{\log(nA)}{\mu} + \frac{T_n}{\mu} \qquad \text{we define} \qquad T_n = -\log(-\log U) \\
\tau_n &= t_{max} - \frac{\log(-\log U)}{\mu}
\end{aligned}$$

The form (4.20) is convenient for generating random variables that have the same distribution as that of the extinction time: if $\mathbf{U}$ is uniformly distributed in $(0, 1)$ then the random variable $\left(-\log(-\log \mathbf{U}) + \log(nA)\right)/\mu$ is a sample from the distribution. Rearranging (4.20), we find the following.

$$\begin{aligned}
\text{If} \quad \mathbb{P}(\tau_n < t) &= p \\
p &= e^{-nAe^{-\mu t}} \\
\frac{1}{p} &= e^{nAe^{-\mu t}} \\
log\frac{1}{p} &= nAe^{-\mu t} \\
log\left(log\frac{1}{p}\right) &= log(nA) - \mu t \\
\frac{1}{\mu}log\left(log\frac{1}{p}\right) + t &= \frac{log(nA)}{\mu} \\
\frac{1}{\mu}log\left(log\frac{1}{p}\right) + t &= t_{max} \\
\text{then} \qquad t &= t_{\max} - \frac{1}{\mu}\log(\log(1/p)).
\end{aligned} \tag{4.23}$$

In Figure 4.7 we display the cumulative distribution of the extinction time τ_{1000}. The Figure also indicates t_1, t_{50} and t_{99}, defined as the values of t such that $\mathbb{P}(\tau_n < t)$ is equal to 0.01, 0.50 and 0.99, calculated using (4.23). The constant A is calculated using the same parameter values as in Figure 4.6.

	formula	number
single-cell time to reach the death pool t_k	$\frac{1}{\delta}\log(4k)$	(4.3)
single-cell survival	$\mathrm{e}^{-\mu(t-t_k)}$	(4.7)
fraction of single cells surviving	$A\mathrm{e}^{-\mu t}$	(4.9)
drug effect A	$\frac{4^{\mu/\delta}+\frac{1}{4}\frac{\mu}{\delta}}{1+\frac{\mu}{\delta}}$	(4.10)
single-cell mean time to extinction	$\frac{1}{\delta}(\log 4-\frac{3}{4})+\frac{1}{\mu}$	(4.12)
single-cell variance time to extinction	$\frac{1}{\mu^2}$	
fraction of population surviving	$(A\mathrm{e}^{-\mu t})^n$	
population mean extinction time	$\frac{1}{\mu}(\log nA+\gamma)$	(4.17)
population variance extinction time	$\frac{1}{\mu^2}\frac{\pi^2}{6}$	

Table 4.1: Main formulas for sustained single dose. The time t_k is given by equation (4.3), A is a function of μ/δ given by equation (4.10) and $\gamma = 0.577\ldots$

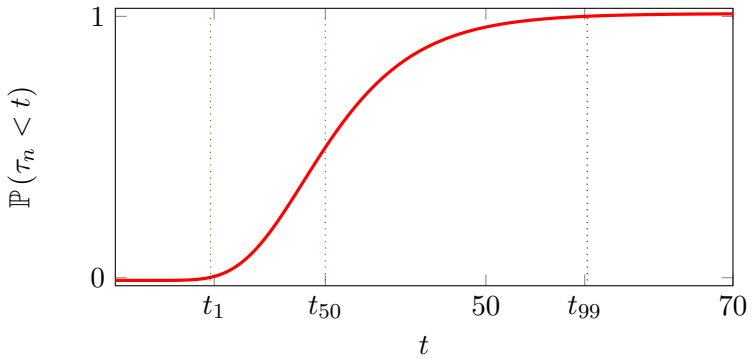


Figure 4.7: The cumulative density function of the extinction time of a cohort of $n = 1000$ cells. The meaning of the notation on the x axis is: the probability of the extinction time less than t_{100p} is p. Using equation (4.23), $t_1 = t_{\max} - 1.50/\mu$, $t_{50} = t_{\max} + 0.37/\mu$ and $t_{99} = t_{\max} + 4.6/\mu$, where $t_{\max} = \frac{1}{\mu} \log nA$.

4.3 Multiple-dose treatment with cell division

In Section 4.3, we consider a model where cells with sufficiently large regulator values may divide. Under the influence of repeated drug doses and recovery periods, whether extinction is the ultimate fate of the population depends on the balance between cell death and division. That balance, in turn, depends on the distribution of regulator values in the surviving cell population, which becomes biased towards larger values.

In multiple-dose treatment, the drug is given in doses of duration T_d each, followed by a recovery period of duration $T - T_d$. Thus, one cycle takes time T. In the example of Figure 4.9, the recovery period is twice the dose duration.

Let examine how each cell's regulator value changes. While a dose is being administered, the drug's effect is similar to that described by (4.2):

$$k_i(t) = k_0 \mathrm{e}^{-\delta t} \qquad 0 \le t \le T_d.$$

During recovery periods, on the other hand, a cell's regulator value increases towards its natural value k_0 with rate α. In the first recovery period

$$k_1(t) = k_0 - (k_0 - k_1(T_d))\mathrm{e}^{-\alpha(t-T_d)} \qquad T_d < t < T.$$

The term $(k_0 - k_1(T_d))$ describes the decrease of the k value after the treatment period. $\mathrm{e}^{-\alpha(t-T_d)}$ decreases as times becomes larger, hence the whole term $(k_0 - k_1(T_d))\mathrm{e}^{-\alpha(t-T_d)}$ becomes close to zero and the cell regains its initial value.

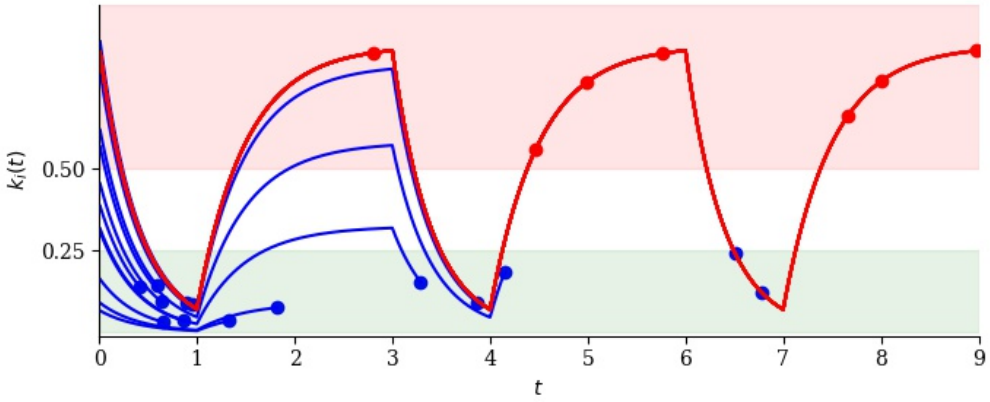


Figure 4.8: Illustrating the effect of multiple cycles of drug dose and recovery on a small cell population. Blue lines represent cells that die before the end of the third cycle and read lines represent cells that survive to the end of the third cycle. Blue dots indicate the death of a cell, which happens with rate μ to cells with regulator values smaller than 0.25 (green shading). Red dots indicate cell division, which happens with rate λ to cells with regulator values greater than 0.5 (red shading). In the initial cell population, regulator values are uniformly distributed between 0 and 1. After three rounds of dose and recovery, all remaining cells are descended from an initial cell with initial regulator value close to 1. The parameter values are $\mu = 1$, $\lambda = 0.4$, $\delta = 2.5$, $\alpha = 2$, $T = 3$ and $T_d = 1$.

In general, during the nth dose,

$$k_i(t) = k_n \mathrm{e}^{-\delta(t-nT)} \qquad nT < t < nT + T_d.$$

In the nth recovery period

$$k_i(t) = k_0 - (k_0 - k_i(nT + T_d))\mathrm{e}^{-\alpha(t-nT-T_d)} \qquad nT + T_d < t < (n+1)T.$$

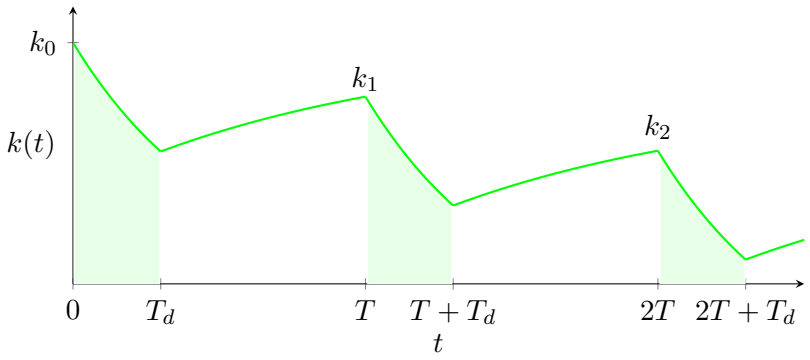


Figure 4.9: The regulator value of an individual cell under multiple-dose treatment. Drug doses are administered for time T_d (indicated by green shading) followed by a recovery period. The total cycle time is T. The nth peak value is given by (4.25).

As shown in Figure 4.9, we denote the nth peak regulator value by k_n. That is,

$k_n = k_i(nT)$. Then

$$k_1 = k_0(1 - \mathrm{e}^{-\alpha(T-T_d)}) + k_0\mathrm{e}^{-\delta T_d}\mathrm{e}^{-\alpha(T-T_d)}.$$

Given $n \geq 1$, we find that k_{n+1} depends on k_n and k_0 as follows:

$$k_{n+1} = Ak_0 + Bk_n, \tag{4.24}$$

where $A = 1 - \mathrm{e}^{-\alpha(T-T_d)}$ and $B = \mathrm{e}^{-\delta T_d}\mathrm{e}^{-\alpha(T-T_d)}$. We solve this difference equation by repetitive substitution. The solution of (4.24) is

$$k_n = K + (k_0 - K)B^n \qquad \text{where } K = \frac{A}{1-B}k_0. \tag{4.25}$$

As $n \to \infty$, $k_n \to K$. We observe that the asymptotic peak value, K, is a function of T and T_d, δ and α, that depends on the dosing duration and effectiveness, and on the extent of recovery after each dose.

Now we consider the effect of multiple doses on the size of the cell population. Recall that cells with $k_i(t) < 0.5$ divide with rate λ and cells with $k_i(t) < 0.25$ dies with rate μ. Thus, regulator values increase, and the number of cells may increase, in the recovery intervals between doses. In the examples shown in Figure 4.10, $T = 3T_d$.

If the drug-induced death rate μ is sufficiently large, the increase in cell numbers during recovery periods is not sufficient to make up, on average, for the loss of cells during the preceding doses. Then, all cells will eventually be killed. However, the number of doses for complete extinction varies from realisation to realisation, as illustrated in the lower part of Figure 4.10. In the Figure, extinction of the cell population (indicated by a red dot) occurs before the end of the tenth dose in two of the realisations shown. On the other hand, if μ is smaller than a threshold, the population of cells will increase in the long run because more cells divide during the recovery periods that are killed during the doses. This is the situation illustrated in the upper part of Figure 4.10. In this parameter range, it is possible to observe temporary decrease of average regulator values, and even extinction of the population, in some realisations.

Can we calculate the threshold value of μ that defines ultimate extinction? We begin by noting that an individual cell's fate depends on its initial regulator value. Firstly,

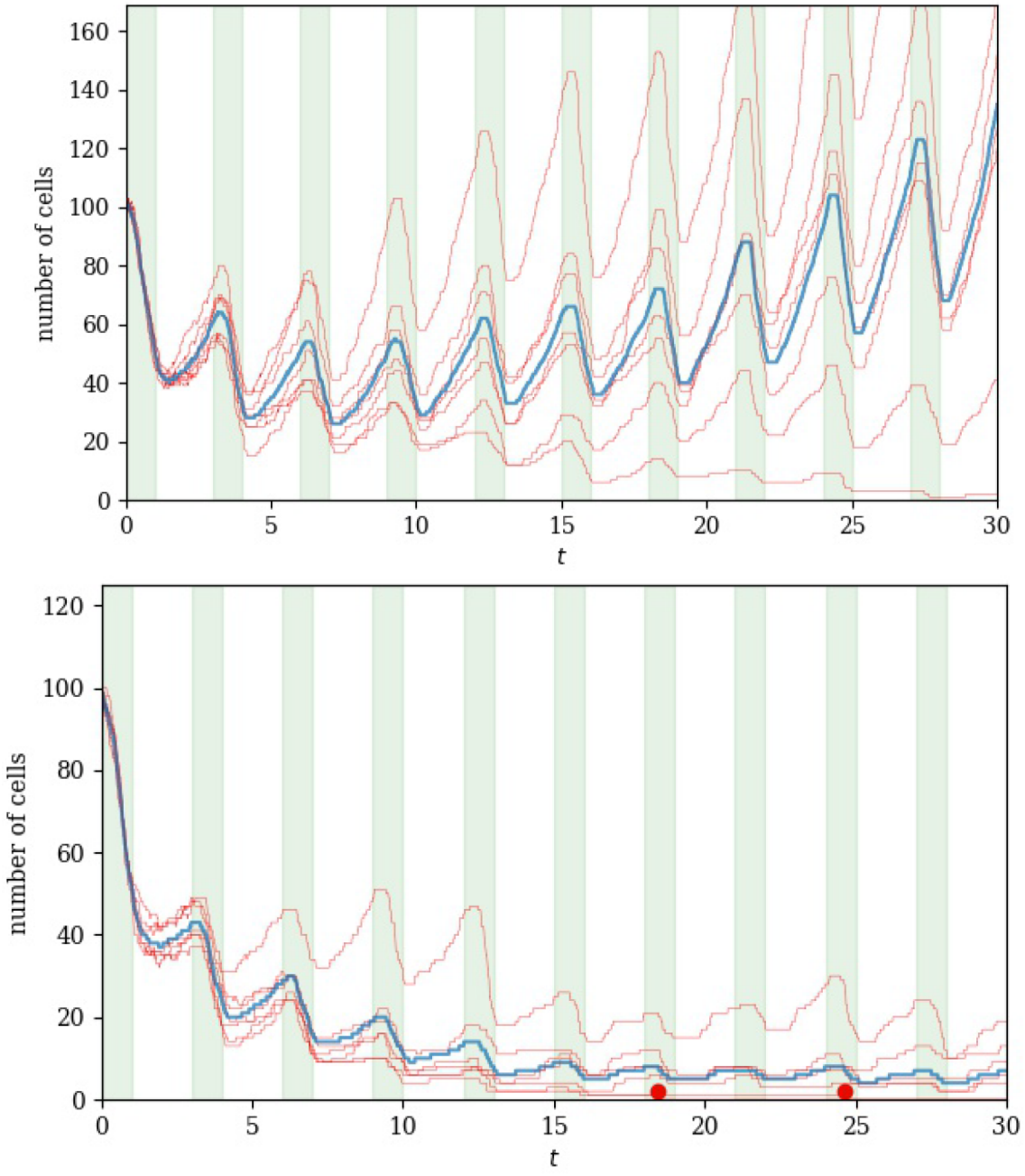


Figure 4.10: Number of cells as a function of time. Red lines are individual realisations and the ensemble average is shown in blue. The intervals when the drug is applied are shaded green. Red dots indicate extinction times. $\mu = 1$ and $\delta = 2.5$, $T = 3$ and $T_d = 1$. Top: $\lambda = 0.4$. Bottom: $\lambda = 0.25$. The initial population of 100 cells has regulator values chosen uniformly in $[0, 1]$.

cells with higher $k_i(0)$ spend less time in the death pool. Secondly, cells 'remember' their initial regulator value in the sense that $k_i(t)$ relaxes towards $k_i(0)$ in each recovery phase. A dramatic example of this is seen in Figure 4.8 where, after three cycles, all surviving cells are descended from the initial cell with highest regulator value. We provide an estimate of the threshold in the next section.

symbol	dimensions	interpretation
μ	T^{-1}	cell death
λ	T^{-1}	cell division
δ	T^{-1}	drug action
α	T^{-1}	relaxation
T	T	cycle time
T_d	T	dose time

Table 4.2: Parameters for multiple-dose treatment.

The distribution of regulator values in the population of cells before the first dose is uniform in $(0, 1)$. Each cycle of dose and recovery favours cells with larger regulator values (that spend less time in the death pool and more time in the division pool). This selection effect is an adaptation of the population akin to the development of drug resistance (even though it remains true that the drug can, given enough time, kill all cells). Indeed, we may observe in the lower figure of Figure 4.10 that the first few drug doses do reduce the cell population significantly, but the cell population that survives is able to recover.

4.4 Estimate of critical value

Given λ, T and T_d, what is the minimum value of μ, noted here as μ_c, necessary to guarantee eventual extinction? Figure 4.11 summarises numerical results at different values of μ. Blue dots represent the mean number of cells still alive after 100 cycles of dose and recovery. If μ is sufficently large then the mean number of cells is small (and all cells are eliminated in some realisations). At smaller values of μ, the population grows in the long run. In a deterministic model, there is a sharp transition between behaviours in the late-time limit. In a stochastic model, the probability of extinction is non-zero even when

μ is below the threshold. In both cases, the greater the number of doses, the sharper the transition appears.

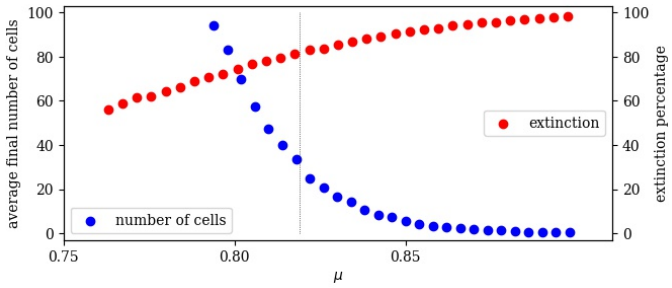


Figure 4.11: Each dot represents the mean number of cells still alive after 100 doses. The dashed vertical lines are the approximation (4.26). Each dot represents the average over 10000 realisations. The initial number of cells is 100 and the parameter values are $\lambda = 0.25$, $\alpha = 2$, $\delta = 2.5$, $T = 3$ and $T_d = 1$.

With these caveats, it is useful to calculate an approximation for the critical value of μ and examine its parameter dependence. We may estimate the critical value from the relation

$$\mu \times \text{mean time spent in death pool} \quad = \quad \lambda \times \text{mean time spent in division pool.}$$

Applying this relation is not straightforward because it requires knowledge of the distribution of regulator values after many cycles of dose and recovery. We obtain an estimate of μ_c by considering the times T_1, T_2, T_3 and T_4 that characterise entry to and exit from the death and divison pools, as shown in Figure 4.12. Using the approximations given in Figure 4.12, we estimate that the critical value, μ_c, satisfies the condition

$$\mu_c(T_3 - T_2) = \lambda(T_1 + T - T_4) \tag{4.26}$$

The time durations for the dosing process can be estimated as follows: T_1 is the time from treatment start until the cell has reached a k value of 0.5. T_2 is the time when the k value is 0.5. T_3 is the time when the cell leaves the death pool and its k value is 0.25. T_4 is the time the recovering cell reaches the k value of 0.5.

$$\mu_c = \frac{\lambda(T_1 + T - T_4)}{T_3 - T_4} + \frac{\lambda(1/\delta log2 - T_d - 1/\delta log2 + T)}{T_d + 1/\alpha log(4/3) - 1/\delta log6}$$

We note, firstly, that μ_c is proportional to λ, the division rate of cells when they are in the division pool. Next, we observe that μ_c is a decreasing function of δ and of T_d/T. That

is, faster-acting drugs, and drugs that are applied for a longer fraction of the total cycle time, kill cancer cells more rapidly. The dependence on the parameter α that describes the rate of relaxation of regulator values in the recovery periods, is more complicated because speed of recovery affects times spent in both death and division pools.

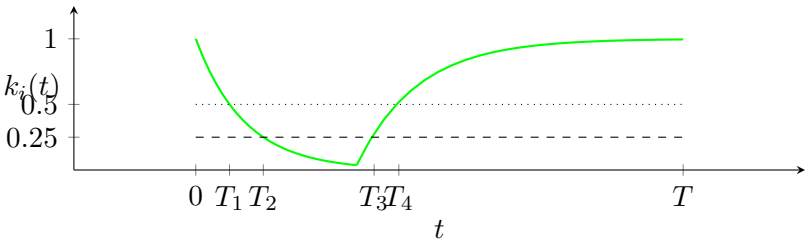


Figure 4.12: Constructing the approximation. Dotted line: $k = 0.5$. Dashed line: $k = 0.25$. We use the following approximations: $T_1 = \frac{1}{\delta}\log 2$, $T_2 = \frac{1}{\delta}\log 4$, $T_3 = T_d + \frac{1}{\alpha}\log\frac{4}{3}$, $T_4 = T_d + \frac{1}{\alpha}\log 2$.

4.5 Discussion

The effect of a drug on tumour cells can be described using ODEs or agent based models. While ODEs often have the advantage of easy implementation and analysis, they do not naturally capture stochasticity or heterogeneity. Agent based models overcome those shortcomings in computational models where agents and their interactions are governed by stochastic rules, perhaps requiring high computational power and running times (Truong et al. 2022). Here, based on a published agent based model (Truong et al. 2022) where a heterogeneous cancer cell population is treated with a MEK inhibitor decreasing phosphorylated ERK and causing cell death, we use stochastic modelling and analysis as a bridge between different types of models. We summarise the dynamics in simple mathematical expressions. A summary of the formulas can be found in Table 4.1. With the advance of molecular biology and the development of therapies that target intracellular signalling pathways (Felten et al. 2023; St Clair 2009; Tsimberidou 2015), it becomes more important to consider heterogeneity of target cells (Dewachter et al. 2019).

Our model allows us to describe the fate of a single cell under a continuous drug treatment and can be extended to a heterogeneous population of cells. Heterogeneity originates in the initial conditions: each cell's starting k value is chosen randomly between 0 and 1. The effect of the drug is to decrease each cell's k with timescale $1/\delta$. As the population changes in size and distribution, cells with $k < 0.25$ are in the death pool, with death rate μ. Cells with $k > 0.5$ are in the division pool, with division rate λ. The threshold of the division and death can is adjustable according to the pathway of interest. $1/\delta$ describes the potency of the drug and can be measured by the activity of the targeted protein. For example to detect phosphorylated ERK, cancer cell culture were grown with and without drug treatment and immunoblots of those cultures were created incubated with antibodies to activated phosphorylated ERK1 and ERK2 and total ERK1 and ERK2. Antibodies are quantifiable with the appropriate anti-mouse or anti-rabbit horseradish peroxidase–conjugated secondary antibody by enhanced chemiluminescence (Yeh et al. 2009). The death and division rate can be obtained from tumour cell cultures under drug treatment in different concentration and without. For example, Yang et al (J. Yang et al. 2021) used time-resolved microscopy to track the temporal change of the number of live and dead tumor cells *in vitro*. Another option are xenograph models where tumour cells are injected subcutaneoulys or in the same organ as the tumour's origin .

The change in tumour size can be measured by surgically removing the tumour for *ex vivo* weighing, *in vivo* tumor volume measurement using calipers, within internal organs by employing magnetic resonance imaging, computed tomography, or ultrasound (Bonato et al. 2024). Additionally, organoids can be employed to measure longitudinal changes in tumour size (Bonato et al. 2024).

The survival probability of each cell and the whole population can be calculated to obtain the time of population extinction. We identified a relation between the population size and time to extinction. This helps to predict the required treatment duration for a given cell population and their size. The cumulative density of extinction times was obtained which indicates how long to treat until probability of extinction reaches certain number. In addition, we simulated a multiple dose treatment where the cells are allowed to recover between the cycles and divide. We found a formula describing how the regulator value changes over time depending on treatment duration and the drug potency δ. This provides a tool to predict how adjusting the treatment duration and chosing drugs with different potencies influences treatment success. Using the method, characteristics of the population such as a critical division rate for uncontrolled growth or a successful treatment can be identified. This gives insight into what death rate is needed to drive a population into extinction depending on the birth rate of that population and treatment duration. Given a tumour cell population with a certain birth rate, adjustments of the death rate or treatment duration can be done for a successful treatment.

Stochastic PKPD models are important to describe small populations (for example, the small residual cancer cell population after effective immunotherapy, or the small initial population early in infection). Because the dynamics in every individual is partly driven by random fluctuations, the same set of parameter values and initial conditions leads to multiple different outcomes. Change over time is dependent on a series of consecutive probabilistic events. Stochastic models are also applicable *in vitro* because drug effects are complex and probabilistic at the single-cell level. Even genetically identical cells in a uniform environment differ in their response to drugs due to stochasticity in gene expression levels and other biochemical phenomena (R. Wang et al. 2017). Mutation-based categorization incompletely recapitulates the complexity and diversity of cancer subtypes, and sub-classification of tumours based on gene expression profiles potentially improves clinical decisions (Melo et al. 2013).

Chapter 5

A hybrid PKPD agent based model of the tumour immune interaction and effects of anti cancer combination therapy

5.1 Introduction

The generation of a potent anti tumour immune reaction is a multiple step process. First, antigen presenting cells, *eg.,* dendritic cells, capture tumour antigens and process them with the major histocompatibility complex (MHC) class I and II pathway. The resulting peptide epitopes are displayed on the cell surface to stimulate $CD8^+$ and $CD4^+$ T cells. Usually this process, called priming, happens in the regional lymph node. Then, tumour specific T cells differentiate into effector cells by receiving a combination of signals from the T cell receptor which establishes specificity and several costimulatory molecules. The T cell receptor and associated CD4 or CD8 molecules bind MHC-peptide complexes presented on dentritic cells. Transmembrane proteins of the B7 and tumour necrosis factor receptor (TNFR) families, and also receptors for some cytokines, such as interleukin-12 (IL-12), receive costimulatory signals. This includes the costimulatory molecule CD28, which binds B7-1 (also known as CD80) and B7-2 (also known as CD86) on dentritic cells, and activating receptors like 4-1BB (also known as CD137 and TNFRSF9), OX40 (also known as TNFRSF4), and glucocorticoid induced TNFR related protein (GITR;

also known as TNFRSF18) (Vanneman and Dranoff 2012). After activation, cytotoxic T cells can use two mechanisms for killing the targeted cell: granule exocytosis and death ligands. The granule exocytosis pathway is initiated by secreting perforin which makes the tumour plasma membrane permeable for granzymes. Once entered, the cancer cell granzymes cleave critical intracellular substrates controlling cell death and survival. Death ligands are proteins expressed by cytotoxic lymphocytes on the cell surface or secreted to exosomes and trigger target cell death via death receptors (Martínez-Lostao et al. 2015).

To maintain an active immune response, T cells must avoid negative regulatory signals that trigger their inactivation or induce tolerance programs such as anergy or exhaustion. Major negative costimulatory receptors expressed on the T cell surface are CTLA4 and programmed death 1 (PD1). Immunosuppressive networks within the tumour micro environment consist of soluble factors and regulatory cell populations. Myeloid-derived suppressor cells (MDSCs) expand during cancer, inflammation and infection, and have an ability to suppress T-cell responses by using soluble factors such as IL-10 and TGFβ and induction of regulatory T cells (Treg) (Gabrilovich and Nagaraj 2009). Tumour-associated macrophages (TAMs) cause immunosuppression by producing cytokines, chemokines, growth factors, and triggering the inhibitory immune checkpoint proteins release in T cells (Lin et al. 2019). One important cell type for regulating and suppressing the immune system are regulatory T (Treg) cells, a sub-population of T cells. They are crucial for peripheral tolerance, preventing autoimmunity, and limiting chronic inflammatory diseases. However, they can also suppress desired immune reaction i.e. anti tumour immunity. Regulatory T cells use multiple mechanisms for suppressive effects: suppression by inhibitory cytokines, suppression by cytolysis, suppression by metabolic disruption and suppression by modulation of dendritic-cell (DC) maturation or function. Treg cells can secrete inhibitory cytokines such as interleukin-10 (IL-10), transforming growth factor-β (TGF β), and IL-35. Another mechanism is to mediate cytolysis via granzyme A and/or granzyme B and perforin. Targeted cells can also be killed or suppressed by metabolic disruption. Rapid consumption of IL-2 by CD25+ Treg cells lead to cytokine-deprivation-mediated apoptosis. In addition regulatory T cells can generate adenosine pericellular and transfer cyclic AMP intracellularly through membrane gap junctions which expose the target cell to two potently inhibitory molecules. Suppression can also happen via modulation of DC maturation or function. Interaction with the cytotoxic T-lymphocyte antigen 4 (CTLA4)–CD80/CD86 induce the release of a potent regulatory molecule, indoleamine 2,3-dioxygenase (IDO).

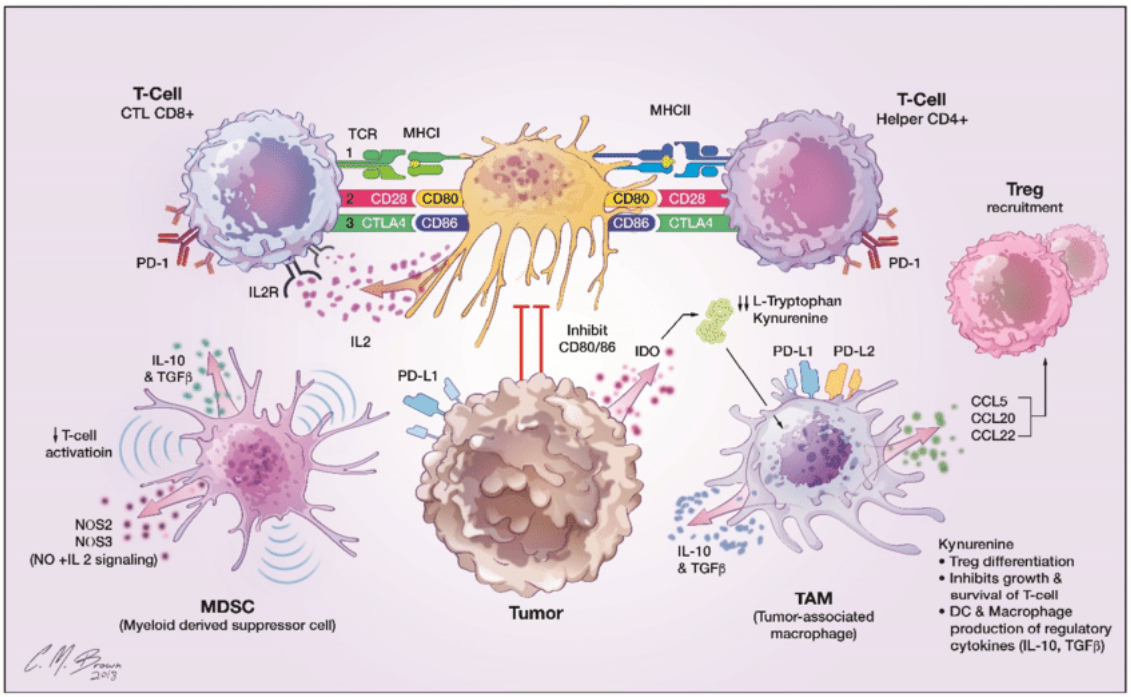


Figure 5.1: Immune response against tumour cells (Figure taken from (Leibold et al. 2019) with permission of the author).

This causes the catabolism of tryptophan into pro-apoptotic metabolites. DC maturation and function is inhibited by lymphocyte-activation gene 3 (LAG3) binding to MHC class II molecules. The modulation of dendritic cell co-stimulation also reduces the $CD8^+$ T cell proliferation (Vignali et al. 2008; Dowling et al. 2018).

Tumour survival and escape depend on various factors, including the immune response, concentration gradients (for example, those of oxygen and drug in the tumour micro environment, TME), and on tumour cell characteristics (such as the location, cell-cycle phase and PDL1 expression status). In healthy tissue, the PD1 receptor and its ligand PDL1 prevent excessive immune activity and serve as an immune checkpoint pathway to maintain 'self' tolerance. Under prolonged immune stress, PDL1 expression can be induced on cancer cells and other cells in the TME, producing an immunosuppressive environment (Han et al. 2020). Cell death due to radiotherapy can be induced by a combination of direct radiation-mediated cytotoxicity and the stimulation of an anti tumour immune response. Immune effector cells and suppressor cells can be subject to radiation-mediated cytotoxicity but more can be attracted by radiotherapy-induced cancer cell death (Alfonso et al. 2021). In addition, radiotherapy-induced cancer cell death can cause PDL1 overexpression (Sato et al. 2019). Furthermore, the efficacy of radiotherapy depends on the oxygenation status of the tumour cells as hypoxic cells are less

radiosensitive. Cell-cycle regulation also plays an important role. For example, the G2-M phase is more sensitive to radiation than the G1 phase. Radiation can activate cell repair mechanisms such as the intracellular p53 and p21 pathways which results in cell cycle delay, accumulation in the G1 or G2 phase, and synchronisation (Powathil et al. 2013). DNA damage response inhibitors counteract these repair mechanisms by inhibiting ataxia telangiectasia mutated and rad3-related kinase (Checkley et al. 2015).

Similarly to radiotherapy, the cell cycle plays a critical role in chemotherapy efficacy. For example, the drugs paclitaxel and docetaxel cause cells to accumulate at the G2-M phase, while flavopiridol treatment results in G1 and G2 phase accumulation (Powathil et al. 2013; Nehme et al. 2001). Additionally, chemotherapy causes higher immunogenicity by increasing the potential for cancer cells debris to be recognised by the immune system but can also lead to an immunosuppressive TME by overexpression of PDL1 on cancer cells and immune cell death (Fabian et al. 2021; Ng et al. 2018). PDL1 or PD1 antibodies such as pembrolizumab block the interaction between PD1 and its ligands, PDL1 and PDL2. This blockade prevents effector cell exhaustion and facilitates tumuor killing and immune-mediated rejection (Lindauer et al. 2017). Considering the characteristics of each treatment option, synergism could be identified. For example, radiation-induced cell cycle delay can help various cell cycle phase-specific drugs to induce a higher cell kill, DNA damaging treatments such as radiotherapy or chemotherapy could be combined with DNA damage response inhibitors, and administration of PD1 antibodies can improve the immune response after radiation or chemotherapy treatment.

Therefore, with the increasing complexity of mono and combination therapies, it is critical to understand those interactions, the heterogeneity of the tumour, the emergent behaviour in the TME, and the therapeutic effect of drug dose and schedule on a given tumour. Pharmacokinetic and pharmacodynamic models are used to predict dose and scheduling. However, those ordinary differential equation models often require many states and an even larger number of parameters to capture the complex behaviour of many cell types. Additionally, spatial temporal dynamics, heterogeneity, emergent behaviour are often not incorporated into standard ODE PK-PD modelling approaches (Truong et al. 2022). Therefore, in this chapter, we developed a multiscale hybrid agent based PDE ODE model that incorporates tumour immune cell interaction which can be easily extended with different treatment modules such as radiotherapy, PD1 antibody, chemotherapy and DNA damage response inhibitor treatment.

5.2 Model structure

The hybrid multiscale immuno-oncology model consists of an agent based modeldescribing the interaction between cancer cells and immune cells which is extended with ODEs and PDEs that simulate the oxygen distribution in the TME and different treatment modalities such as PD1 antibody treatment, radiotherapy, chemotherapy, and DNA damage response inhibitor.

5.2.1 Agent based model

The ABM consists of cancer cells, immune effector and suppressor cells as agents. The behaviour of each agent and interaction with each other and the environment is governed by a set of rules and attributes. Each cell has its location on a square 3D grid with (x,y,z) coordinates as attributes (see Figure 5.2). In order to avoid long running times, the number of cells is scaled, so that one cell in the simulation refers to 10^6 cells in reality. We assume that the tumour doesn't include extracellular structures such as fibroblast, vessels, or other cells, and that the tumour mass is densely packed without void space. Therefore, we can consider that a 0.1 cm^3 tumour would be 10^8 cells in reality or 100 cells in the simulation (Del Monte 2009). A 1 cm^3 tumour would be 1000 cells in the simulation and a 100 cm^3 tumour would be 10 000 cells in the simulation. The minimum detectable size in clinical conditions depends on various factors such as the tumour to background ratio, location of the tumour, imaging isotope and can range from 0.2-2 cm^3 (Erdi 2012). The lethal size is defined as a 1000 cm^3 tumour which would be 10^6 cells in the simulation (Norton 1988; Brunton and Wheldon 1980). The immune cell cancer cell proportion in the TME is 1:100 (Gong, Milberg et al. 2017).

Cancer cells

Attributes of cancer cells are the PDL1 expression status, cell cycle phase, and time to division. They can consume oxygen, mutate, divide, become quiescent, necrotic, and die due to natural death or due to therapy. A cancer cell ages and goes through the cell cycle phases. After birth a cell is in the G1 phase for 11h, then it passes through the phases for a certain time span: S phase for 8h, G2 phase for 4h, and M phase for 1h. Once a cancer cell is in the M phase, it is ready for division. After division, the cell cycle starts again (Cooper 2000). If the oxygen level is below a threshold or if there is no space next to a cancer cell for division, it goes into the quiescent/resting phase (G0 phase) until the environmental

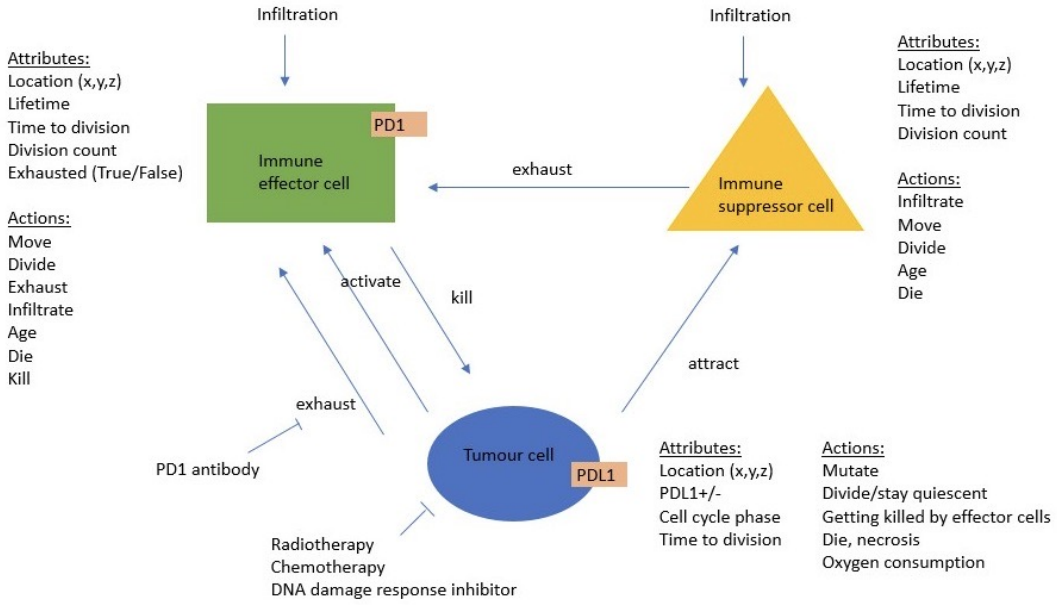


Figure 5.2: Schematic representation of the ABM and the attributes and actions of each cell type involved.

condition has changed. Oxygen levels can be increased and space can become available due to cells being eliminated by immune cell killing, radiotherapy, chemotherapy, DNA damage response inhibitor treatment, necrosis due to low oxygen levels, or natural death. When division happens, a cell which is ready for division will be randomly chosen, and the daughter cell will be randomly placed in one of the free positions next to the mother cell. In a square grid 3D environment there are 26 possible places as there a $3x3$ cube with the tumour cell in the middle. The PDL1 status of the mother cell will be inherited by the daughter cell. The simulation starts with PDL1-negative cancer cells, and in the course of the simulation, cancer cells can be randomly chosen to become PDL1-positive. The rates are calculated using the Gillespie algorithm. The oxygen levels in the environment are driving the cell behaviour. Cancer cells consume oxygen at a certain rate. Quiescent cells consume less oxygen than dividing cells.

Immune cells

To simplify the immune reaction, immune cells are summarised as effector and suppressor cells, that either act as cytotoxic cells which eliminates cancer cells or regulatory cells which suppresses the immune response. Attributes of immune cells are a certain lifetime (age), time to division, division count, and exhaustion status for effector cells. Actions include infiltration into the TME, moving, dividing, ageing, death, and for the effector cells

exhaustion and killing of cancer cells. In addition, immune suppressor cells can exhaust immune effector cells. Immune cells randomly infiltrate a free space in the environment. After infiltration, the division count, time to division, and age start at 0. While being in the TME, immune cells can move to one of 26 free possible locations next to them. Immune effector cells are attracted by cancer cells, and they are more likely to move to a location which is closer to a cancer cell. Immune suppressor cells have the goal to exhaust immune effector cells and are more likely to move closer to an effector cell position. Exhaustion happens with a certain rate determined by the Gillespie algorithm when an immune effector cell is next to a PDL1+ cancer cell or an immune suppressor cell. After exhaustion, the effector cell becomes inactive and is not able to move, kill, or divide. The exhausted cells will be cleared from the environment once the lifespan is reached. Effector cells can kill cancer cells when they are next to a cancer cell. The kill rate is determined by the Gillespie algorithm. Immune cells age with the time in the simulation. They die and get cleared from the system when their age is above the lifetime or due to cell death because of radiation or chemotherapy. Immune cell division happens at a certain rate when the time to division is reached and the division count is not exhausted. The daughter cell will be placed randomly in one of the free 26 possible locations next to the mother cell. The division count and time to division of the mother cell will be updated after the division event and the daughter cell inherits the division count, and starts the time to division, and lifespan at 0 (Gong, Milberg et al. 2017).

5.2.2 Gillespie algorithm and rates

The actions are chosen by the Gillespie algorithm (Truong et al. 2022) and the probabilities for an event depend on the reaction rates and cell numbers. Cancer cell division happens at a rate equal to the division rate per cell (called c_div_rate in the code below) times the number of existing cancer cells in the environment ($len\_all\_c\_cells$ in the code below), which makes division more likely the higher the cancer cells numbers are. A PDL1 expression event is more likely when more cancer cells are being eliminated by effector cells, chemotherapy or radiotherapy (see c_mutate in the code below). The cancer killing probability increases with the number of effector cells, while the natural cancer cell death rate depends on the number of cancer cells (see $c\_killed\_by\_eff\_cells$ and c_death in the code below). The immune effector infiltration rate increases with the number of cancer cells, immune effector cells in the TME and eliminated cancer cells by the immune reaction

(see *eff_infiltrate* in the code below) while the suppressor infiltration rate depends on the number of tumour cells that have been killed by effector cells (see *sup_infiltrate* in the code below). Immune cell division and moving rates are higher with a larger number of immune cells (see *eff_divide* in the code below). The effector exhaustion rate depends on the number of suppressor cells, PDL1+ cancer cells and the receptor occupancy with the antibody (see *eff_exhaust* in the code below). Those rates and their dependencies can be changed to account for different tumour types. More details about the parameter values can be found in the table A.1 in the appendix.

```
#cancer cell
    c_divide=c_div_rate*int(len_all_c_cells[-1])
    c_mutate=(l.kill_count+l.IR_count+l.chemo_count)*c_mutate_rate
    c_killed_by_eff_cells=c_killed_rate*int(len_eff_cells[-1)
    c_death=death_rate*int(len_all_c_cells[-1])
#suppressor cell
    sup_move=int(len_sup_cells[-1])*sup_move_rate
    sup_infiltrate=l.kill_count*sup_infil_rate
    sup_divide=int(len_sup_cells[-1])*sup_div
#effector cell
    eff_exhaust=(int(len_sup_cells[-1])+len_PDL1_pos[-1])
*eff_exhaust_rate*p.exhaust_possible*(100-R0_L[-1])/10000
    eff_move=int(len_eff_cells[-1])*eff_move_rate
    eff_infiltrate=int(len_all_c_cells[-1]+len_eff_cells[-1]
    +l.kill_count/2)*eff_infil_rate
    eff_divide=int(len_eff_cells[-1])*eff_div
```

Oxygen in the environment

The oxygen concentration in the TME is modelled with a PDE, where (x, y, z) are the spatial coordinates in a three dimensional environment, t is the temporal variable, D is the diffusion coefficient of oxygen, and U describes the oxygen level in the environment

$$\frac{\partial U}{\partial t} = D\left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}\right),$$

which can be written as

$$\partial_t U = D\nabla^2 U = D\left(\partial_{xx} + \partial_{yy} + \partial_{zz}\right) U.$$

This can be discretised using a forward Euler in time and a second-order central finite difference (Özişik et al. 2017), such that

$$\begin{aligned}\frac{U^{n+1}_{i,j,k} - U^n_{i,j,k}}{\Delta t} =& D\Big(\frac{U^n_{i-1,j,k} - 2U^n_{i,j,k} + U^n_{i+1,j,k}}{(\Delta x)^2} + \frac{U^n_{i,j-1,k} - 2U^n_{i,j,k} + U^n_{i,j+1,k}}{(\Delta y)^2} \\ &+ \frac{U^n_{i,j,k-1} - 2U^n_{i,j,k} + U^n_{i,j,k+1}}{(\Delta z)^2}\Big),\end{aligned}$$

which implies that

$$\begin{aligned}U^{n+1}_{i,j,k} =& U^n_{i,j,k} + D\Delta t\Big(\frac{U^n_{i-1,j,k} - 2U^n_{i,j,k} + U^n_{i+1,j,k}}{(\Delta x)^2} + \frac{U^n_{i,j-1,k} - 2U^n_{i,j,k} + U^n_{i,j+1,k}}{(\Delta y)^2} \\ &+ \frac{U^n_{i,j,k-1} - 2U^n_{i,j,k} + U^n_{i,j,k+1}}{(\Delta z)^2}\Big).\end{aligned}$$

.

Here, n denotes the step for the difference of the time, i denotes the step for the difference in the x coordinate, j denotes the step for the difference in the y coordinate, and k denotes the step for the difference in the z coordinate.

We are simulating the oxygen diffusion with the python code below.

```
def diffusion(u0,dt,D,dg):
    ''' one step of diffusion equation '''
    u = u0.copy()
    u[1:-1, 1:-1, 1:-1] = u0[1:-1, 1:-1, 1:-1] + D * dt * (
          (u0[2:, 1:-1, 1:-1] - 2*u0[1:-1, 1:-1, 1:-1]
          + u0[:-2, 1:-1, 1:-1])/(dg*dg)
          + (u0[1:-1, 2:, 1:-1] - 2*u0[1:-1, 1:-1, 1:-1]
          + u0[1:-1, :-2, 1:-1])/(dg*dg)
          + (u0[1:-1, 1:-1, 2:] - 2*u0[1:-1, 1:-1, 1:-1]
          + u0[1:-1, 1:-1, :-2])/(dg*dg)
```

```
        )
    return u
```

u0[1:-1, 1:-1, 1:-1] simulates the amount of oxygen in the environment which diffuses in the 3D grid depending on the diffusion rate D and time step dt. The 3D grid is simulated with the expression in the brackets after D*dt. Each of the terms simulate the diffusion along the x, y or z axis. The term (u0[2:, 1:-1, 1:-1] - 2*u0[1:-1, 1:-1, 1:-1] + u0[:-2, 1:-1, 1:-1])/(dg*dg) describes the diffusion along the x axis, (u0[1:-1, 2:, 1:-1] - 2*u0[1:-1, 1:-1, 1:-1] + u0[1:-1, :-2, 1:-1])/(dg*dg) describes the diffusion along the y axis, and (u0[1:-1, 1:-1, 2:] - 2*u0[1:-1, 1:-1, 1:-1] + u0[1:-1, 1:-1, :-2])/(dg*dg) describes the diffusion along the z axis.

Active dividing cancer cells consume oxygen with the consumption rate ko while the oxygen consumption of quiescent cells is less by a factor of qc. The parameters are assumed and can be changed to simulate different degrees of oxygenation in the tumour micro environment and oxygen consumption rates of the cancer cells. We set the initial condition of oxygen as 1 and measure the concentration of oxygen in the TME relative to its initial condition. Cancer cells go into the division, quiescent, and necrotic stage according to the available oxygen level in the TME. Table 5.1 shows the parameters for the oxygen diffusion and consumption model.

Parameter	Value
Diffusion coefficient D	1 $[\mu m^2/d]$
Division range	$100 - 50\%$
Quiescence range	$50 - 30\%$
Death range	$30 - 0\%$
Oxygen consumption rate for dividing cancer cells ko	0.001 $[\%/d]$
Factor for oxygen consumption of quiescent cells qc	0.01

Table 5.1: Parameters of the oxygen diffusion and consumption.

5.2.3 Treatment

PD1 antibody

The PD1 antibody treatment module is implemented with a PKPD model from the literature (Elassaiss-Schaap et al. 2017), where a system of ODEs is used to calculate the receptor occupancy of the PD1 antibody pembrolizumab on the PD1 effector cell receptor

which will determine the effector cell exhaustion rate.

The monoclonal PD1 antibody is administrated i.v. into the central (blood) compartment as the concentration $C_1(t)$, where it can be distributed to the periphere compartment as the concentration $C_2(t)$ and redistributed to the central compartment with rates K_{12}, K_{21}. The concentrations in this model are in the units $nmol/l$. Elimination from the central compartment can be linear with the rate K or as a nonlinear saturable rate with the Michaelis-Menten kinetic $\frac{V_{max}}{K_M+C_1}$. Inside the central compartment, the antibody can bind to the PD1 receptors in the blood (e.g on immune effector cells) denoted with $C_{PD1b}(t)$ which leads to the bound drug-receptor complex $PD1_b$. Binding and detaching rates are K_{onPD1} and K_{offPD1}. The complex can be degraded with rate K_{deg}. The tumour plasma flow transports the unbound antibody to the tumour compartment at the rate PLQ. The tumour compartment consists of the vasculature, the endosomal layer and the interstitium. The monoclonal antibody inside the vasculature ($C_{vs}(t)$) is transported from the tumour vasculature to the interstitium via the neonatal fragment crystallizable-receptor (FcRn) salvage pathway (Ryman and Meibohm 2017). The antibody will be engulfed and released by the endosomal cell through pinocytosis with rate $CLup$. The unbound drug ($C_{ub}(t)$) binds to the FcRn receptor with rate K_{onFcRn}, and dissociates with rate $K_{offFcRn}$. Unbound proteins or unbound antibodies will be degraded with rate K_{deg}, while the the FcRn receptor antibody complex will not be degraded by lysosomes and a fraction of the bound drug ($C_b(t)$) on the FcRn receptor will be released by the vesicle to either the vascular side, as $C_{vs}(t)$, with rate $FR * CLup$, or the interstitial side, as $C_{is}(t)$, with rate $(1 - FR) * CLup$. Pinocytosis can also transport the antibody from the interstitium ($C_{is}(t)$) with rate $CLup$ to the endosomal compartment. In addition, the lymph flow can transfer the drug from the tumour vasculature ($C_{vs}(t)$) to the tumour interstitium ($C_{is}(t)$) with rate $(1 - vref) * L$. The lymph flow can further eliminate the drug from the vascular compartment with rate L or from the interstitium with rate $(1 - vref) * L$. Inside the interstitial tumour compartment, the antibody concentration $C_{is}(t)$ binds to the PD1 receptor on immune effector cells ($CPD1_t$) with rate K_{onPD1} and dissociates K_{offPD1}. The bound drug-receptor complex $PD1_t$ can initiate hyperbolic feedback and cause a receptor upregulation ($MPD1_t(t)$) on the immune effector cell. The receptor occupancy of the drug in the tumour compartment is calculated with the relation of antibody-drug complex ($PD1_t(t)$) to the concentration of unoccupied receptors ($CPD1_t(t)$). The subscript t in $PD1_t$, $CPD1_t$, and $MPD1_t(t)$ denotes that those receptors are in the tumour

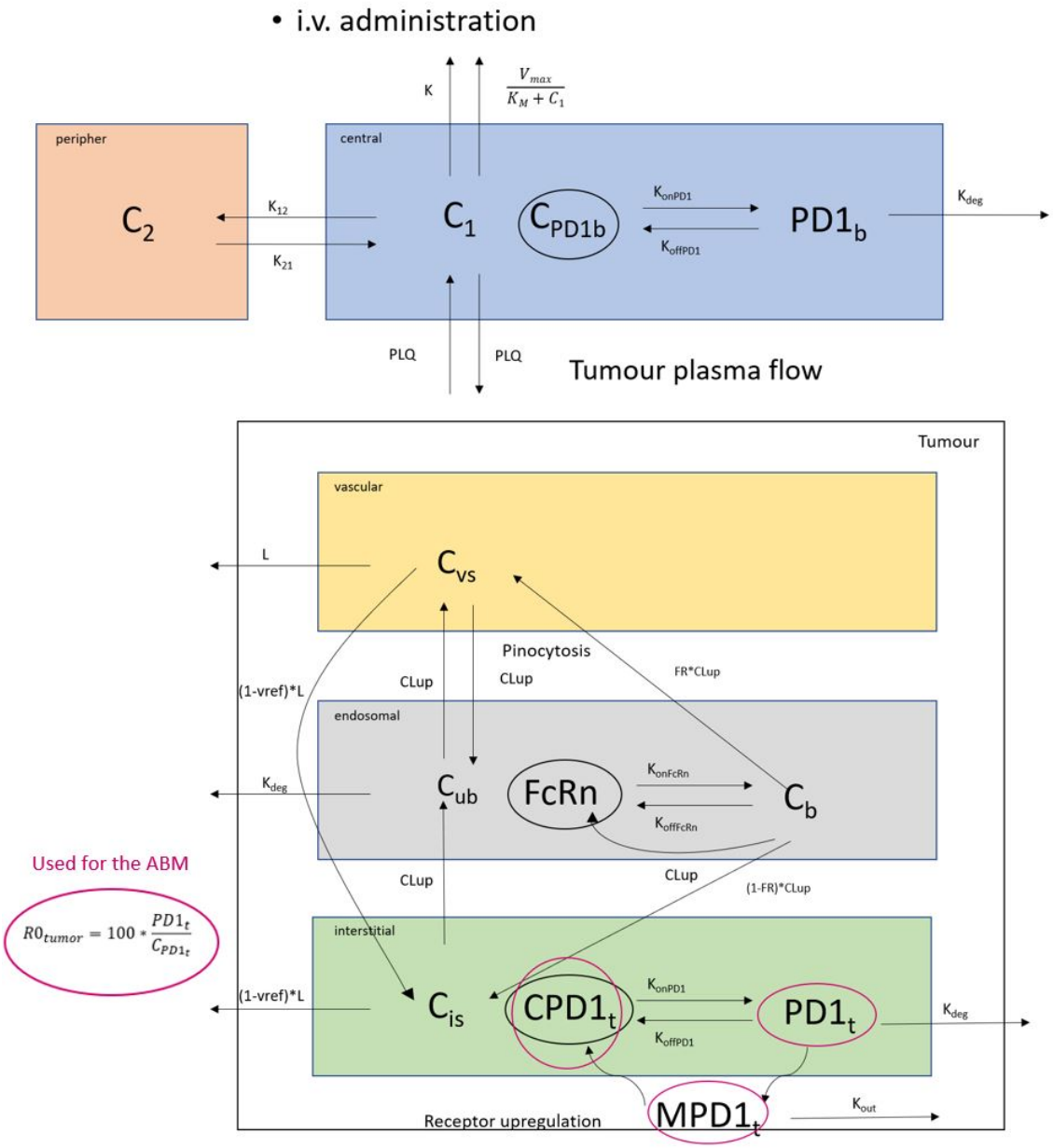


Figure 5.3: Schematic of the PKPD model for pembrolizumab

intestitium. Figure 5.3 shows the schematic of the PKPD model for pembrolizumab.

The model consist of the following ODEs:

- Central compartment:

$$V_1 \frac{dC_1}{dt} = -K(C_1 V_1) - \frac{V_{max}}{K_M + C_1} C_1 - PLQC_1 + PLQC_{vs} - K_{12} C_1 V_1 + K_{21} C_2 V_2 - K_{onPD1}(C_{PD1_b} - PD1_b) + K_{offPD1} PD1_b V_1,$$

- Peripheral compartment:

$$V_2 \frac{dC_2}{dt} = K_{12} C_1 V_1 - K_{21} C_2 V_2,$$

- Vascular space tumour:

$$V_{vs}\frac{dC_{vs}}{dt} = PLQC_1 - (PLQ - L)C_{vs} - (1 - vref)LC_{vs} - CL_{up}C_{vs} + CL_{up}FRC_b,$$

- Endosomal space monoclonal antibody (mAb) unbound to FcRn:

$$\frac{dC_{ub}}{dt} = \frac{CL_{up}}{V_{es}}(C_{vs} + C_{is}) - K_{onFcRn}C_{ub}FcRn + K_{offFcRn}C_b - K_{deg}C_{ub},$$

- Endosomal space mAb bound to FcRn:

$$\frac{dC_b}{dt} = -\frac{CL_{up}}{V_{es}}C_{vs} + K_{onFcRn}C_{ub}FcRn - K_{offFcRn}C_b,$$

- Endosomal FcRn:

$$\frac{dFcRn}{dt} = \frac{CL_{up}}{V_{es}}C_{vs} - K_{onFcRn}C_{ub}FcRn + K_{offFcRn}C_b,$$

- Interstitial compartment:

$$V_{is}\frac{dC_{is}}{dt} = (1 - vref)LC_{vs} - (1 - vref)LC_{is} - CL_{up}C_{is} + CL_{up}(1 - FR)C_b - K_{onPD1}C_{is}V_{is}(C_{PD1_t} - PD1_t) + K_{offPD1}PD1_tV_{is},$$

- Drug receptor binding in the tumour:

$$\frac{dPD1_t}{dt} = K_{onPD1}C_{is}(C_{PD1_t} - PD1_t) - K_{offPD1}PD1_t - K_{degPD1}PD1_t,$$

- Drug receptor binding in blood:

$$\frac{dPD1_b}{dt} = K_{onPD1}C_1(C_{PD1_b} - PD1_b) - K_{offPD1}PD1_b - K_{degPD1}PD1_b,$$

- Tumour PD-1 receptor upregulation and elimination (amount):

$$\frac{dMPD1_t}{dt} = k_{in}\left(1 + E_{maxtp}\frac{PD1_t}{EC_{50tp} + PD1_t}\right) - k_{out}M_{PD1t},$$

- Receptor occupancy:

$$R0_{tumour} = 100 * \frac{PD1_t}{C_{PD1_t}}.$$

Radiotherapy

In the radiotherapy module, the survival probability of each cell is simulated with a modified linear square model and oxygen modification factor (OMF) according to Powathil et al (Powathil et al. 2013).

$$OMF = \frac{OER(pO_2)}{OER_m} = \frac{1}{OER_m} \frac{OER_m pO_2(l) + K_m}{pO_2(l) + K_m}$$

Here, $pO_2(x)$ describes the oxygen concentration at location $l(x, z, y)$. The ratio of the radiation doses needed for the same cell kill under anoxic and oxic conditions is OER, the oxygen enhancement ratio. $OER_m = 3$ is the maximum ratio, where m denotes the maximum. $K_m = 3mm$ Hg is defined as the pO_2 at half the increase from 1 to OER_m.

The modified linear square model is used to calculate the survival probability of each cell:

$$S(d) = exp[\gamma(-\alpha OMFd - \beta(OMFd)^2)]$$

$S(d)$ describes the survival probability of a cell which has a certain oxygen modification factor due to its position and receives a radiation dose d. α and β are sensitivity parameters. The sensitivity parameter γ reflects the different susceptibility of each cell kind and cancer cell cycle phase. In particular:

$$\gamma = \begin{cases} 1, & \text{for cancer cells in S-G2-M phase (Powathil et al. 2013),} \\ 0.5, & \text{for cancer cells in the G1 phase (Powathil et al. 2013),} \\ 0.25, & \text{for cancer cells in the G0 phase (Powathil et al. 2013),} \\ 1, & \text{for immune effector cells [assumed],} \\ 0.6, & \text{for immune suppressor cells [assumed].} \end{cases}$$

The sensitivity parameter γ ranges from 0 to 1. The S, G2, and M phases are the most radiosensitive phases with $\gamma = 1$, while the resting phase G0 is the least affected phase by radiation with $\gamma = 0.25$. The G1 phase has a γ value of 0.5. Immune effector cells are more radiosensitive than immune suppressor cells (Liu et al. 2015). Hence, it is assumed that γ is 1 for effector cells and 0.6 for suppressor cells.

Considering that damage caused by low-dose radiation (less than 5 Gy), can be repaired within hours, a modified linear square model is used (Powathil et al. 2013). This leads to a modified survival probability:

$$S^*(d) = \begin{cases} S, & d > 5, \\ \\ S + (1 - S) \times 0.5, & d \leq 5. \end{cases}$$

The cell cycle delay in the G1 and G2 phase after radiation is drawn from a uniform distribution between 1-9h (Powathil et al. 2013). A cell survives if a random number drawn from a uniform distribution in (0-1) is smaller than the modified survival probability $S^*(d)$ and dies otherwise.

Immune cell infiltration is modelled deterministically according to Alfonso et al (Alfonso et al. 2021) and depends on the number of radiotherapy doses N_{RT}, time of irradiation t_j, number of cancer cells killed by the irradiation K_{Tj}, a stimulation decay c, and immune cell recruitment factors δ_E for immune effector cells E_i and δ_S for suppressor cells S_i.

$$E_i = \sum_{j=1}^{N_{RT}} K_{Tj}(\delta_E e^{-c(t_i - t_j)}) \tag{5.1}$$

$$S_i = \sum_{j=1}^{N_{RT}} K_{Tj}(\delta_S e^{-c(t_i - t_j)}) \tag{5.2}$$

Each radiation dose N_{RT} cause the death of a certain cell number K_{Tj}. The more cells are being killed by radiation, the more debris will be created which attracts more immune cells. Additionally, the time point of radiation t_j is important. Immediately after time point t_j, the immune cells attraction is the highest and decrease with the increase of time as it is being modelled with an exponential decay $e^{-c(t_i - t_j)}$. The recruitment factors δ_E and δ_S and the stimulation decay c can be used to modify the immune cell infiltration

after radiation.

Chemotherapy

For the chemotherapy module, we simulated treatment with docetaxel with a kinetic/pharmacodynamic (K/PD) model from Frances et al (Frances et al. 2011). In particular, we consider

$$\frac{dD}{dt} = -k_d D + u_D(t)D(0)$$

where $D(t)$ describes the docetaxel amount in g at time t, k_D the 'biological constant' describing the elimination of the drug out of the body to control the dose history profiles, $u_D(t)$ the dosing schedule, and D(0) the initial dose. To take into consideration that the drug distribution differs throughout the 3D lattice, the drug concentration is scaled based on the predicted oxygen gradient. To take emerging resistance into account, the killing rate $f_D(t)$ depends on the time t, the amount of drug in the system $D(t)$, the efficacy rate p_D, and the resistance parameter r_D:

$$f_D(t) = p_D e^{-r_D t} D(t).$$

The chemotherapy will affect immune cells and cancer cells in the G2 and M phases. The death probability is calculated by the killing rate $f_D(t)$ divided by the maximum killing rate. A random number is drawn from a uniform distribution in (0,1), if it is smaller than the survival probability, the cell will die and survive otherwise.

DNA damage response inhibitor

The DNA damage response inhibitor treatment module is based on a PKPD model from Terranova et al (Terranova et al. 2021):

$$\frac{dCEN}{dt} = -(q + cl)\frac{CEN}{v1} + q\frac{PER}{v2}$$

$$\frac{dPER}{dt} = q(\frac{CEN}{v1} - \frac{PER}{v2})$$

where CEN is the drug amount in mg/m^2 given i.v. with a body surface area of 1.82 m^2, resulting in a drug amount in mg in the central compartment. From there the drug amount can be distributed and redistributed to the periphere compartment with rate q where it will be the drug amount PER. The drug is cleared from the body with rate

cl. $v1$ is the volume of the central compartment and $v2$ is the volume of the periphere compartment. To take into account that the drug that reaches the cancer cells depends on the location, the drug amount predicted by the PK ODE model is scaled based on the predicted oxygen gradient. The drug effect $E(l,t)$ at time t and at a certain location $l(x,y,z)$ is modelled by using the Emax model

$$E(l,t) = Emax\frac{CEN(l,t)^h}{EC_{50}^h + CEN(l,t)^h}$$

where $Emax$ denotes the maximal drug effect. Emax is 1 when the repair inhibition is complete. EC_{50} is the drug concentration that achieves half of the maximal drug effect ($0.5*E_{max}$) and h is the Hill coefficient.

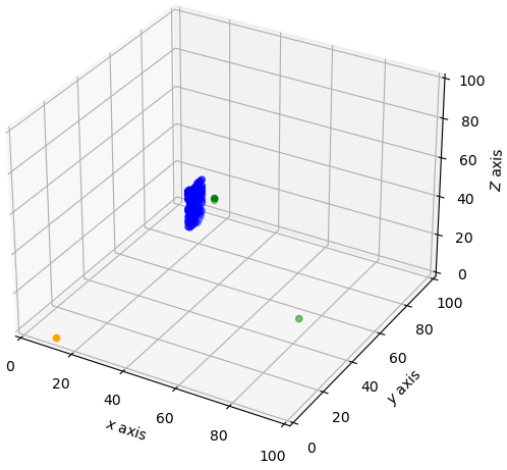
The drug only affects cancer cells in the S phase. A random number is drawn and compared with survival probability E(l,t). If that number is bigger than E(l,t), the cell dies.

5.3 Results

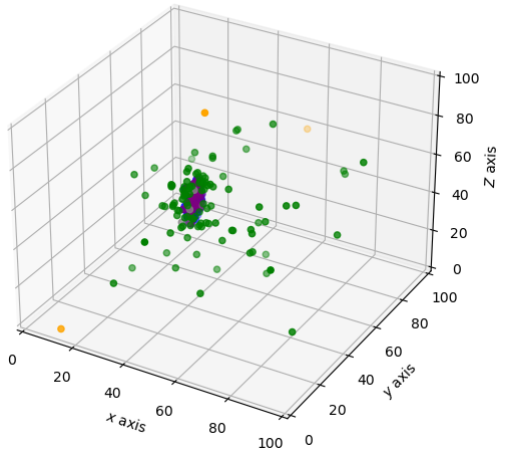
We simulate different tumours with their parameters with and without treatment. The python code can be found at https://github.com/VanThuyTruong/ABM_PKPD.

5.3.1 Tumour immune cell interaction without treatment

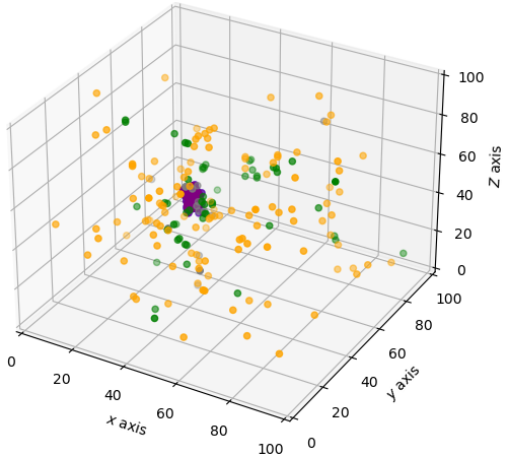
Simulating a tumour starting with 100 cancer cells without treatment intervention, we can see that the cancer grows and attracts immune cells (Figure 5.4a-5.4b). Effector cells eliminate some cancer cells and as a result, the cancer will express PDL1. Interactions between immune effector cells and PDL1+ cancer cells and infiltrated suppressor cells cause effector cell exhaustion (Figures 5.4c-5.4f). Over time the TME becomes more and more immunosuppressive (see Figures 5.4c-5.4f and Figures 5.5a-5.5c). At the end of simulation, the exhausted effector cells are gathered around the tumour as a 'protective' layer which prevents the active effector cells to come close enough to the tumour to eliminate cells (Figures 5.5c-5.5d). Overall, an exponential growth can be seen (Figure 5.5e).



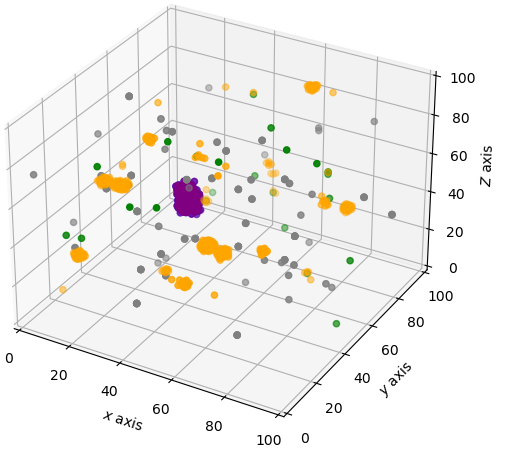
(a) TME at t=0.8d



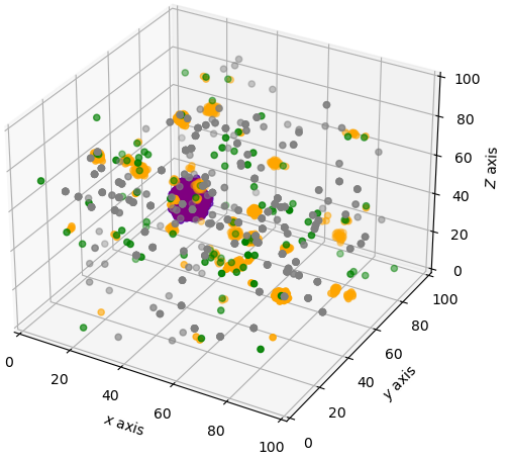
(b) TME at t=20d



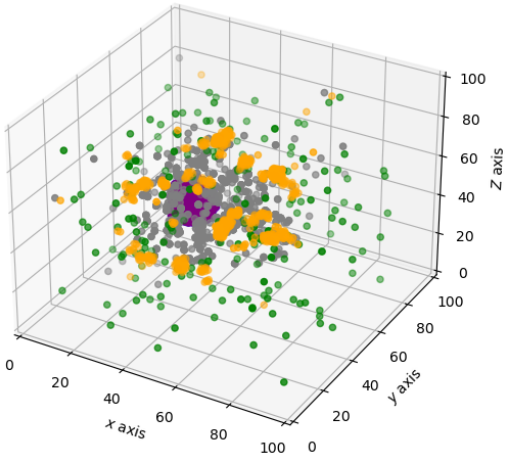
(c) TME at t=48d



(d) TME at t=97d

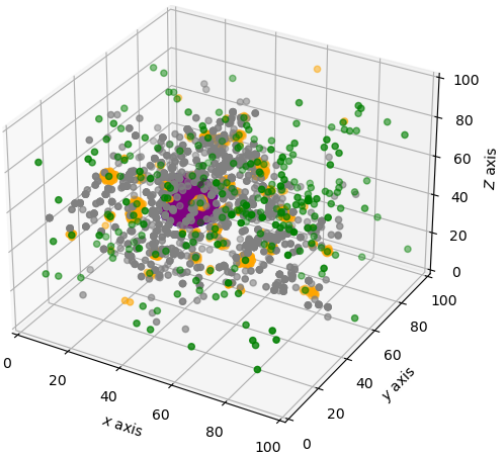


(e) TME at t=154d

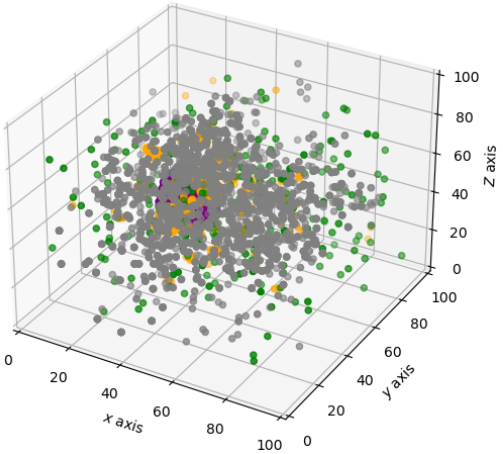


(f) TME at t=190d

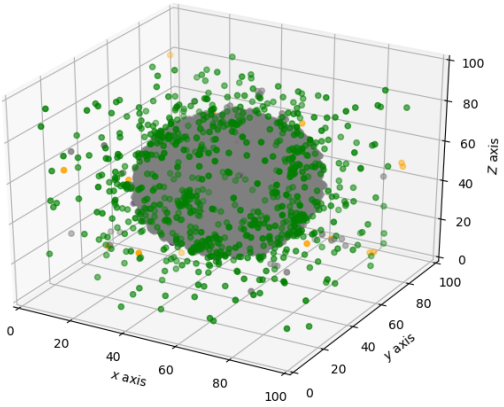
Figure 5.4: Figures 5.4a – 5.4f show the interaction of tumour and immune cells over time.



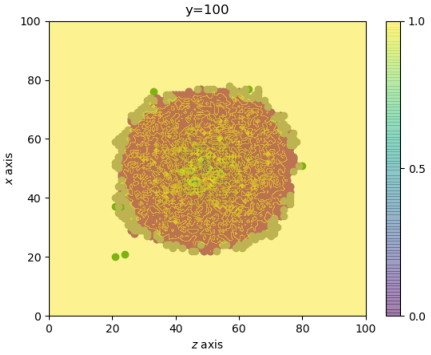
(a) TME at t=214d



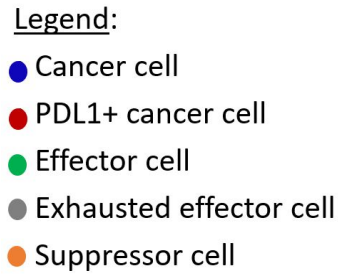
(b) TME at t=248d

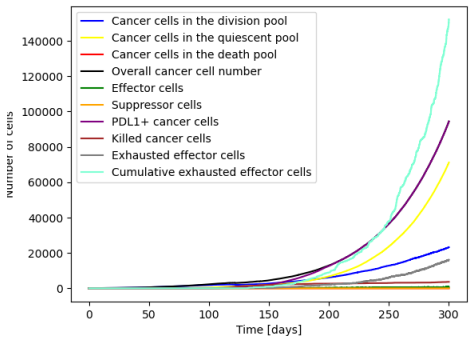


(c) TME at t=300d



(d) Slice through the tumour at end of simulation





(e) Cell count over time

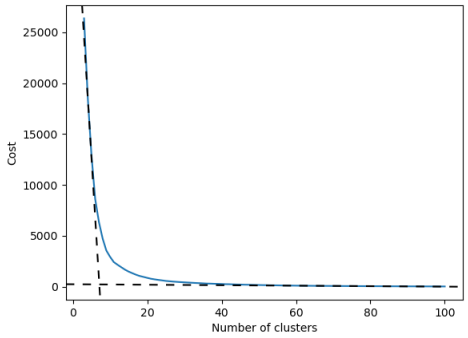
Figure 5.5: Figures 5.5a – 5.5d show the interaction of the tumour cells with the immune system over time and at the end of the simulation. Figure 5.5e shows the cell count over time.

Analysis of the cell cell interaction without treatment

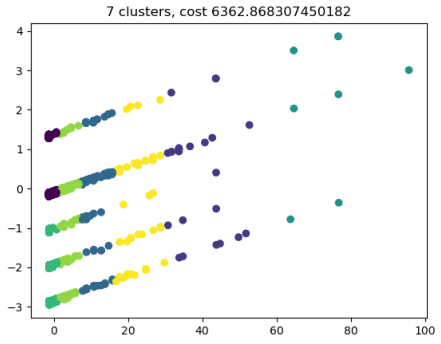
We use PCA and kmeans to analyse the data produced by the agent based model to understand the behaviour of the different kinds of agents. The model is run without treatment. For each cancer cell, we collect information about their position, status (active dividing, quiescent, cell cycle phase, death due to necrosis or being eliminated by immune effector cells), their time to division, PDL1 status, and their effector cell and suppressor cell encounter. The collected information about the effector cells are their current position, their initial position in the tumour microenvironment, their distance and number of times they have moved, their age, time to division, division count, the number of encounters with cancer cells and suppressor cells and their status (exhausted, alive, dead). The suppressor cells were not analysed as their number is low in comparison to the cancer and effector cells. Hence, they are assumed to have a similar pattern of behaviour. The collected data set is multi-dimensional and we identify the first two principal components by using PCA. Those two principal components describe the variation in the data set with dimensions thus making the data set easier to visualise. After utilizing PCA, we plot the data in two dimensions and use the k-means clustering algorithm to identify clusters of data points that are closely related, to further classify and understand the data set.

We simulated different numbers of clusters (K) for the cancer cell population and plotted them against the cost (see Figure 5.6a). By extending the downwards slope from $K = 1$ and putting a tangent on the cost function for larger K values we create a point of intersection which gives us 7 as the optimal number of clusters for the cancer cell population. Those clusters contain cancer cells that are similar in their behaviour and attributes (for example, their position, status, and encounter with other cells). Those cancer cell subgroups are shown in different colours in Figure 5.6b. The same method was applied to the immune effector cells. The results show that the optimal number of clusters for the effector cells population is 6 (see Figures 5.7a and 5.7b). If we simulate those interactions with a system of ODEs, 7 equations for the different cancer cell subgroups, 6 equations for the effector cell subgroups and 1 equation for the suppressor cells would be required. Additionally, more equations might be needed to capture different compartments for the interaction of cells in different locations of the tumour micro environment. To sum up, this analysis shows that ABM is suitable to simulate the complex behaviour of different kinds of cells with a rules and attributes. Instead of having equations for each subtyp of cancer and effector cell, they are characterised with attributes in the ABM.

Rules determine actions such as mutation of cancer cells, exhaustion of effector cells or the movement of cells to a certain location. The cells have (x, y, z) coordinates as an attribute for their position instead of having a large number of compartments for different locations in the tumour micro environment. Additionally, unexpected emerging behaviour can occur due to the interaction of the different kinds of cells and the stochasticity in the underlying Gillespie algorithm.

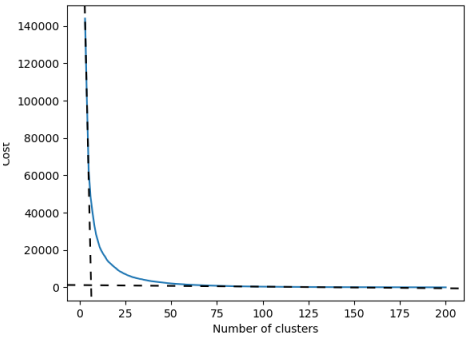


(a) The Elbow method shows that the optimal number of clusters for the cancer cell population is 7.

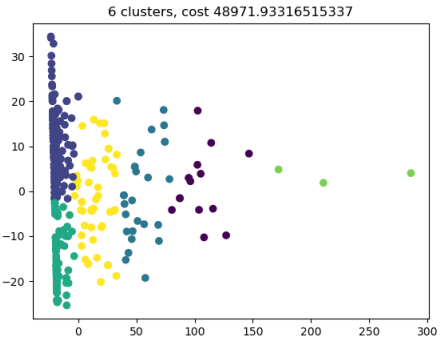


(b) Visualisation of the cancer cell subgroups after reduction to two dimensions with PCA and grouping them in 7 clusters with the kmeans algorithm, shown here with different colouring

Figure 5.6: The analysis of the cancer cell behaviour show that they can be divided in 7 subgroups.



(a) The Elbow method shows that the optimal number of clusters for the effector cell population is 6.

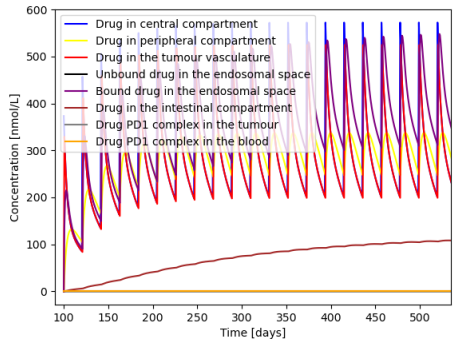


(b) Visualisation of the effector cell subgroups after reduction to two dimensions with PCA and grouping them in 6 clusters with the kmeans algorithm, shown here with different colouring

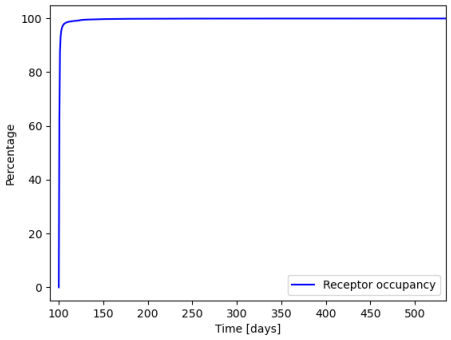
Figure 5.7: The analysis of the effector cell behaviour show that they can be divided in 6 subgroups.

5.3.2 PD1 antibody treatment

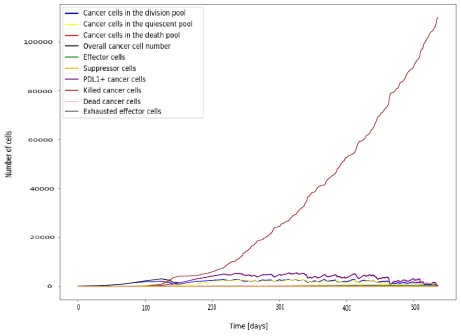
A PD1 antibody treatment with a dose of 2 mg/kg Q3W is simulated. The pharmacokinetics of the antibody, including the drug distribution in the central and peripheral compartment, tumour vasculature, endosomal space, intestinal compartment and the drug PD1 complex, can be seen in 5.8a. This treatment schedule leads to almost 100% receptor occupancy (see Figure 5.8b) which prevents immune effector cell exhaustion. To show the importance of the immune cell infiltration, a tumour parameter set with a slow immune cell infiltration is chosen. Starting treatment at day 100 with 105 infiltrated effector cells and 1.5% PDL1 mutated cancer cells, the cancer cell growth decreases by 51% in the first two months and tumour extinction happens at day 535 (Figure 5.8c). Starting at day 200 day with the same treatment and 100% mutated cancer cells, 1400 exhausted effector cells, and 310 effector cells, extinction happens at day 345 (Figure 5.8e). This simulation shows the importance of immune effector cell infiltration. The treatment is more effective at a later time point despite the larger cancer cell and exhausted effector cell count because more immune effector cells are in the TME at that time which can contribute to tumour cell elimination. Figures 5.8d and 5.8f, which are snapshots of the TME taken at the same time point, but with different treatment start times, emphasise this. In the case of early PD1 treatment (Figure 5.8d) less immune effector cells have infiltrated the TME. Therefore, even though immune cell exhaustion does not happen since the PD1 receptor is blocked before the majority of cancer cells become PDL1 positive, the time to tumour extinction is longer (after 535 days) in comparison to the case of a later treatment starting at 200 days (extinction happens at day 345). In addition, by keeping the tumour small less tumour antigen is produced and therefore less immune effector cells are being attracted into the tumour micro environment.



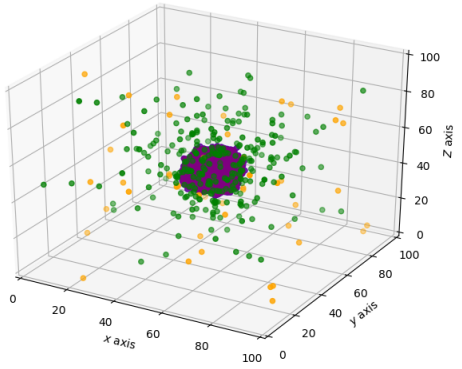
(a) PK model (drug concentration in different compartments)



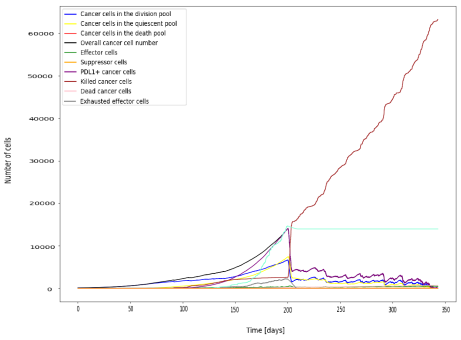
(b) PD model (receptor occupancy over time)



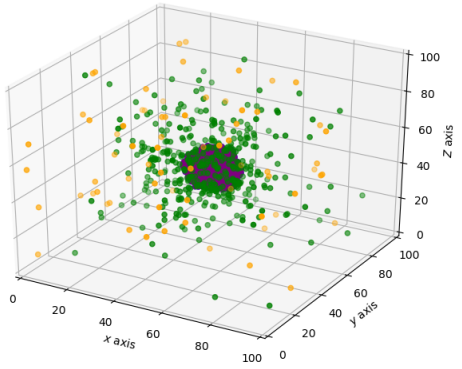
(c) Cell count over time with a treatment of 2 mg/kg Q3W starting at day 100



(d) TME at t=300d with a treatment of 2 mg/kg Q3W starting at day 100



(e) Cell count over time with a treatment of 2 mg/kg Q3W starting at day 200

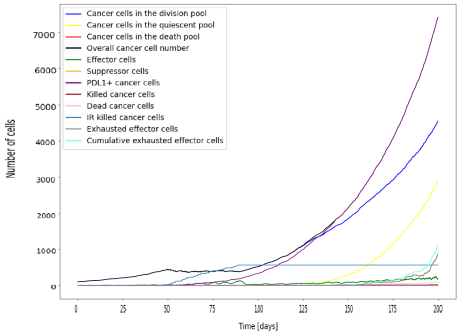


(f) TME at t=300d with a treatment of 2 mg/kg Q3W starting at day 200

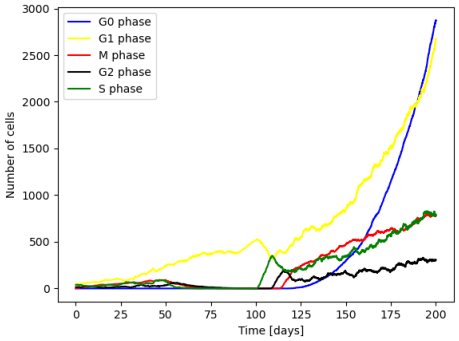
Figure 5.8: Figures 5.8a – 5.8f show the PKPD of a PD1 antibody treatment and the interaction of tumour and immune cells over time with different treatment schedules.

5.3.3 Radiotherapy

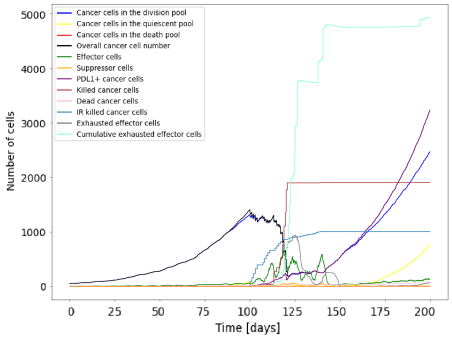
Investigating radiotherapy also shows the importance of the immune reaction. The radiation causes cancer cell death and creates debris which increases the immune effector cell infiltration. Radiotherapy is given as a treatment of 2.5 Gy 5 days a week for 7 cycles. Starting at day 50, less immune effector cells have infiltrated the TME. Therefore, even with the increased immune response after radiation the tumour cell population reaches a steady state during treatment and growth is continuing after end of treatment (see Fig. 5.9a). In comparison, at day 100 more immune effector cells are in the TME. The combination of radiotherapy induced cytotoxicity and immune cell infiltration decreases the tumour cell count by around 96% at the end of treatment (Figure 5.9c). Due to radiation and cancer cell elimination by immune cells, PDL1 positive mutants emerge and cause effector cell exhaustion. Despite this decrease the tumour population can recover and increases exponential after the treatment finishes. Those simulations also show that radiotherapy affects the cell cycle phases differently and causes cell cycle delays and redistribution (see Figures 5.9b, 5.9d).



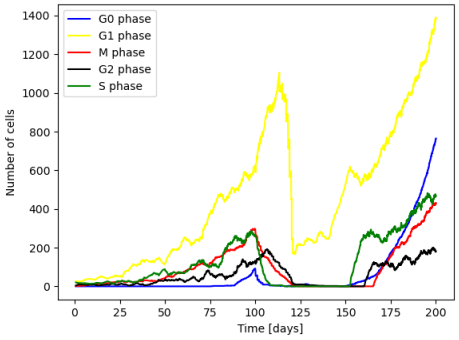
(a) Starting radiation at day 50, the tumour cell population reaches a steady state during the treatment period and regrows after end of treatment. The radiation mediated immune cell infiltration is not high enough to have a significant contribution to decrease the tumour cell count.



(b) Cell cycle phases over time with a radiation treatment of 2.5 Gy 5 days/week for 6 weeks starting at day 50



(c) Starting radiation at day 100, the tumour cell population decreases due a combination of direct radiation-mediated cytotoxicity and the stimulation of an anti tumour immune response which is stronger than in the case of an early treatment start as more immune cells have infiltrated the tumour microenvironment at day 100.

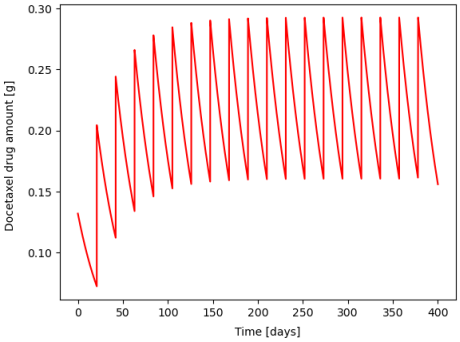


(d) Cell cycle phases over time with a radiation treatment of 2.5 Gy 5 days/week for 6 weeks starting at day 100

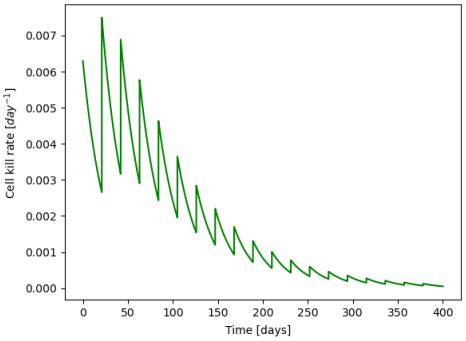
Figure 5.9: Figures 5.9a – 5.9d show the effect of radiotherapy with a weekly dose of 2.5 Gy/day for 5 days and 7 cycles starting at day 50 and day 100.

5.3.4 Chemotherapy

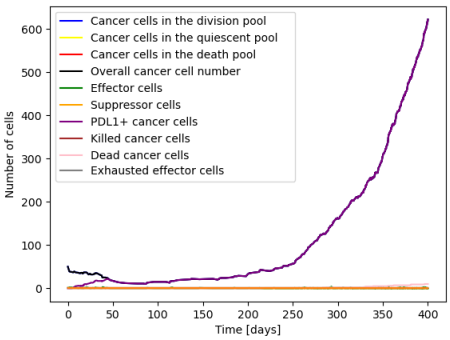
Looking into chemotherapy with docetaxel 0.132g Q3W we can see that the time point for starting the treatment and the tumour size is crucial. The administration of 0.132g docetaxel Q3W with its concentration over time is shown in Figure 5.10a. This causes a kill rate which decreases over time due to the emerging resistance of the cells (Figure 5.10b). Starting the treatment at day 0 initially decreases the cell count, and causes a plateau in the cancer cell count until resistance emerges and the tumour can grow exponentially to around 600 cells at the end of the simulation (Figure 5.10c). Starting treatment at day 50, a sharp decrease after the initial dose emerges as docetaxel cause cell death of cancer cells in the drug-sensitive G2 and M phase. The cancer cell number decreases steadily until resistance emerges around day 220 and the tumour grows again until over 500 cells at the end of the simulation (Figure 5.10d). The same treatment beginning at day 100 has a bigger impact on the cancer cell number. Initially, there is a sharp decline in the cell number after the first dose which follows a time span of approximately 20 days with an almost constant cell number and then a sharp decline again (Figure 5.10e). Looking at the plots with the cell cycle phases over time, we can see that after the initial cell kill of cells in the G2 and M phase, the cells in the dormant state become active again as the cell death frees up space and oxygen supply can increase. Dormant cells can enter the cell cycle and increase the cells in the G2 and M phases which are susceptible by chemotherapy (Figure 5.10f).



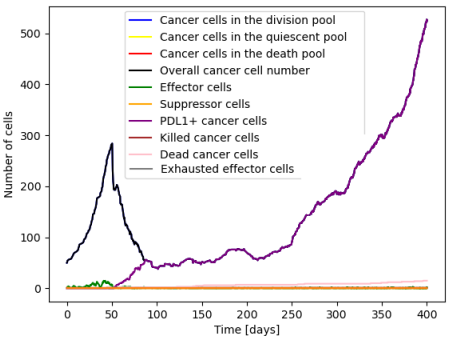
(a) Docetaxel concentration over time



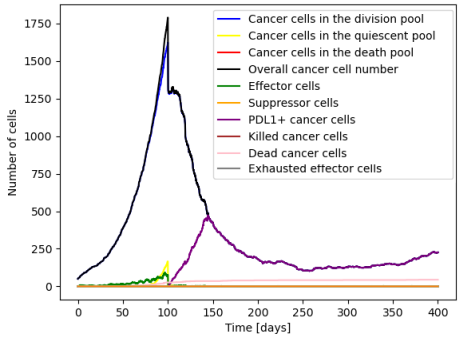
(b) Docetaxel kill rate over time including emerging resistance



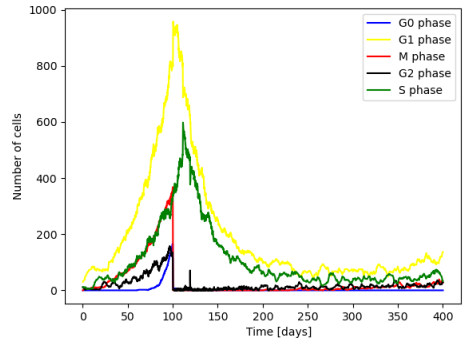
(c) Cell count over time with a treatment of 0.132g docetaxel Q3W starting at day 0



(d) Cell count over time with a treatment of 0.132g docetaxel Q3W starting at day 50



(e) Cell count over time with a treatment of 0.132g docetaxel Q3W starting at day 100

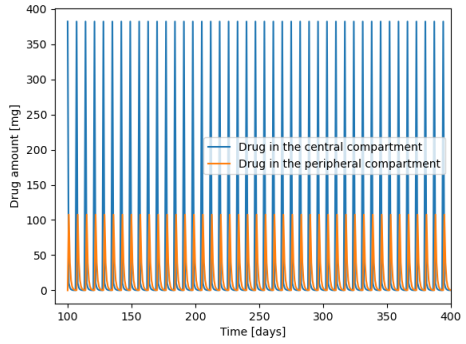


(f) Cell cycle phases over time with a treatment of 0.132g docetaxel Q3W starting at day 100

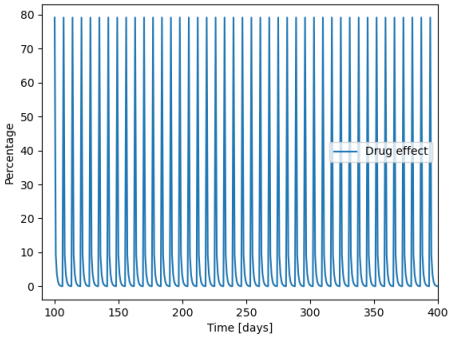
Figure 5.10: Figures 5.10a – 5.10f show the chemotherapy treatment with 0.132g docetaxel Q3W with the resulting drug concentration, kill rate, cell count and cell cycle phases.

5.3.5 DNA damage response inhibitor

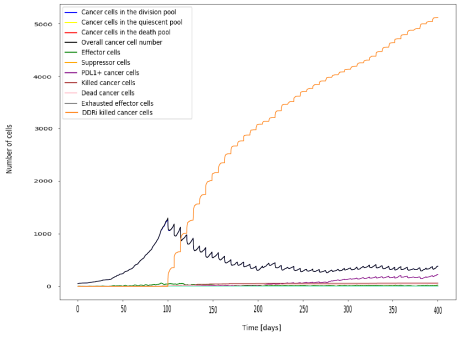
Investigating the DNA damage response inhibitor treatment we can see the drug concentration in different compartments from a weekly dose administration of 210 mg/m^2 (Figure 5.11a) and the resulting DNA repair inhibition (Figure 5.11b). Figure 5.11c shows that treatment starting at day 100 causes a decline in the cell count from around 1400 cells to a steady state of around 500 cells. We can see in Figure 5.11d that after treatment start at day 100, the cell number in the S phase oscillate as the treatment affects this cell cycle phase. But due to the weekly treatment schedule, cells from the G1 phase enter the S phase. This causes an increase which will lower once the treatment is given. Figure 5.12a displays not only a decline in cancer cell numbers after the start of treatment at day 200 but also shows a decrease in exhausted effector cells. The plot with the rates (Figure 5.12c) emphasises this by showing that the effector cells exhaustion decreases after treatment initiation at day 200. The cancer cell count reaches a steady state of around 500 cells. Looking at Figures 5.11c and 5.12a we can see that the DNA damage response inhibitor treatment only suppresses the cancer growth but does not eliminate the tumour because it only lowers the exhaustion rate but does not increase immune effector cell infiltration (see Figure 5.12c for the rates). Comparing the plot of the TME (Figure 5.12d) with the TME during PD1 antibody treatment (Figures 5.8d and 5.8f) we can see that there is not a sufficient number of immune effector cells to eliminate the cancer.



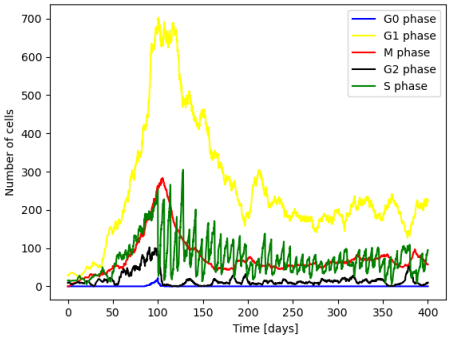
(a) DNA damage response inhibitor concentration over time in different compartments



(b) Drug effect (DNA repair inhibition) over time

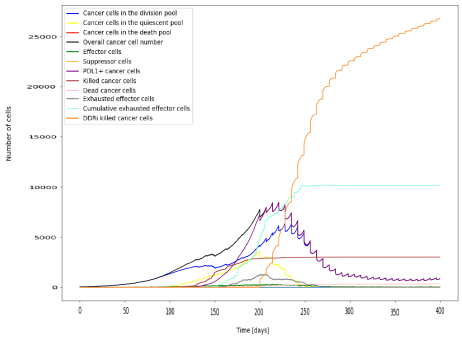


(c) Cell count over time with a treatment of weekly 210 mg/m^2 DNA damage response inhibitor starting at day 100

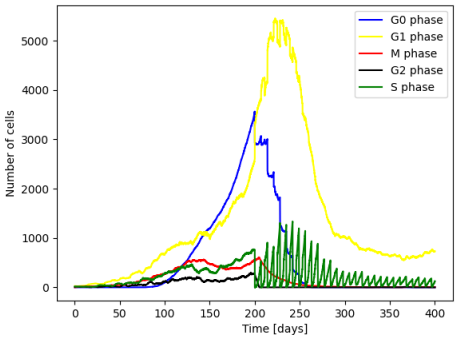


(d) Cell cycle phases over time with a weekly treatment of 210 mg/m^2 DNA damage response inhibitor starting at day 200

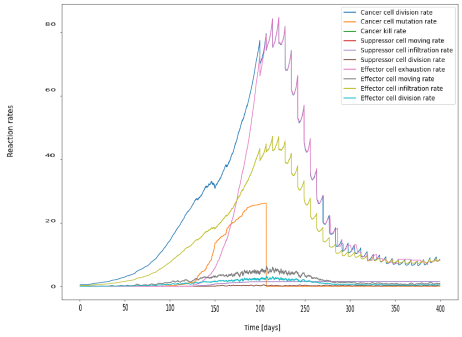
Figure 5.11: Figures 5.11a – 5.12d show the DNA damage response inhibitor treatment with weekly 210 mg/m^2 with the resulting pharmacokinetic and pharmacodynamic, cell count, rates of cell interaction, and the TME.



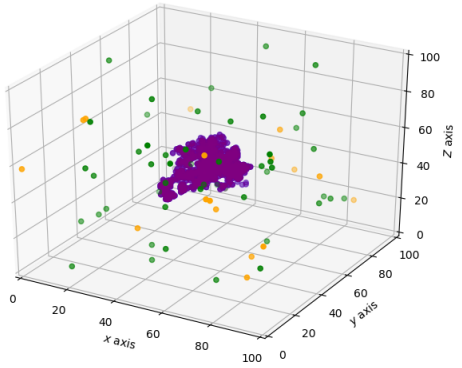
(a) Cell count over time with a treatment of weekly 210 mg/m^2 DNA damage response inhibitor starting at day 200



(b) Cell cycle phase distribution over time with a treatment of weekly 210 mg/m^2 DNA damage response inhibitor starting at day 100



(c) Cell interaction rates over time with a treatment of weekly 210 mg/m^2 DNA damage response inhibitor starting at day 200



(d) Tumour micro environment at the end of simulation with a treatment of weekly 210 mg/m^2 DNA damage response inhibitor starting at day 200

Figure 5.12: Figures 5.11a – 5.12d show the DNA damage response inhibitor treatment with weekly 210 mg/m^2 with the resulting pharmacokinetic and pharmacodynamic, cell count, rates of cell interaction, and the TME.

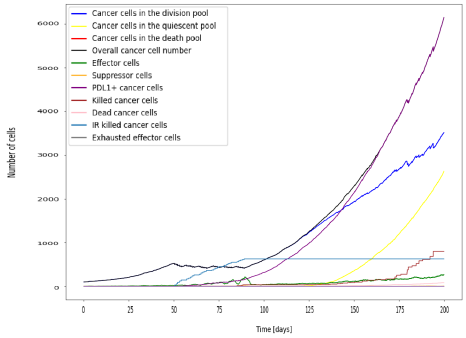
5.3.6 Combination therapy

To understand interactions between different treatment options, we simulate combination treatment. Looking at the combination of PD1 antibody with radiotherapy, we see that the right schedule is important. To introduce the subsequent analysis, the PD1 antibody treatment is given prior to radiotherapy. In Figure 5.13a we can see that the outcome of the combination therapy with 2 mg/kg PD1 antibody Q3W starting at day 20 with radiation of 2.5 Gy 5 d/week starting at day 50 for 6 weeks is similar to the case of radiation mono therapy with the same schedule (see 5.9a for comparison). In contrast, combining PD1 antibody with a later radiation treatment starting at day 100 causes tumour extinction (see Figure 5.13b). This is caused by the higher immune cell infiltration at the later time. In addition, radiotherapy causes debris which attracts immune effector cells to the tumour micro environment. Therefore, the cancer population extinction is caused by a combined effect of radio toxicity and immunogenetic cancer cell elimination. This example shows that an established immune response to cancer is crucial to aid treatment success. More information about adapting radiotherapy to the immune response can be found in the review article of Galluzzi et al. (Galluzzi et al. 2023).

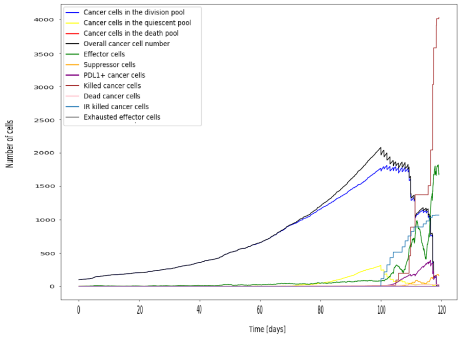
The combination of a DNA damage response inhibitor with radiotherapy demonstrates the importance of finding the optimal schedule. The simulation shows that treatment with weekly 210 mg/m^2 DNA damage response inhibitor starting from day 100 and 2.5 Gy radiation for 5 days a week starting from day 150 for 6 weeks is detrimental. When starting the DNA damage response inhibitor early at day 100, the immune response has not been established and only a low number of immune effector cells have infiltrated the TME (Figure 5.13c). Decreasing the cancer cell count with the DNA damage response inhibitor lowers the antigenicity and immune cell infiltration as in this model the immune effector cell infiltration rate depends on the number of cancer cells in the TME. When the radiation starts at day 150, immune cells and cancer cells are eliminated due to radiotoxicity. Radiation-induced cancer cell death leads to an increase in PDL1 expression. This causes a higher immune effector cell exhaustion rate (see Figure 5.13e). In combination with the decreased immune effector cell count and the increased immune cell exhaustion, the equilibrium of cancer growth and elimination by the immune system is disturbed and the tumour can grow. This cause a peak in the number of exhausted effector cells around 190 days and a peak in the cancer cell count around day 210. The increased immune cell infiltration due to the cancer cell debris caused by radiotherapy between day 150 and 200

decreases the cancer cell count after the peak around day 210 and leads to an equilibrium again from approximately day 225. Figure 5.13d shows the impact of treatment on the cell cycle phases. DNA damage response inhibitor treatment decreases the cells in the S phase in an oscillating manner as it is given weekly. Between 150-200 days, radiotherapy decreases the cell count in all phases.

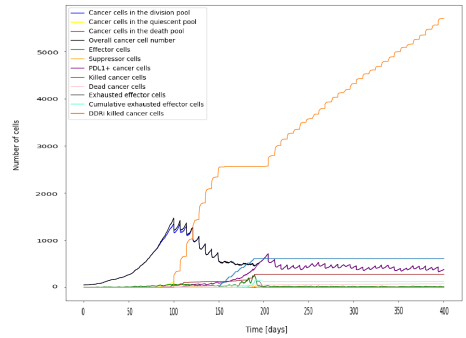
Looking at the combination of 0.132g Q3W docetaxel treatment starting at day 0 with 2.5 Gy 5 days a week radiation starting at day 100 for 6 weeks, we can see that the effect of radiation is negligible. Until the end of radiation, approximately 50 cancer cells have died due to the radiotherapy (Figure 5.13f). Once resistance to the chemotherapy emerges the cancer grows exponentially to around 600 cells at the end of the simulation. In contrast, radiation alone with the same administration schedule is able to decrease the cancer cell number by around 96% (see Figure 5.9c). This example demonstrates the important contribution of the immune system to tumour extinction. Because ongoing chemotherapy has eliminated immune cells, radiation is less effective.



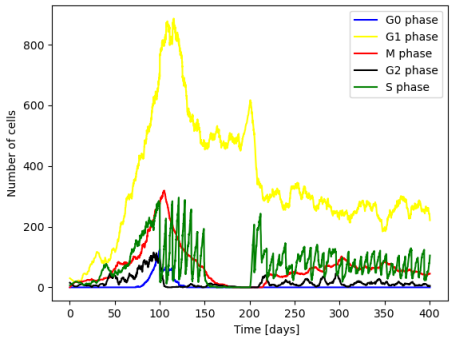
(a) Combination of 2mg/kg Q3W PD1 antibody treatment starting at day 20 with radiation of 2.5 Gy 5 days a week starting at day 50 for 6 weeks



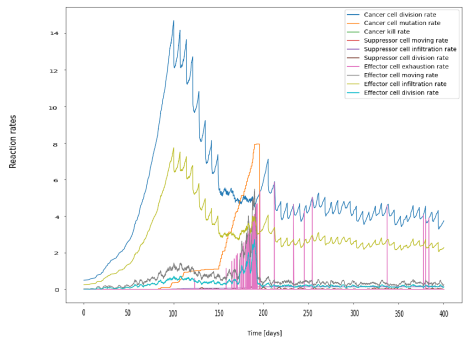
(b) Combination of 2mg/kg Q3W PD1 antibody treatment starting at day 80 with radiation of 2.5 Gy 5 days a week starting at day 100 for 6 week



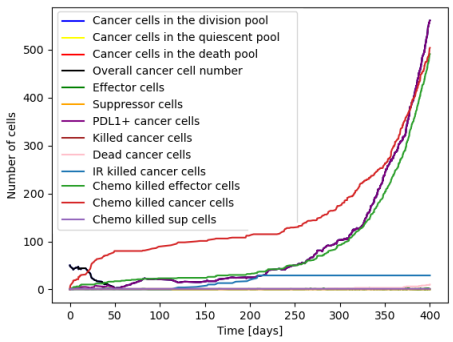
(c) Combination therapy with 210 mg/m^2 DNA damage response inhibitor starting from day 100 and radiation of 2.5 Gy 5 days a week starting at day 150 for 6 weeks



(d) Reaction rates of the ABM for the treatment case in Figure 5.13c



(e) Treatment with daily 50 mg/kg DNA damage response inhibitor starting from day 100 and 2.5 Gy radiation for 5 days a weeks starting from day 150 for 6 weeks



(f) Combination therapy with 0.132g docetaxel Q3W starting from day 0 and radiation of 2.5 Gy 5 days a week starting at day 100 for 6 weeks

Figure 5.13: Figures 5.13a – 5.13f show different options for combination treatment and their impact on the cell count and cell interaction rates over time.

5.4 Discussion

We have developed a three dimensional hybrid multi scale agent based ODE PDE model including tumour immune interaction, cell cycle phases, oxygen and drug diffusion dynamics, the pharmacodynamics of chemotherapy, targeted therapies, immunotherapy, radiotherapy and the respective systemic exposure levels of pharmacological treatments (described by pharmacokinetic modelling). The aim of this model is to support a more detailed understanding of the spatial-temporal dynamic interactions between cancer cells, the relevant immune cells, and especially targeted therapies' molecular moieties, which are all expected to interact simultaneously in the tumour microenvironment. This interaction can then be further investigated in the context of combination therapies, thus making some inroads in the mechanistic understanding of positive and negative synergies when multiple therapies are administered within this complex system. The key scientific question this work addresses is the role of temporal sequencing in combination therapies, which require the simultaneous modelling of many relevant components of the system to be properly understood.

For the chosen parameter sets, we have simulated the trajectories of tumours without therapy and with monotherapy or combination therapy. The simulations show the interplay between the cell types and which components are contributing to tumour extinction and their time dependency. The agent-based framework has the advantage of providing spatial interaction and being a rule-based stochastic bottom up method. Heterogeneity can be easily simulated. Therefore, emerging behaviour from the interplay between the different cell types and the consequences of the therapy can be observed and provide plausible mechanisms for unforeseen outcomes (e.g.treatment failure and spatial distribution of immune cells). This makes the framework suitable for simulating complex biological systems with sub-scale components (molecular, cellular, tissue, organism) and inherent emerging behaviour with limited empirical data and most relevant to elucidate unexpected behaviors (micro-known to produce macro-unknown).

Adjustments were made to simplify the model and aid in easier interpretation of the results: The cells are on a lattice set-up, immune cells are simplified and the vasculature was omitted. This could come with some limitations: having a lattice constrains the size of cells. One lattice can only accommodate a certain number of cells while in reality cancer cells could make clusters with different sizes. In addition, scaling the number of

cells in the simulation versus the number of cells in reality could cause inaccuracies in the cell-cell interaction. Further, in our simulation moving and placing daughter cells after division is only possible on discrete lattice points while in reality immune cells could move at a continuous distance. Off-lattice models could be used but those require additional computational power as collision and overlapping placements of cells need to be avoided. In this model immune cells are modelled as either suppressor or effector cells while in reality the TME consists of different types of cells such as CD8+ T cells, natural killer cells, regulatory T cells, and myeloid-derived suppressor cells (Y. Xing et al. 2021). The immune response is further simplified by omitting T-cell priming, trafficking, recruitment to the TME, and cytokines. Furthermore, the tumour vasculature is not simulated in the examples as immune effector cells can randomly enter the TME at a free location on the lattice. This could be addressed in future by limiting the entry points to certain locations of the grid. Despite these simplifications, the model can capture characteristics of cancers such as growth, PDL1 expresssion rate, antigenicity, and consequences of different treatment options on the TME and cell-cell interaction and provides insight in emerging behaviour.

Despite the limitations, our model captures experimental observations from the literature. Our simulations indicate that high immune cell infiltration improves the success of PD1 antibody treatments. In the model infiltrating immune effector cells cause PDL1 positive mutated cancer cells which drives immune effector cells exhaustion. Once the PD1-PDL1 exhaustion mechanism is counteracted by the PD1 antibody, in tumour simulations with a high immune effector cell infiltration, the immune system can erradicate the tumour. In their study Kitano et al (Kitano et al. 2017) found that expression of PD-L1 and PD-1 in early breast cancer is associated with higher number of tumour infiltrating lymphocytes and pathological complete response. Additionally, PD-L1 expression in cancer cells and PD-1 expression in tumour infiltrating lymphocytes were significantly correlated. While data show that PD-L1 and PD-1 are associated with poor clinicopathological features, paradoxically they are indicators of good prognosis, higher tumour infiltrating lymphocyte expression and pathological complete response (Kitano et al. 2017). Topalian et al. (Topalian et al. 2016) noted that PD-L1 can be driven by oncogenes or adaptive immune responses and that PD-L1 testing can guide treatment sequencing for cancers. FDA approval of a PD-L1 IHC test for pembrolizumab in NSCLC was based on a study showing higher response rates in PD-L1 positive tumors (Topalian et al. 2016).

Factors like tumor mutational burden and CD8+ T cell density relate to PD-L1 expression and might serve as easier-to-detect biomarkers. Multi-factor biomarker panels may offer better predictive value than single markers (Topalian et al. 2016). Obeid et al. (Obeid et al. 2016) took tissue microarrays of metastatic melanoma samples from 147 patients and quantified them for CD8C, CD45, CD4C, CD3, CD163, CD20, CD138, FoxP3, PD-1, PD-L1 and PD-L2 markers by immunohistochemistry. Xing et al (X. Xing et al. 2018) found that gastric carcinoma patients with high levels of PD-L1 expression had better survival rates. Those with higher T cell infiltration also showed increased levels of PD-L1, PD-L2, and PD-1 expression, which correlated with favorable outcomes, suggesting the presence of an adaptive immune resistance mechanism. Conversely, patients with low-density CD3 T cell infiltration and no PD-L1 expression in tumor cells had the worst outcomes across different pathological tumor node metastasis stages, indicating a possibly inactive immune status. These findings underscore the importance of assessing PD-L1 expression in all tumor contexts and thoroughly characterizing the immune microenvironment in gastric cancer. Studies on NSCLC (Hu-Lieskovan et al. 2019) found PD-L1, tumour mutational burden, and CD8+ T cells are associated with positive responses to PD-1 blockade, with PD-L1 being the strongest predictor of overall survival. Ott et al. (Ott et al. 2019) demonstrated that a T-cell–inflamed gene-expression profile, PD-L1, and TMB predict responses to pembrolizumab across various tumors, highlighting that combined biomarkers can identify patients more likely to respond to anti-PD-1 therapies. This comprehensive review of PD-L1 testing emphasises its importance in guiding treatment decisions and improving patient outcomes in multiple cancer types.

Similar to PD1 antibody treatment, our simulations demonstrate that the immune system is crucial for the success of radiotherapy, which is supported by existing literature. Host immune status plays a critical role in determining the effectiveness of radiation treatment. Radiation triggers the release of danger signals and chemokines that attract inflammatory cells to the tumor microenvironment, including antigen-presenting cells that activate cytotoxic T-cell function. Conversely, radiation can also draw immunosuppressive cells into the tumor microenvironment (Weichselbaum et al. 2017).

A study by Stone et al. (Slone et al. 1979) in a mouse fibrosarcoma model established that the host's immune status affects the efficacy of radiation-induced antitumor effects. Subsequent research by Lee et al. (Y. Lee et al. 2009) showed that T cells are essential for tumor regression following high-dose radiation (15–20 Gy). In their study,

mouse melanoma B16 tumors in immunocompetent hosts responded to high-dose radiation, while tumors in nude mice (without T cells and some B cells) or in wild-type hosts with depleted cytotoxic CD8+ T cells did not respond to ionizing radiation (Y. Lee et al. 2009). Additionally, radiation promoted antigen-specific T-cell priming. This effect could be suppressed chemotherapy such with paclitaxel or dacarbazine, indicating the need to carefully consider the timing and type of chemotherapy used in combination with radiation to avoid immunosuppressive effects (Y. Lee et al. 2009; Weichselbaum et al. 2017). An analysis of The Cancer Genome Atlas data (Weinstein et al. 2013) across seven cancers by Wen et al (Wen et al. 2020) revealed complex associations between immune infiltration and radiotherapy outcomes. A prognostic model based on immune infiltration status was developed. Overall, immune infiltration was found to be highly relevant to radiotherapy outcomes, highlighting the importance of understanding tumor-associated immune cell infiltration for identifying biomarkers and therapeutic targets (Wen et al. 2020).

Future work could involve calibrating the model to specific cancer types and patients to generate predictions for clinical hypothesis testing. Emerging technologies, such as multiregion sequencing, single-cell sequencing, analysis of autopsy samples, and longitudinal liquid biopsy analysis, offer potential methods to gain insight into the complex architecture of cancers (Dagogo-Jack and Shaw 2018). For instance, Topalian et al. (Topalian et al. 2016) characterised melanoma, gastric carcinoma, and breast carcinoma with PDL1+ tumor cells, with and without immune cell infiltration, using immunohistochemistry. Obeid et al. selected metastatic melanoma samples from 183 resections obtained from 147 patients, utilizing ample clinical follow-up and surgical pathology material to obtain core samples from at least 3–4 tumor regions for tissue microarray construction. They quantified those tissue microarrays for CD8, CD45, CD4, CD3, CD163, CD20, CD138, FoxP3, PD-1, PDL1, and PDL2 markers using immunohistochemistry. Similarly, Xing et al. (X. Xing et al. 2018) performed immunohistochemistry on a tissue microarray of 1,014 gastric carcinoma specimens using PD1, PDL1, and PDL2 antibodies, along with T cell markers CD3 and CD8, quantified through automated image analysis.

The immune response to radiotherapy can be measured using various imaging methods, such as MRI, single-photon emission CT, and PET, which enable visualization of the distributions and dynamics of immune cell populations (Grassberger et al. 2019). Candidate immune-based biomarkers of response to radiotherapy include: clinical blood count measurements (e.g., lymphocyte, neutrophil count), circulating immune cell sub-

populations (e.g., T cell subtypes, myeloid-derived suppressor cells, tumor-antigen-specific lymphocytes), cell surface markers (e.g., PD-1, PDL1, FAS ligand, tumor-antigen-specific cytotoxic T lymphocytes), functional assays of lymphocyte activity (e.g., in vitro antigen stimulation, response to vaccine challenge, natural killer cell activity assay), tumour-infiltrating lymphocytes (e.g., CD8+ tumor-infiltrating lymphocytes, CD8+PD-1+ TILs, tumor-antigen-specific TILs), circulating inflammatory markers (e.g., C-reactive protein, lactate dehydrogenase), circulating cytokines (e.g., interleukins, TGFβ), and humoral markers (Grassberger et al. 2019).

Due to the explicit spatial setup in our ABM, pathological images from tumor biopsies and model outputs can be compared. Patient data can serve as initial conditions for simulations, as well as for calibration and validation purposes. Detected objects, such as cancer cells and immune cells, can be mapped with corresponding agents in the model (Gong, Anders et al. 2019).

Future work could involve adjusting parameter values to fit the model to specific tumor types and simulating the tumor microenvironment architecture in greater detail, such as including vasculature. Additionally, the process of finding the optimal treatment for a given tumor could be automated using artificial intelligence methods, such as reinforcement learning.

Chapter 6

Combining a hybrid PKPD agent based model with a reinforcement learning algorithm to optimise cancer combination treatment

6.1 Introduction

Combination therapy is a regimen that combines two or more therapeutic agents. The rationale behind using this approach includes several factors. Firstly, some treatments may have complementary mechanisms of action, leading to enhanced therapeutic effects when used together. The combined effect of multiple agents can be greater than the sum of their individual effects. For example, research in cancer therapy became focused on investigating combination therapies that target different pathways to create an enhanced effect. Secondly, using a single medication can lead to the development of resistance as the constant exposure to a single compound induces cancer cells to initiate alternative salvage pathways. By using a combination of drugs that act through different mechanisms, it is possible to produce an effective treatment response in fewer cycles and therefore minimise the development of resistance. Thirdly, due to additive effects of multiple drugs, a lower dose is required. This can have a reduced toxicity as lower doses of multiple drugs can be better tolerated than higher dose of a single medication. Lastly, by combining multiple drugs, a broader spectrum of mechanism of actions can be archived. Targeting multiple as-

pects simultaneously can be useful in diseases with a complex pathophysiology or potential evasion mechanism (Mokhtari et al. 2017). While combination therapy can be highly effective, therapies such as cancer treatment involving drugs with narrow therapeutic windows, high inter-individual variability, or combination of interacting drugs require a schedule that considers the effects of multiple drugs. Treatment schedules need to be accounted for altering status such as mutation caused by previous treatment or changing immune cell infiltration. Further, different sensitivity to drugs for example due to location or cell cycle phase have to be considered. Currently in oncology clinical trials drugs are often given simultaneously without considering how effects of the individual drug on the tumour cells and tumour micro environment might enhance or decrease treatment outcome. For example in the KEYNOTE trial pembrolizumab was administered for up to 35 cycles and chemotherapy for the first 4 cycles (Paz-Ares et al. 2018). Or the DNA damage response inhibitor berzosertib is combined with different chemotherapeutic from day 1 (Terranova et al. 2021). Reinforcement learning has the potential to improve combination therapy dosing by incorporating interacting components in a flexible manner and enhance the full potential of drug combinations (Ribba et al. 2022). Instead of a concurrent treatment regime, the algorithm can analyse interactions of the drug and tumour micro environment and suggest the best time point for administration of multiple drugs.

Reinforcement Learning belongs to the class of Artificial Intelligence decision-making algorithms and is closely related to optimal control theory. It imitates learning by interaction with the environment with the aim of identifying the best action in a given situation to maximise the long-term reward (Sutton and Barto 2018). This algorithm differs from supervised learning techniques by focusing on immediate feedback from the environment instead of correcting input/output pairs or direct instruction on what to achieve, thus being able to find unexpected solution. During the learning phase the learner explores the environment by taking random actions and collects resulting rewards. It also exploits the learned knowledge and chooses the best action. The aim is to find a balance between exploration and exploitation (Ribba et al. 2022; Jalalimanesh et al. 2017). The foundation of Reinforcement Learning is built upon a Markov decision process, comprising four crucial components: States (S), Actions (A), Transition Probability, and Reward function (R). Transition probability represents the likelihood of moving from the current state to the next state, while the reward function provides the gain upon taking action in the current state ($S \times A \rightarrow R$) (Ziyan Wang et al. 2020). The decision for a future action only de-

pends on the current state as this state alone holds sufficient information to determine the optimal sequence of actions that maximise long-term rewards. Hence a Markov process is applied (Ribba et al. 2022).

The solution to the Markov decision process is a policy (π) that outlines the best actions for the agent to take in each state, along with their corresponding real values (V).

If the transition probabilities and rewards are unknown, the Markov decision process problem becomes a reinforcement learning problem. Reinforcement learning falls under the realm of machine learning, involving an agent that optimises decision-making through trials within a fixed environment with reward functions (Ziyan Wang et al. 2020). The primary goal in reinforcement learning is to maximise the value function for a controller or agent to direct the state of the environment (system) towards a target. Controller actions that lead to appropriate changes in the system receive rewards, while inappropriate actions are penalised, and these experiences accumulate in the value function. Reinforcement learning concepts can be applied to clinical decision processes in precision medicine, where medical professionals act as controllers, patients or patient populations represent the system, treatment decisions serve as actions, disease states or biomarkers as states, and the reward function denotes clinical utility (McComb et al. 2022). The algorithm's aim is to determine an optimal policy, which is a sequence of state-action pairs generating the highest cumulative reward. It is insufficient to focus solely on maximizing rewards at each step; instead, the emphasis lies on the long-term cumulative reward. An optimal policy might involve actions that do not lead to the highest instantaneous reward but ultimately maximise rewards in subsequent actions. This concept can be linked to a tennis player intentionally losing a game on the opponent's service to conserve energy and perform better in the following game, where they serve for the set.

Reinforcement learning problems are often solved using value (or action-value) functions, commonly known as "Q" which represents the expected cumulative reward from a specific state or state-action pair (Ribba et al. 2022). Once the Q values are known, the algorithm selects the action with the highest Q value for a given state. If the Q value is unknown, it is learned through experience. The algorithm undergoes a training period where it explores the environment randomly and collects rewards for the chosen actions. The epsilon-greedy algorithm is used to balance out exploration and exploitation, allowing the system to randomly explore the environment to collect unknown rewards or punish-

ments while also exploiting the known best Q values (Sutton and Barto 2018). Agent based models often involve stochastic simulations with a large number of states, making it challenging to extract a finite Markov decision process. Consequently, model-free reinforcement learning algorithms like Q-learning are well-suited for optimization. In Q-learning, we define a set of state spaces (S), a set of possible actions (A), and a reward function R: S $\times$ A $\rightarrow$ R (Jalalimanesh et al. 2017). After each iteration of the algorithm, the system selects an action (A'), leading to a new state (S'). The value of the new state determines the reward (R) of the action (A) in the current state (S). Running the algorithm involves filling the Q(S,A) matrix by aggregating the rewards for all state-action pairs. Ultimately, the algorithm converges to the best policy, represented by the maximum elements in the Q(S,A) rows (Jalalimanesh et al. 2017). The algorithm uses the Bellmann equation to update the Q values during the learning process.

$$Q(S,A) = Q(S,A) + \alpha(R + \gamma maxQ(S',A') - Q(S,A)) \tag{6.1}$$

Q(S,A) is the Q-value for taking action A in state S. It represents the expected cumulative reward when starting in state S, taking action A, and then following the optimal policy thereafter. α is the learning rate, which determines the weight of the new information when updating the Q-value. R is the immediate reward received after taking action A in state S. γ is the discount factor, which controls the importance of future rewards compared to immediate rewards. It is a value between 0 and 1. The key idea in Q-learning is to iteratively update the Q-values based on the Bellmann equation. The algorithm starts with arbitrary Q-values and then interacts with the environment by taking actions and observing rewards. After each action, the Q-value for the state-action pair is updated using the Bellmann equation, incorporating the observed reward and the estimated value of the next state (obtained by taking the maximum Q-value over all possible actions in the next state). Through repeated interactions and updates, the Q-values gradually converge towards the optimal Q-values, which represent the maximum expected cumulative reward achievable from each state-action pair. As the algorithm continues to explore and exploit the environment, the Q-values get refined, allowing the algorithm to make better decisions and follow an optimal policy that maximises long-term rewards.

Many examples using reinforcement learning to optimise treatment can be found in literature. Moodie et al applied Q-learning to estimate optimal dynamic treatment rules

from observational data (Moodie et al. 2012). Krakow and colleagues used this method to find personalised treatment recommendations from cohort and registry data (Krakow et al. 2017). Radiotherapy was optimised with Q learning by Jalalimanesh et al (Jalalimanesh et al. 2017). Escandell-Montero et al. optimised of anemia treatment in hemodialysis patients via reinforcement learning (Escandell-Montero et al. 2014). Gaweda et al. individualised anemia management using reinforcement learning (Gaweda et al. 2005). Moore and colleagues showed how reinforcement learning can be used in propofol anesthesia (Moore et al. 2014). Zhao et al use this method to discover individualised treatment regimens. The optimal policy is learned from a single training set of finite longitudinal patient trajectories (Zhao et al. 2009). This is the first report describing the integration of an agent based system pharmacology model with reinforcement learning methods. Given the mechanistic nature of the model, insights into tumour immune cell interaction and effects of combination treatment is provided while the reinforcement algorithm is used to find the optimal treatment schedule for a given cancer with its parameters.

6.2 Methods

We use the tumour immune interaction model from our previous chapter where an agent-based model is used to simulate tumour and immune effector and suppressor cells. The behaviour of each agent and interaction with each other and the environment is governed by a set of rules and attributes. Cancer cells can divide and attract immune effector and suppressor cells into the tumour micro environment. Immune effector cells can eliminate cancer cells which causes a higher mutation rate from the PDL1 negative phenotype to the PDL1 positive cancer cells. PD1-PDL1 interaction leads to immune cell exhaustion. The PD1 antibody treatment is implemented with a PBPK model (Elassaiss-Schaap et al. 2017) . The survival probability of each cell after radiation is simulated with a modified linear square model and the cell cycle delay after radiation and immune cell infiltration are considered (Powathil et al. 2013; Alfonso et al. 2021). We simulated docetaxel treatment with a K/PD model including an efficacy rate and a resistance parameter (Frances et al. 2011). Chemotherapy affects immune cells and cancer cells in the G2 and M phases. The DNA damage response inhibitor treatment only affects cancer cells in the S phase and is based on a PKPD model (Terranova et al. 2021). To take into account that the drug that reaches the cancer cells depends on the location, the drug amount predicted by the PK ODE models is scaled based on the oxygen gradient PDE.

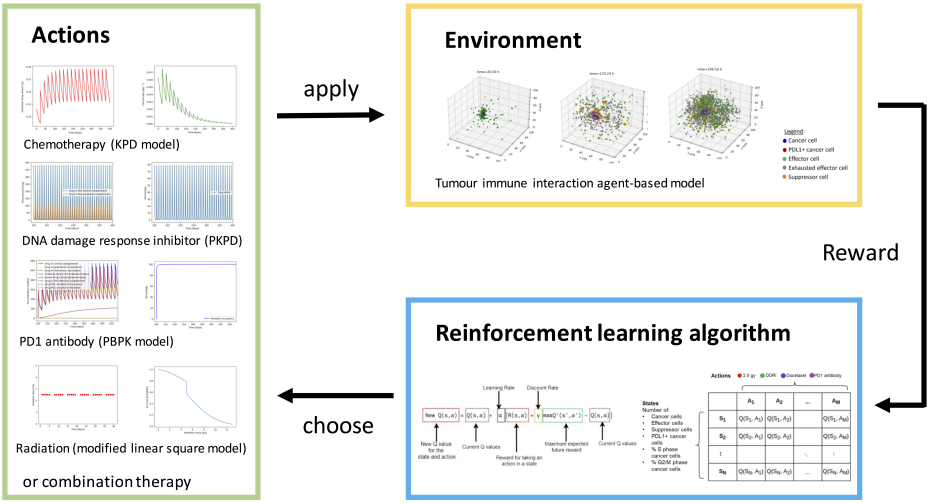


Figure 6.1: Schematic representation of the integration of PKPD, PBPK, modified linear squared model and ABM with the reinforcement learning algorithm

The underlying PKPD agent based model and the reinforcement algorithm are programmed in Python 3.8. To find the optimal combination therapy we firstly set up a Q table and define the relevant states. Number of cancer cells, immune effector and suppressor cells are important factors for the tumour immune architecture contributing to the dynamic interaction. Further, the percentage of cancer cells in the drug and radiation sensitive G2, M and S phase needs to be considered. To limit the number of states, we define the number of states for the cancer cells (*c_cells* in the code snippet below) as 100. The cancer cell number is counted in steps of 1000 cells. So, the first state contains 0-1000 cancer cells, the second has 1001-2000 cells and the last state consist of 100,000 or more cells. There are 100 states for each of the immune cells types (*sup* and eff in the code snippet below). They are counted in steps of 50 cells. The first state has 0-50 cells, the second one contains 51-100 cells and the state with the highest immune cell count consist of 500 or more cells. The percentage of cancer cells in the susceptible cell phase G2, M and S phase are summarised in 10% steps (states S and $G2M$).

Therefore, we set up the Q table as follows:

```
c_cells = 100
S= 10
G2M=10
sup=100
```

```
eff=100
action_in=7
q_values = np.zeros((c_cells ,S,G2M,sup,eff, action_in))
```

The Q table *q_values* is a six dimensional table filled with zeros initially and will be updated with the Q values $Q(S, A)$ during the runs. We set the number of actions that the algorithm can choose to 7. Therefore, the action row in the Q table has 7 columns. Possible actions are:

1. Radiation of 2.5 Gy/day (limitation: 5 $days/week$ for a maximum of 30 doses all together)
2. Radiation and PD1 antibody 2 mg/kg (limitation Q3W)
3. Radiation and DDRi 210 mg/m^2 (limitation Q1W)
4. No action
5. DDRi 210 mg/m^2 (limitation Q1W)
6. Docetaxel 0.132 g (limitation Q3W)
7. PD1 antibody 2 mg/kg (limitation Q3W)

Due to toxicity we have constrains regarding the schedule. Radiation is given as 2.5 Gy/day with the limitation of 5 $days/week$ for a maximum of 30 times all together during the simulation period. PD1 antibody is given as a dose of 2 mg/kg with a minimum break of 3 weeks until the next dose. DNA damage response inhibitor is dose at 210 mg/m^2 with a minimum break of 1 week between the doses. The treatment actions are chosen based on current therapeutic regimes. It is easily possible to modify the doses, schedules or add more treatment options depending on the interest of the user. The running time needs to be taken into consideration. With an increases number of treatment options (actions) the running time to learn from the environment and to fill the Q table will get larger.

After an action is chosen, a reward based on the number of eliminated cancer cells in that time step (cancer cells being eliminated by effector cells (*cancer.kill_by_eff* in the code snippet below), by radiation (*cancer.IR_kill*), by DNA damage response inhibitor (*cancer.DDRi_kill*), by natural death (*cancer.natural_death*), by chemotherapy (*cancer.chemo_kill*)) and immune effector cells (being eliminated by radiation ($eff.IR\_kill$), by chemotherapy ($eff.chemo\_kill$)) at that time point is calculated. The reward will be

higher if cancer cells are being killed by treatment or by immune effector cells. The reward takes delayed killing into account which can happen when immune effector cells need to be attracted into the environment by previous actions. Toxicity of treatment is discouraged by decreasing the reward by the number of eliminated immune effector cells by the treatment. The factors rx and ry can be used to encourage the algorithm to take a less toxic approach by increasing ry or a more aggressive treatment by increasing rx. The ultimate goal is to eradicate the cancer, therefore the reward for cancer elimination is set as 10000.

```
reward=rx*(cancer.kill_by_eff+cancer.IR_kill+cancer.DDRi_kill
+cancer.natural_death+cancer.chemo_kill)
        -ry*(eff.IR_kill+eff.chemo_kill)
if cancer_number==0:
    reward+=reward+10000
```

Since we set rx and ry as 1, the reward can range from the maximum number of cancer cells possible in the simulation (this number is set to 10^6 cells) as the maximum and the minimum is set by how many immune effector cells can be eliminated. Since the ratio between immune cells and cancer cells is 1:100, the minimum is -10^4.

We use the epsilon greedy algorithm which is a strategy to determine whether the Q learning algorithm should explore the environment randomly or choose the action from the Q table with the highest return. It works as follows: The function draws a random number. If that number is smaller than epsilon, the most favourable possible action in that state will be chosen, otherwise a random action will be chosen. The function *get_next_action* in the code snippet below shows how this is being coded. The value for epsilon depends on the environment. A small epsilon encourages exploration over exploitation of the known rewards. With a larger number of possible actions and outcomes the epsilon value can be smaller. Therefore, during the initial phase of the training with unknown outcomes the epsilon value can be smaller and increase with larger experience. In our example, the epsilon value starts at 0.0001 for the first run and increases in steps of 0.0001 to 1 for the last run.

```
def get_next_action(c_q1, S_q1, G_q1,sup_q1,eff_q1, epsilon,
q_values,action_in):
  #if a randomly chosen value between 0 and 1 is less than epsilon,
  #then choose the most promising value from the Q-table for this state.
```

```
if np.random.random() < epsilon:
  return np.argmax(q_values[c_q1, S_q1, G2M_q1,sup_q1,eff_q1])
else: #choose a random action
  return np.random.randint(action_in)
```

After each action the reward is used to update the Q table according to the Bellmann equation (Equation 6.1). Firstly, as seen in the code snipped below the old Q value Q(S,A) is located in the Q table with the state S and action A. The state contains a certain cancer cell number (c_q1), percentage of cancer cells in the S phase (S_q1), percentage of cancer cells in the G2 and M phase (G2M_1), the number of immune suppressor cells (sup_q1), immune effector cells (eff_q1). The action is described with the action_index. The temporal difference is the difference between the current estimate for a Q-value for a state-action pair and the updated Q-value. The temporal difference consist of the immediate reward, the maximum reward in the future state times the discount factor. The discount factor is a value between 0 and 1 that represents the importance of future rewards compared to immediate rewards. If it close to 0, the agent becomes myopic and tries to maximise immediate rewards. The new Q value is updated using the old Q value, the learning rate and the temporal difference. The learning rate scales the impact of the new information relative to the current Q-value. A high learning rate means the algorithm adapts quickly to new experiences but may forget older information, while a low learning rate makes the algorithm slower to adapt but more resistant to changes in the environment. It is essential to balance the learning rate to ensure the algorithm converges to an optimal policy. A too high learning rate can lead to oscillations or divergence, while a low learning rate can result in slow learning and potentially getting stuck in suboptimal policies.

```
old_q_value=q_values[c_q1, S_q1, G2M_q1,sup_q1,eff_q1,action_index]
temporal_difference = reward
+ discount_factor * np.max(q_values[c_q2, S_q2,G2M_q2,sup_q2,eff_q2])
- old_q_value
new_q_value = old_q_value + (learning_rate * temporal_difference)
q_values[c_q1, S_q1, G_q1,sup_q1,eff_q1, action_index] = new_q_value
```

A run starts at day 0 and ends at day 400 or earlier when the tumour has gone extinct. The algorithm is run for a large number of iteration, for example 10000 runs. This provides the opportunity to collect enough experience about the environment and

beneficial actions. The epsilon value increases with every run, hence exploration is more favoured at the beginning and during the later runs exploitation of the collected experience is dominating. After a large number of runs, the algorithm finds an optimal solution and rewards are less negative and the chosen treatments are similar as exploitation of the collected knowledge is preferred over random exploration of the environment.

6.3 Results

We simulated a treatment starting at 50 days and 100 days to mimic an early tumour detection treatment paradigm, starting later when a more developed tumour environment exists to understand potential differentiation of treatment needed. After 50 days the tumour grows to a size of 150-350 cells in the different runs as the agent-based model is stochastic and each run is slightly different. After 100 days the tumour mass reaches a size of 800-1800 cells. 10000 runs take approximately 24h. During the simulation the epsilon values increases so that in the first runs the algorithm is encouraged to explore the environment and learn about the consequences of the different treatment options on the tumour immune interaction while during the later runs the algorithm is more likely to use the experience from previous runs and choose the optimal action for a long term reward. We display the results and behaviour of selected runs for a treatment starting at day 50 in Figures 6.2 to 6.5. We can see during the runs 2-18, that all 4 possible treatment options are chosen (see Figure 6.2). In some cases the tumour goes extinct, but in other cases despite the treatment the tumour grows to approximately 780 cells from 350 cells at the beginning of treatment in run 5 or to approximately 800 cells from about 400 cells at day 50 in run 18. The rewards that the algorithm received after an action can vary from approximately -24 till 146. During the runs 3833-3881 the epsilon value has increased to approximately 0.38. The algorithm has a high likelihood to explore the environment but sometimes base its action on the previous learning. We can see in Figure 6.3 that the algorithm has learned to control the tumour growth and in no case the cancer cell number increases over the number at 50 days. In the simulations 7744-7767 (Figure 6.4) we can see that with an epsilon of approximately 0.77 the algorithm has learned to control the tumour growth and how to obtain tumour extinction. In comparison to the previous runs the rewards are less negative. The lowest reward is -6 instead of -10 in the runs with an epsilon around 0.3 or -24 in the initial runs. It can be observed that an early radiotherapy with our given tumour parameter set is disadvantageous: in the run 7745 radiation leads to a peak in the

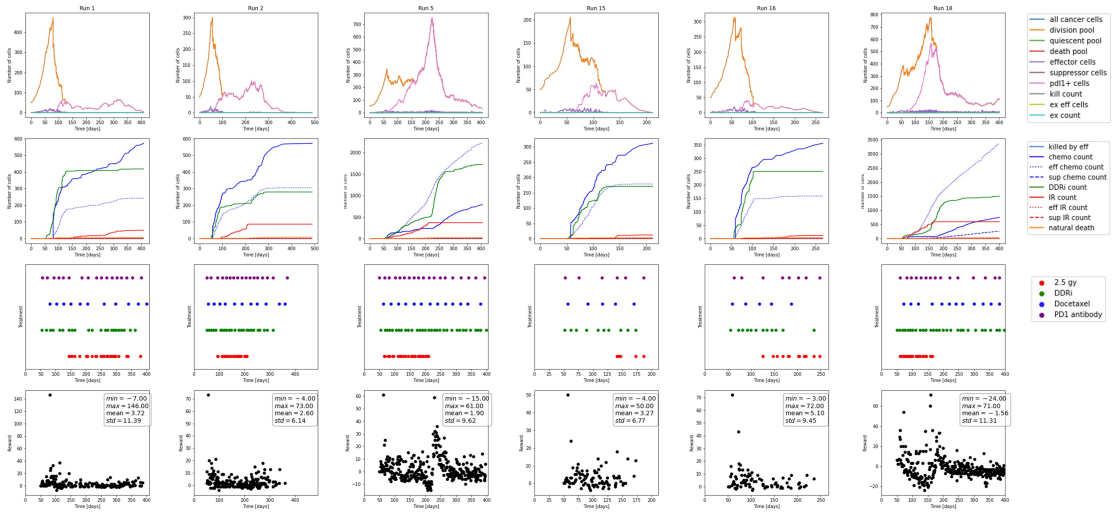


Figure 6.2: Training runs 1-18 for a treatment starting at day 50

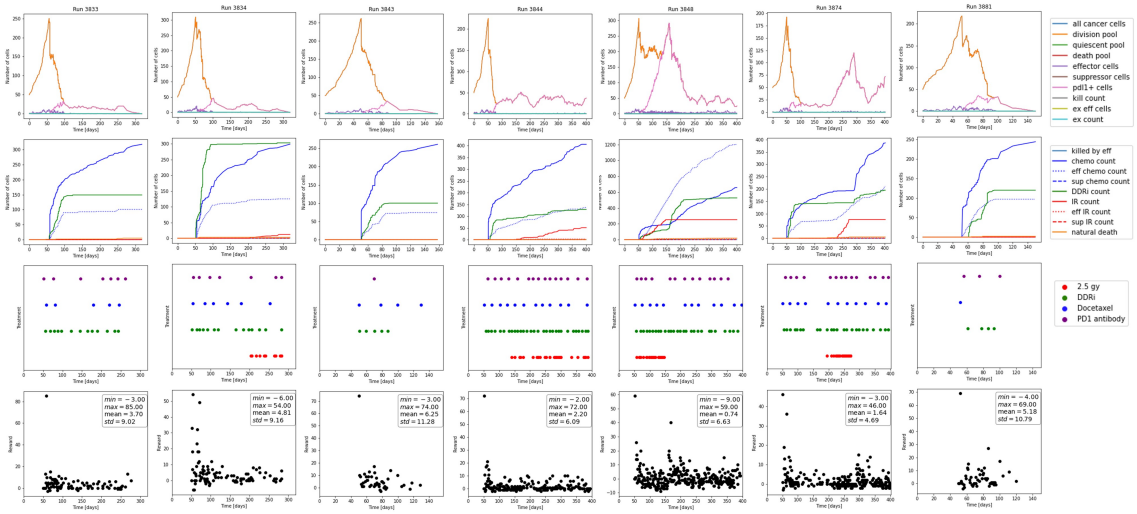


Figure 6.3: Training runs 3833-3881 for a treatment starting at day 50

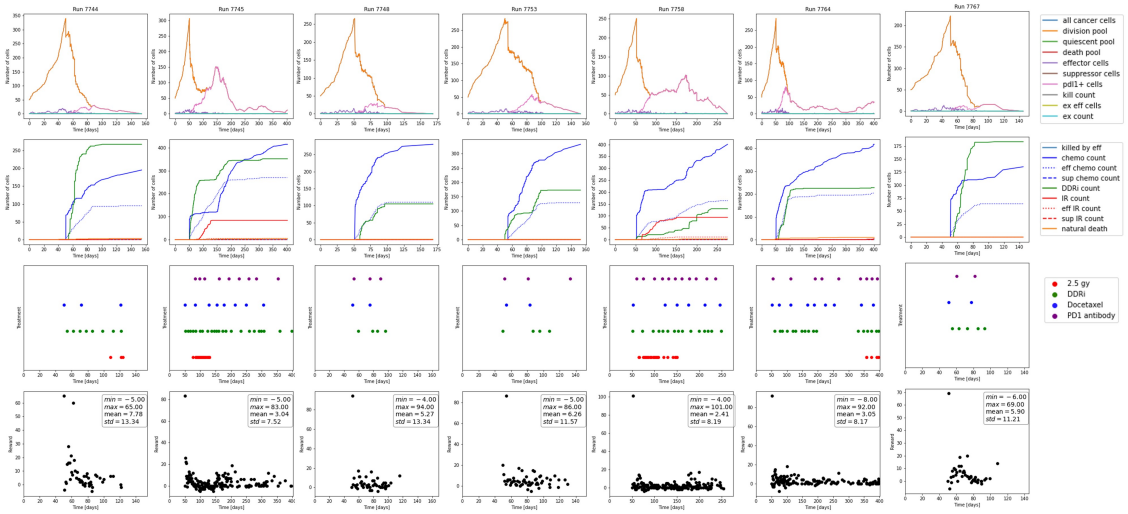


Figure 6.4: Training runs 7744-7767 for a treatment starting at day 50

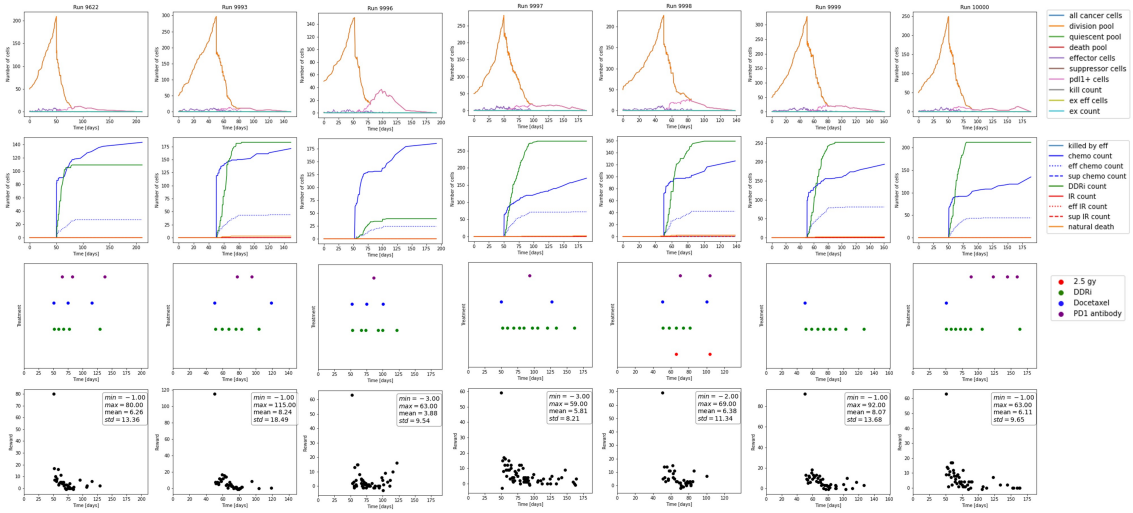


Figure 6.5: Training runs 9622-10000 for a treatment starting at day 50

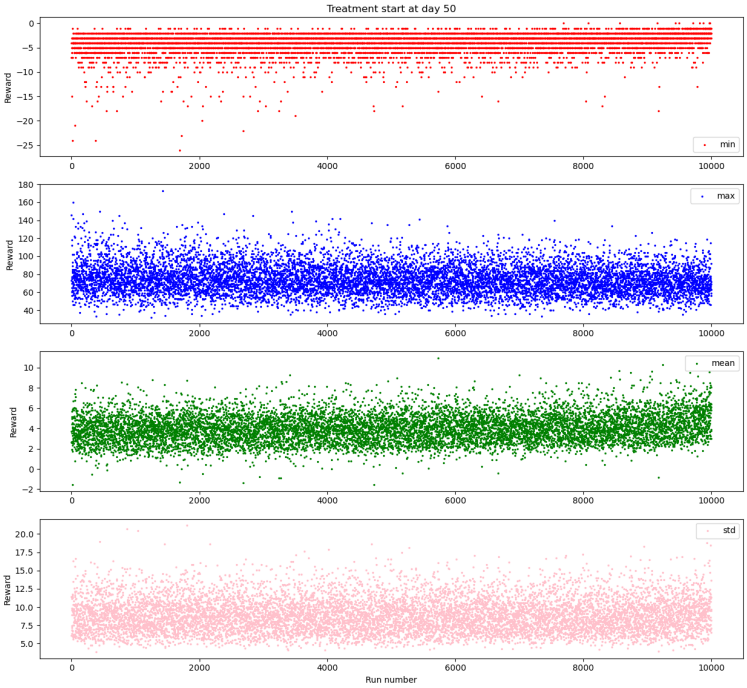


Figure 6.6: Summary of the given rewards for run 1-10000 with a treatment starting at day 50

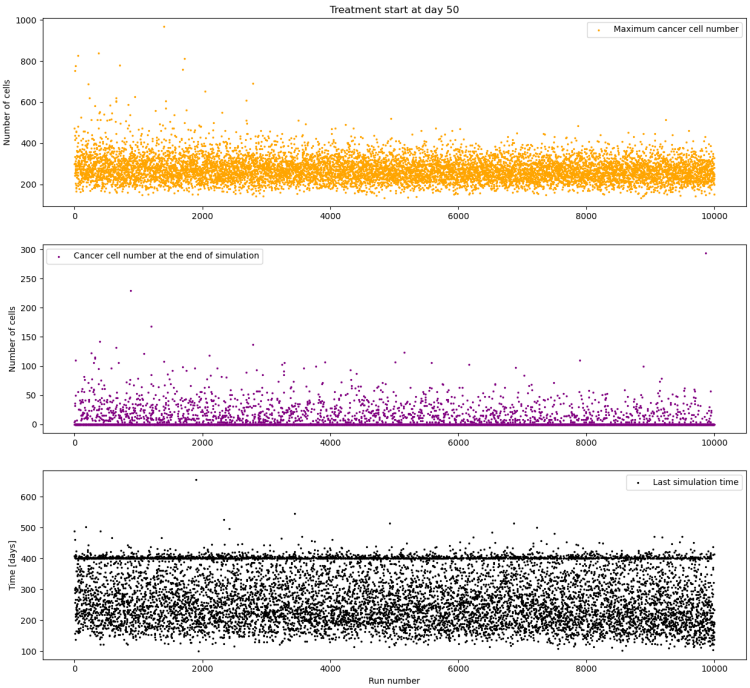


Figure 6.7: Summary of the cancer cell numbers and simulation time for run 1-10000 with a treatment starting at day 50

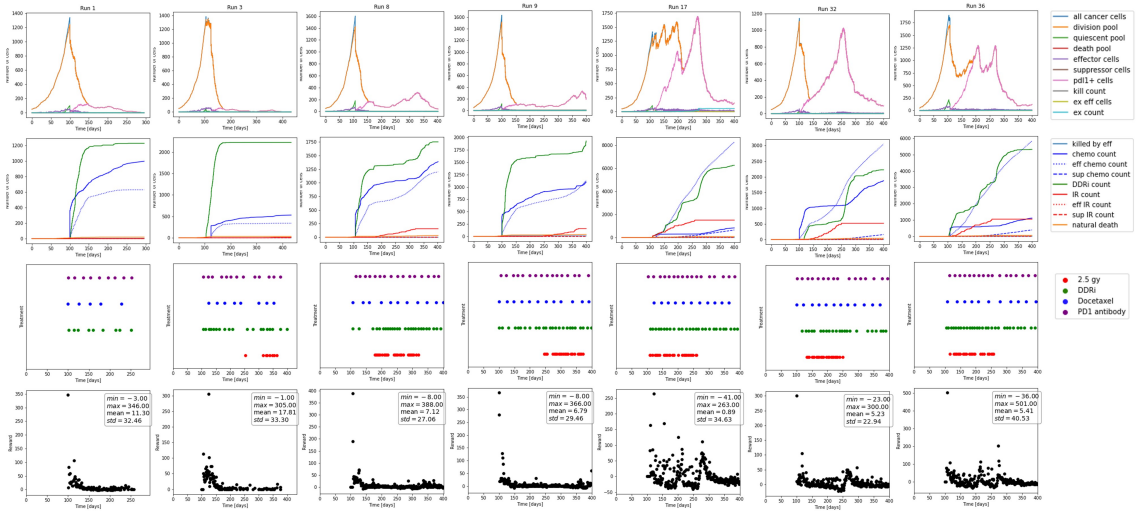


Figure 6.8: Training runs 1-36 for a treatment starting at day 100

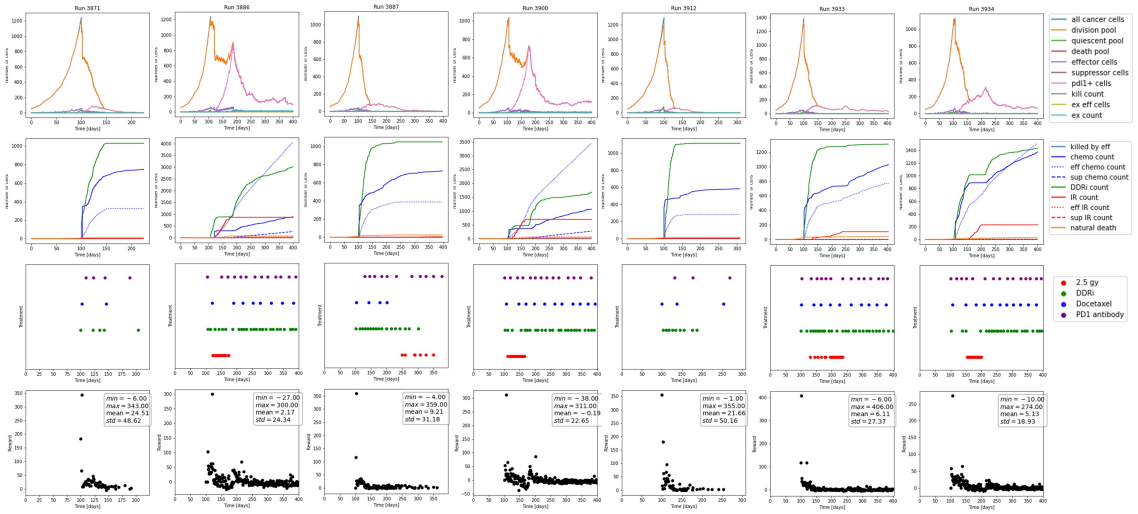


Figure 6.9: Training runs 3871-3934 for a treatment starting at day 100

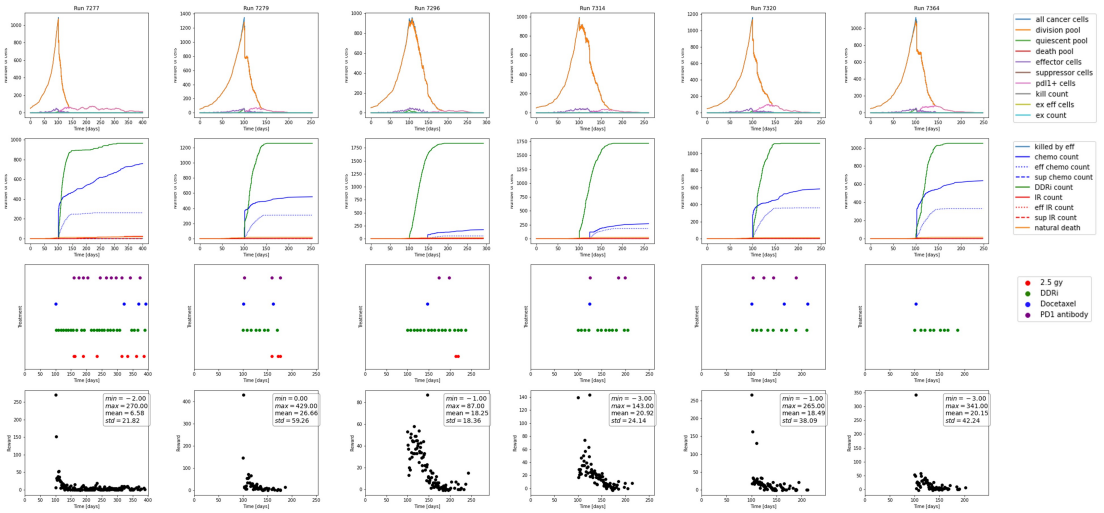


Figure 6.10: Training runs 7277-7380 for a treatment starting at day 100

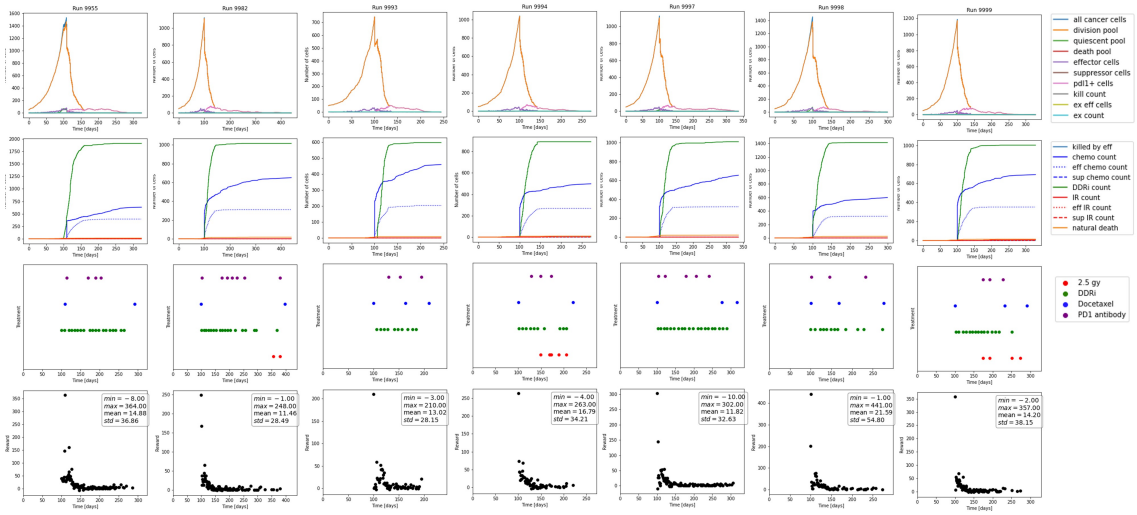


Figure 6.11: Training runs 9955-9999 for a treatment starting at day 100

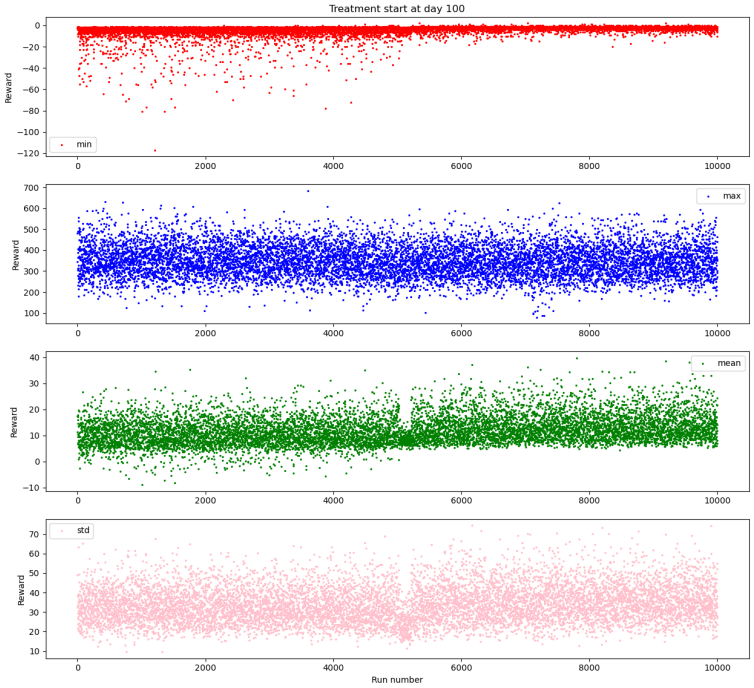


Figure 6.12: Summary of the given rewards for run 1-10000 with a treatment starting at day 100

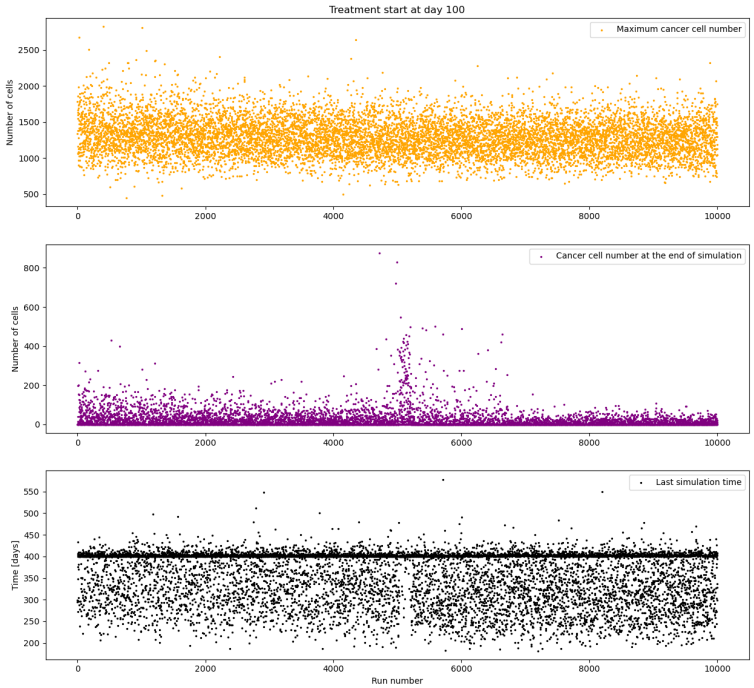


Figure 6.13: Summary of the cancer cell numbers and simulation time for run 1-10000 with a treatment starting at day 100

cancer cell number at approximately day 150 and in the run 7758 radiotherapy causes an increase in the cancer cell number and a peak around day 170. Radiotherapy treatment failure can have multiple reasons: Due to the dose limitations and tumour size, there are cells left after the end of radiotherapy so that the tumour can regrow (Recht et al. 1988). In addition, radiotherapy causes PDL1+ mutation and eliminates immune effector cells (Sato et al. 2019). This can cause a disbalance between cancer cell elimination by the immune system due to the reduced number of immune effector cells and effector cell exhaustion. Further, the cancer repopulation rate could be an important factor as the number of infiltrated immune effector cells after radiation might not be large enough to eradicate a large tumour (Begg 2012). In the later runs from 9622-10000 with an epsilon of 0.96-1, we can see that for the given tumour parameters a combination of chemotherapy, DNA damage response inhibitor and PD1 antibody is preferred. Each run is different due to the stochasticity of the agent-based model, but a trend can be observed. Chemotherapy is given first to decrease the cancer cell number, DNA damage response inhibitor is given as a maintenance therapy and PD1 antibody is given from around day 80. In one case radiotherapy is given as two isolated doses to decrease the cancer cell count. Looking at all plots for the treatment start at day 50 in the second row, we can see that in successful cases with tumour cell extinction, either chemotherapy or DNA damage response inhibitor therapy are the main driver of cell death. When reviewing the summary plots in Figure 6.6 and 6.7, it shows that in the course of the simulations, the minimum and mean rewards increases while the maximum rewards decreases. The maximum cancer cell number, cancer cell number at the end of simulation and simulation time decreases.

When treatment starts at day 100, more immune cells have infiltrated the tumour micro environment. Looking through the plots (Figures 6.8-6.11), we can see similar patterns as in the case of an earlier treatment at day 50. Here, an early start of radiation is also disadvantageous. As the dose of 2.5 Gy is not high enough to kill all the cancer cells, the consequences are a regrowth and PDL1 mutation (see runs 8, 17, 32, 36 in Figure 6.8, and runs 3886, 3900, 3934 in Figure 6.9). Similarly, to the treatment start at day 50, it can be observed that radiation is more advantageous as isolated doses when the tumour cell count is already low (see Figure 6.10 runs 7279, 7296, Figure 6.11 runs 9982, 9994, 9999). Another similarity is that in the successful treatment schedules, chemotherapy is given at treatment start to decrease the cell count and DNA damage response inhibitor is used as a maintenance therapy. Differences are that PD1 antibody is given in almost all

successful treatment schedules at treatment start and that the main driver of cancer cell elimination is the DNA damage response inhibitor (see second rows in Figures 6.8-6.11). Those simulations also show that, while radiotherapy increases the immune cell infiltration, in combination with chemotherapy it causes toxicity and immune cell kill can be higher than the cancer cell elimination (see Figure 6.8 runs 17, 32, 36. Figure 6.9 runs 3886, 3900, 3934). When reviewing the summary plots in Figure 6.12 and 6.13, we can see similar patterns as in the treatment case starting at day 50. Since the number of cancer cells are larger than in the other treatment case, the numbers for the rewards and cells are higher. But in the course of the simulations a similar pattern can be observed: the minimum and mean rewards increase while the maximum rewards decrease. The maximum cancer cell number, cancer cell number at the end of simulation and simulation time decreases.

The Q table reflects the rewards for the actions in each state. Advantageous actions will have a high Q value and hence are more likely to be chosen when the algorithm is in the exploitation phase where it choose its actions based on the learned experience from the environment.

6.4 Discussion

This is the first report describing the integration of an agent based system pharmacology model with reinforcement learning methods. As such, it provides a novel approach for optimizing treatment combinations based on a combination of tumour and micro environment characteristics, therapeutic profiles, and mechanisms of action. This approach allows for the determination of optimal treatment regimens that balance efficacy and toxicity, taking into account various factors that impact treatment outcomes.

One notable strength of this approach is its flexibility. The ability to extend the model to accommodate different doses for each treatment type or a variety of therapeutic interventions enhances its applicability across diverse clinical scenarios. Additionally, the adaptability of the model to account for variations in tumour immune composition, such as 'cold' or 'hot' tumours, further enhances its potential clinical relevance. Further PK parameters could be adjusted to specific patient populations. This adaptability underscores the importance of tailoring treatments to immunological profiles and interaction of combined drugs.

This approach shows that there are multiple solutions to treatment optimization,

and that multiple treatment regimens can be optimal and also differ substantially among themselves. Solutions to the therapeutic regimen optimization problem can be found in a reasonable amount of time and require a relatively small number of assumptions. To train the algorithm, a small positive number is set as a threshold and the algorithm is repeated until the Q value for the states is stable and the difference of the old and updated value is smaller than that small threshold (see section 2.3.7). In our example, we run the algorithm with 10000 repetitions each to find the solutions for two cases. In the first case the treatment is started at day 50 mimicking an early tumour detection treatment paradigm with a low immune cell infiltration into the tumour micro environment. We can see that for the given tumour parameters the algorithm favours a combination of chemotherapy, DNA damage response inhibitor and PD1 antibody. Chemotherapy is given first to decrease the cancer cell number, DNA damage response inhibitor is administered as a maintenance therapy and PD1 antibody is given from around day 80 when more immune cells have infiltrated the tumour micro environment. Looking through the simulations, it can be seen that either chemotherapy or DNA damage response inhibitor therapy are the main driver of cell death. In the second case we have simulated treatment starting at day 100 where a more developed tumour micro environment with immune cell infiltration exist. The algorithm favours a similar treatment pattern as in the first case. Chemotherapy is given at treatment start to decrease the cell count and DNA damage response inhibitor is used as a maintenance therapy. Differences are that PD1 antibody is given at treatment start to maintain immune function as immune cells have already infiltrated the tumour micro environment. An additional difference is that the main driver of cancer cell elimination is the DNA damage response inhibitor.

A limitation of this method is the speed of execution. With an increased number of actions (for example multiple treatment options or multiple dosing regimen for a treatment option) or an increased number of states (for example considering multiple attributes for the tumour population such as location, vasculature, different mutations) the Q table will get large and training time will increase. Depending on available computational resources, a range of actions and states has to be considered. To increase the speed of execution, the training runs could be paralleled while updating the same Q table.

With future improvements in imaging tumours and their micro environment this model can aid in treatment individualisation where the algorithm finds the best treatment combination and schedule a given patient. Importantly, this work challenges the com-

mon perception that artificial intelligence techniques exclusively rely on large datasets for training. While artificial intelligence techniques often require large data sets, for example observational data was used by Moodie et al for their Q-learning algorithm (Moodie et al. 2012) or Krakow and colleagues used cohort and registry data (Krakow et al. 2017), we have shown that reinforcement learning can also be combined with a mechanistic model. The approach generates its own data, enabling the algorithm to learn and make recommendations. This mechanism-driven data generation not only enhances interpretability but also creates a feedback loop wherein treatment recommendations can be simulated within the mechanistic model, fostering a deeper understanding of the intricate interplay between different components. This provides mechanistic insight into the AI ML black box.

Chapter 7

Concluding remarks

In this thesis, mathematical and computational methods have been used to study cancer development, their characteristics, interactions with the tumour micro environment and effects of mono and combination therapy. The interplay between those different components are important to understand for a successful treatment outcome. Cancer is a complex and heterogeneous disease, encompassing multiple scales of biological organization, ranging from intracellular molecular pathways to interactions between cells and organ and tissues scales. Therefore, mathematical models can help to understand the disease mechanism to design better treatment schedules and to avoid treatment failure or toxicity. However, currently widely used models in pharmacometrics are based on ODEs. Those often assume a homogeneous cell population and model the time-course of biological measurements continuously. Crucial aspects of cancer, such as its heterogeneity, spatial characteristics, and the presence of random events like mutations or emergent behavior are difficult to capture with ODEs. Furthermore, a large number of parameters and compartments are required to model the interactions between cancer cells and the TME, including immune cells in space. To address these limitations, this thesis has introduced alternative approaches for modelling the complexity of tumours and their response to drug treatments and combined them with existing PKPD and PBPK frameworks.

Therefore, in chapter three, agent based modelling was introduced with a case study focused on MEK inhibitors within the Ras-Raf-MEK-ERK pathway. A PKPD model for the MEK inhibitor cobimetinib was implemented. The treatment of cancer cells using both ODE and ABM approaches was simulated. Single and multiple dose treatments were investigated, individual cell fate, their elimination by the drug, division, and muta-

tion within the population was analysed. The simulation results and the comparison of the two models showed us their differences, strengths, and weaknesses. We learned that ODE models provide insights at the population level, and are well-suited for simulating processes that can be approximated as homogeneous, well-mixed systems with central coordination. ABMs on the other hand offer a deeper understanding of cellular-level behavior and the rules driving it. ABMs are suitable for simulating complex biological systems with heterogeneity, sub-scale components (molecular, cellular, tissue, organism) and inherent emerging behaviour. To take advantage of both model types, we can combine them in hybrid multi scale models, where parts of the model are discretised by using a ABM to study heterogeneity and emergent behaviour. This could be cancer cells that are heterogeneous in their response to treatment, their location, or their mutation status. ODEs and PDEs can be applied to simulate parts of the model that are assumed to be homogeneous without the need for individual representation of entities. Those could be concentration time profiles of drugs or nutrients and their distribution in space.

Chapter four is centered around stochastic equations as a bridge connecting ODE models and ABMs. A simplification of the model in the previous chapter was used to develop a stochastic description of a single cell's fate under continuous drug treatment. That approach was extended to model a population of heterogeneous cells. We analysed the survival probability of each cell, the time to population extinction and its relation with population size. Additionally, multiple dose treatments considering cell recovery and division between cycles were investigated to identify critical division and death rates for uncontrolled growth and successful treatment. This chapter shows us that while stochastic equations lack detailed insight into component interactions, they offer simplified mathematical expressions to summarise dynamics. This can be used to run fast analysis to determine whether a more detailed model such as an agent based model is needed. Further, it could help to run more targeted agent based simulations to gain insight into the situations of interest.

Based on the comparison in chapter three, we developed a hybrid model that combines the advantages of ABMs, ODEs, and PDEs to study the effects of mono and combination treatments on tumour-immune interactions in chapter five. The ABM is used to simulate tumour and immune cells incorporating rules governing their interaction, emergence of behavior, and spatial heterogeneity. Each cancer cell has its cell cycle phase and the different sensitivity to treatment included. For example, docetaxel specifically targets the

G2 and M phases. The most radiosensitive phases are the S, G2, and M phases, whereas the resting phase G0 is less affected by radiotherapy. Additionally, the model incorporates the higher radiosensitivity of immune effector cells in comparison to immune suppressor cells. Effects of different treatment regimen are considered. For example, radiotherapy has not only an effect on the tumour cell population but also changes the behaviour of immune cells. After radiation the infiltration of immune cells increases while also PDL1 mutation in cancer cells are promoted, which subsequently causes immune effector exhaustion. Further treatment options and environmental conditions are explored with ODEs or PDEs: A PBPK model gives insights into PD1 antibody treatment, accounting for drug distribution through different physiological compartments. An ODE based KPD and PKPD model simulates chemotherapy and DNA damage response inhibitors. To consider diffusion and the concentration level of oxygen and drug in the 3D space, we incorporated a PDE. The oxygen gradients drive tumour cell behavior, such as active division, quiescence, and the presence of a necrotic core. Additionally, oxygen supply impacts radiation treatment, which is modelled with a modified linear square model. By incorporating these mechanism into our hybrid model, we are able to gain a detailed understanding of the complex interactions between tumours and the immune system. We have analyzed the effects of different treatments on the interactions between various cell types. As a result, we gained insight in the improved effect and and antagonism when multiple treatment modalities are administered within a complex system. This enables us to improve the combination and scheduling of different therapies. Additionally, we used PCA and the kmeans algorithm to analyse the behaviour of cancer and immune effector cells. The result confirmed that ABM is a suitable method to simulate the complex tumour immune interaction in our model. Due to the large number of subgroups in the cancer and immune effector cell population and their attributes, a large system of ODEs would be required whereas the ABM can capture those dynamics with a smaller set of rules and attributes.

To automatise the optimization of combination treatment schedules we explored the application of reinforcement learning in chapter six. Q learning was employed, where a learning algorithm has the goal to find the best treatment policy by collecting rewards, avoiding punishments and considering the long term outcome. Rewards were defined as the number eliminated cancer cells, while punishments were applied when immune cells were eliminated. Hence, constraints such as toxicity levels and treatment schedule re-

strictions were taken into account during the optimization process. Decisions were taken considering the amount of infiltrated immune cells and the presence of cancer cells in specific cell cycle phases. Various treatment options and combinations are possible actions for the algorithm to choose from. The algorithm explores the environment by randomly choosing actions during the learning phase to collect information about the effect of different therapies and their combination on the tumour immune interaction. This allows the identification of optimal treatment schedules that balance objectives of eliminating cancer cells while minimizing harm to the immune system. The combination of the reinforcement learning algorithm with a mechanistic model consisting of an ABM and drug treatment models based on PKPD, PBPK, ODE and PDE models allows us to gain insight into the interaction of cells and the mechanism of the combination treatment. By feeding the results of the reinforcement learning algorithm back into the model, we can understand the mechanism for a successful treatment combination. This approach also offers the flexibility of extending the model with more treatment options or adjusting the parameters for a given cancer or patient. This chapter shows that through the use of reinforcement learning, we can harness the potential of artificial intelligence to enhance cancer treatment decision-making and improve patient outcomes.

In summary, in this thesis alternative modelling strategies are explored to enhance our understanding of tumour behavior and improve the design and optimization of cancer treatments as mono and combination therapy. It shows that stochastic equations, ABMs and artificial intelligence algorithms such as reinforcement learning can complement currently-used approaches to provide a more detailed insight into disease mechanisms and treatment combinations.

References

Alfonso, Juan CL, Grass, G Daniel, Welsh, Eric, Ahmed, Kamran A, Teer, Jamie K, Pilon-Thomas, Shari, Harrison, Louis B, Cleveland, John L, Mulé, James J, Eschrich, Steven A et al. (2021). "Tumor-immune ecosystem dynamics define an individual Radiation Immune Score to predict pan-cancer radiocurability". In: *Neoplasia* 23.11, pp. 1110–1122.

Allen, Linda (2007). "An introduction to mathematical biology". In.

Allen, Linda JS (2010). *An introduction to stochastic processes with applications to biology*. CRC Press.

An, G, Fitzpatrick, BG, Christley, S, Federico, P, Kanarek, A, Neilan, R Miller, Oremland, M, Salinas, R, Laubenbacher, Reinhard and Lenhart, S (2017). "Optimization and control of agent-based models in biology: a perspective". In: *Bulletin of mathematical biology* 79, pp. 63–87.

An, Gary, Mi, Qi, Dutta-Moscato, Joyeeta and Vodovotz, Yoram (2009). "Agent-based models in translational systems biology". In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1.2, pp. 159–171.

Anderson, Philip W (1972). "More Is Different: Broken symmetry and the nature of the hierarchical structure of science." In: *Science* 177.4047, pp. 393–396.

Arulraj, Theinmozhi, Wang, Hanwen, Emens, Leisha A, Santa-Maria, Cesar A and Popel, Aleksander S (2023). "A transcriptome-informed QSP model of metastatic triple-negative breast cancer identifies predictive biomarkers for PD-1 inhibition". In: *Science Advances* 9.26, eadg0289.

Aykul, Senem and Martinez-Hackert, Erik (2016). "Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis". In: *Analytical biochemistry* 508, pp. 97–103.

Baggio, Jacopo A and Baggio, Rodolfo (2020). *Modelling and Simulations for Tourism and Hospitality: An Introduction*. Channel View Publications.

Bauer, Amy L, Beauchemin, Catherine AA and Perelson, Alan S (2009). "Agent-based modeling of host–pathogen systems: The successes and challenges". In: *Information sciences* 179.10, pp. 1379–1389.

Begg, Adrian C (2012). "Predicting recurrence after radiotherapy in head and neck cancer". In: *Seminars in radiation oncology*. Vol. 22. 2. Elsevier, pp. 108–118.

Bianchi, Federico and Squazzoni, Flaminio (2015). "Agent-based models in sociology". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 7.4, pp. 284–306.

Bonabeau, Eric (2002). "Agent-based modeling: Methods and techniques for simulating human systems". In: *Proceedings of the national academy of sciences* 99.suppl 3, pp. 7280–7287.

Bonato, Vinicius, Tang, Szu-Yu, Hsieh, Matilda, Zhang, Yao and Deng, Shibing (2024). "Experimental design considerations and statistical analyses in preclinical tumor growth inhibition studies". In: *Pharmaceutical Statistics*.

Brunton, GF and Wheldon, TE (1980). "The Gompertz equation and the construction of tumour growth curves". In: *Cell Proliferation* 13.4, pp. 455–460.

Buchanan, A (1847). "Physiologic effects of the inhalation of ether". In: *London Med Gaz* 39, pp. 715–717.

Cess, Colin G and Finley, Stacey D (2020). "Multi-scale modeling of macrophage—T cell interactions within the tumor microenvironment". In: *PLOS Computational Biology* 16.12, e1008519.

Chaplain, Mark AJ and Powathil, Gibin G (2016). "Multiscale modelling of Cancer progression and treatment control: the role of intracellular heterogeneities in chemotherapy

treatment". In: *Research on the PHYSICS OF CANCER: A Global Perspective*. World Scientific, pp. 1–18.

Chatfield, Chris (2018). *Introduction to multivariate analysis*. Routledge.

Checkley, Stephen, MacCallum, Linda, Yates, James, Jasper, Paul, Luo, Haobin, Tolsma, John and Bendtsen, Claus (2015). "Bridging the gap between in vitro and in vivo: dose and schedule predictions for the ATR inhibitor AZD6738". In: *Scientific reports* 5.1, pp. 1–12.

Cockrell, Chase and Axelrod, David E (2019). "Optimization of dose schedules for chemotherapy of early colon cancer determined by high-performance computer simulations". In: *Cancer informatics* 18, p. 1176935118822804.

Cockrell, Robert Chase and An, Gary (2018). "Examining the controllability of sepsis using genetic algorithms on an agent-based model of systemic inflammation". In: *PLoS computational biology* 14.2, e1005876.

Coletti, Roberta, Leonardelli, Lorena, Parolo, Silvia and Marchetti, Luca (2020). "A QSP model of prostate cancer immunotherapy to identify effective combination therapies". In: *Scientific reports* 10.1, p. 9063.

Cooper, Geoffrey M. (2000). *The cell : a molecular approach*. eng. 2nd ed. Washington, D.C: ASM Press. ISBN: 0-87893-106-6.

Cosgrove, J, Butler, J, Alden, K, Read, M, Kumar, V, Cucurull-Sanchez, L, Timmis, J and Coles, M (2015). "Agent-based modeling in systems pharmacology". In: *CPT: pharmacometrics & systems pharmacology* 4.11, pp. 615–629.

Dagogo-Jack, Ibiayi and Shaw, Alice T (2018). "Tumour heterogeneity and resistance to cancer therapies". In: *Nature reviews Clinical oncology* 15.2, pp. 81–94.

Dartois, Céline, Brendel, Karl, Comets, Emmanuelle, Laffont, CM, Laveille, Christian, Tranchand, Brigitte, Mentré, France, Lemenuel-Diot, Annabelle and Girard, Pascal (2007). "Overview of model-building strategies in population PK/PD analyses: 2002–2004 literature survey". In: *British journal of clinical pharmacology* 64.5, pp. 603–612.

Del Monte, Ugo (2009). "Does the cell number 109 still really fit one gram of tumor tissue?" In: *Cell cycle* 8.3, pp. 505–506.

Delgado-SanMartin, Juan A, Hare, Jennifer I, Davies, EJ and Yates, James WT (2017). "Multiscalar cellular automaton simulates in-vivo tumour-stroma patterns calibrated from in-vitro assay data". In: *BMC medical informatics and decision making* 17.1, pp. 1–12.

Dewachter, Liselot, Fauvart, Maarten and Michiels, Jan (2019). "Bacterial heterogeneity and antibiotic survival: understanding and combatting persistence and heteroresistance". In: *Molecular cell* 76.2, pp. 255–267.

Dowling, Mark R, Kan, Andrey, Heinzel, Susanne, Marchingo, Julia M, Hodgkin, Philip D and Hawkins, Edwin D (2018). "Regulatory T cells suppress effector T cell proliferation by limiting division destiny". In: *Frontiers in immunology* 9, p. 2461.

Elassaiss-Schaap, J, Rossenu, S, Lindauer, A, Kang, SP, De Greef, R, Sachs, JR and De Alwis, DP (2017). "Using model-based "learn and confirm" to reveal the pharmacokinetics-pharmacodynamics relationship of pembrolizumab in the KEYNOTE-001 Trial". In: *CPT: pharmacometrics & systems pharmacology* 6.1, pp. 21–28.

Erdi, Yusuf Emre (2012). "Limits of tumor detectability in nuclear medicine and PET". In: *Molecular imaging and radionuclide therapy* 21.1, p. 23.

Escandell-Montero, Pablo, Chermisi, Milena, Martinez-Martinez, Jose M, Gomez-Sanchis, Juan, Barbieri, Carlo, Soria-Olivas, Emilio, Mari, Flavio, Vila-Francés, Joan, Stopper, Andrea, Gatti, Emanuele et al. (2014). "Optimization of anemia treatment in hemodialysis patients via reinforcement learning". In: *Artificial intelligence in medicine* 62.1, pp. 47–60.

Ette, Ene I and Williams, Paul J (2007). *Pharmacometrics: the science of quantitative pharmacology*. John Wiley & Sons.

Fabian, Kellsye P, Wolfson, Benjamin and Hodge, James W (2021). "From Immunogenic Cell Death to Immunogenic Modulation: Select Chemotherapy Regimens Induce a Spectrum of Immune-Enhancing Activities in the Tumor Microenvironment". In: *Frontiers in Oncology* 11.

Felten, Renaud, Scherlinger, Marc, Mertz, Philippe, Chasset, François and Arnaud, Laurent (2023). "New biologics and targeted therapies in systemic lupus: from new molecular targets to new indications. A systematic review". In: *Joint Bone Spine*, p. 105523.

Figueredo, Grazziela P, Siebers, Peer-Olaf, Owen, Markus R, Reps, Jenna and Aickelin, Uwe (2014). "Comparing stochastic differential equations and agent-based modelling and simulation for early-stage cancer". In: *PloS one* 9.4.

Fisher, Rosie, Pusztai, Lazos and Swanton, C (2013). "Cancer heterogeneity: implications for targeted therapeutics". In: *British journal of cancer* 108.3, pp. 479–485.

Frances, Nicolas, Claret, Laurent, Bruno, Rene and Iliadis, Athanassios (2011). "Tumor growth modeling from clinical trials reveals synergistic anticancer effect of the capecitabine and docetaxel combination in metastatic breast cancer". In: *Cancer chemotherapy and pharmacology* 68.6, pp. 1413–1419.

Fu, Jin, Wu, Sheng and Petzold, Linda R (2013). "Time dependent solution for acceleration of tau-leaping". In: *Journal of Computational Physics* 235, pp. 446–457.

Gabrilovich, Dmitry I and Nagaraj, Srinivas (2009). "Myeloid-derived suppressor cells as regulators of the immune system". In: *Nature reviews immunology* 9.3, pp. 162–174.

Galluzzi, Lorenzo, Aryankalayil, Molykutty J, Coleman, C Norman and Formenti, Silvia C (2023). "Emerging evidence for adapting radiotherapy to immunotherapy". In: *Nature Reviews Clinical Oncology*, pp. 1–15.

Gaweda, Adam E, Muezzinoglu, Mehmet K, Aronoff, George R, Jacobs, Alfred A, Zurada, Jacek M and Brier, Michael E (2005). "Individualization of pharmacological anemia management using reinforcement learning". In: *Neural Networks* 18.5-6, pp. 826–834.

Gefen, Orit and Balaban, Nathalie Q (2009). "The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress". In: *FEMS microbiology reviews* 33.4, pp. 704–717.

Gilbert, Nigel (2019). *Agent-based models*. Vol. 153. Sage Publications, Incorporated.

Gladtke, E (1985). "LaudatioProfessor Dr. Dr. Friedrich Hartmut Dost 75 Jahre". In: *Klinische Pädiatrie* 197.04, pp. 375–375.

Gobburu, Jogarao VS (2010). “Pharmacometrics 2020”. In: *The Journal of Clinical Pharmacology* 50.S9, 151S–157S.

Gong, Chang, Anders, Robert A, Zhu, Qingfeng, Taube, Janis M, Green, Benjamin, Cheng, Wenting, Bartelink, Imke H, Vicini, Paolo, Wang, Bing and Popel, Aleksander S (2019). “Quantitative characterization of CD8+ T cell clustering and spatial heterogeneity in solid tumors”. In: *Frontiers in oncology* 8, p. 649.

Gong, Chang, Milberg, Oleg, Wang, Bing, Vicini, Paolo, Narwal, Rajesh, Roskos, Lorin and Popel, Aleksander S (2017). “A computational multiscale agent-based model for simulating spatio-temporal tumour immune response to PD1 and PDL1 inhibition”. In: *Journal of the Royal Society Interface* 14.134, p. 20170320.

Graaf, Piet H van der (2012). *CPT: pharmacometrics and systems pharmacology.*

Grassberger, Clemens, Ellsworth, Susannah G, Wilks, Moses Q, Keane, Florence K and Loeffler, Jay S (2019). “Assessing the interactions between radiotherapy and antitumour immunity”. In: *Nature reviews Clinical oncology* 16.12, pp. 729–745.

Gumbel, Emil Julius (1935). “Les valeurs extrêmes des distributions statistiques”. In: 5.2, pp. 115–158.

Han, Yanyan, Liu, Dandan and Li, Lianhong (2020). “PD-1/PD-L1 pathway: current researches in cancer”. In: *American journal of cancer research* 10.3, p. 727.

Hanahan, Douglas and Weinberg, Robert A (2000). “The hallmarks of cancer”. In: *cell* 100.1, pp. 57–70.

– (2011). “Hallmarks of cancer: the next generation”. In: *cell* 144.5, pp. 646–674.

Helmlinger, Gabriel, Sokolov, Victor, Peskov, Kirill, Hallow, Karen M, Kosinsky, Yuri, Voronova, Veronika, Chu, Lulu, Yakovleva, Tatiana, Azarov, Ivan, Kaschek, Daniel et al. (2019). “Quantitative systems pharmacology: an exemplar model-building workflow with applications in cardiovascular, metabolic, and oncology drug development”. In: *CPT: pharmacometrics & systems pharmacology* 8.6, pp. 380–395.

Heppenstall, Alison J, Crooks, Andrew T, See, Linda M and Batty, Michael (2011). *Agent-based models of geographical systems.* Springer Science & Business Media.

Holford, Nicholas HG and Sheiner, Lewis B (1982). "Kinetics of pharmacologic response". In: *Pharmacology & therapeutics* 16.2, pp. 143–166.

Holford, Nick (2013). "A time to event tutorial for pharmacometricians". In: *CPT: pharmacometrics & systems pharmacology* 2.5, pp. 1–8.

Jalalimanesh, Ammar, Haghighi, Hamidreza Shahabi, Ahmadi, Abbas, Hejazian, Hossein and Soltani, Madjid (2017). "Multi-objective optimization of radiotherapy: distributed Q-learning and agent-based simulation". In: *Journal of Experimental & Theoretical artificial intelligence* 29.5, pp. 1071–1086.

Jones, HM and Rowland-Yeo, Karen (2013). "Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development". In: *CPT: pharmacometrics & systems pharmacology* 2.8, pp. 1–12.

Kather, Jakob Nikolas, Charoentong, Pornpimol, Suarez-Carmona, Meggy, Herpel, Esther, Klupp, Fee, Ulrich, Alexis, Schneider, Martin, Zoernig, Inka, Luedde, Tom, Jaeger, Dirk et al. (2018). "High-throughput screening of combinatorial immunotherapies with patient-specific in silico models of metastatic colorectal cancer". In: *Cancer Research* 78.17, pp. 5155–5163.

Kather, Jakob Nikolas, Poleszczuk, Jan, Suarez-Carmona, Meggy, Krisam, Johannes, Charoentong, Pornpimol, Valous, Nektarios A, Weis, Cleo-Aron, Tavernar, Luca, Leiss, Florian, Herpel, Esther et al. (2017). "In silico modeling of immunotherapy and stroma-targeting therapies in human colorectal cancer". In: *Cancer research* 77.22, pp. 6442–6452.

Kitano, Atsuko, Ono, Makiko, Yoshida, Masayuki, Noguchi, Emi, Shimomura, Akihiko, Shimoi, Tatsunori, Kodaira, Makoto, Yunokawa, Mayu, Yonemori, Kan, Shimizu, Chikako et al. (2017). "Tumour-infiltrating lymphocytes are correlated with higher expression levels of PD-1 and PD-L1 in early breast cancer". In: *ESMO open* 2.2, e000150.

Knight-Schrijver, VR, Chelliah, V, Cucurull-Sanchez, L and Le Novère, N (2016). "The promises of quantitative systems pharmacology modelling for drug development". In: *Computational and structural biotechnology journal* 14, pp. 363–370.

Kotz, Samuel and Nadarajah, Saralees (2000). *Extreme value distributions: theory and applications*. world scientific.

Krakow, Elizabeth F, Hemmer, Michael, Wang, Tao, Logan, Brent, Arora, Mukta, Spellman, Stephen, Couriel, Daniel, Alousi, Amin, Pidala, Joseph, Last, Michael et al. (2017). "Tools for the precision medicine era: how to develop highly personalized treatment recommendations from cohort and registry data using Q-learning". In: *American journal of epidemiology* 186.2, pp. 160–172.

Kun, E, Tsang, YTM, Ng, CW, Gershenson, DM and Wong, KK (2021). "MEK inhibitor resistance mechanisms and recent developments in combination trials". In: *Cancer treatment reviews* 92, p. 102137.

Lamperti, Francesco, Roventini, Andrea and Sani, Amir (2018). "Agent-based model calibration using machine learning surrogates". In: *Journal of Economic Dynamics and Control* 90, pp. 366–389.

Lee, Youjin, Auh, Sogyong L, Wang, Yugang, Burnette, Byron, Wang, Yang, Meng, Yuru, Beckett, Michael, Sharma, Rohit, Chin, Robert, Tu, Tony et al. (2009). "Therapeutic effects of ablative radiation on local tumor require CD8+ T cells: changing strategies for cancer treatment". In: *Blood, The Journal of the American Society of Hematology* 114.3, pp. 589–595.

Leibold, Adam T, Monaco, Gina N and Dey, Mahua (2019). "The role of the immune system in brain metastasis". In: *Current neurobiology* 10.2, p. 33.

Lemaire, Vincent, Bassen, David, Reed, Mike, Song, Roy, Khalili, Samira, Lien, Yi Ting, Huang, Lu, Singh, Aman P, Stamatelos, Spyros, Bottino, Dean et al. (2023). "From cold to hot: changing perceptions and future opportunities for quantitative systems pharmacology modeling in cancer immunotherapy". In: *Clinical Pharmacology & Therapeutics* 113.5, pp. 963–972.

Levy, Gerhard and Gibaldi, Milo (1972). "Pharmacokinetics of drug action". In: *Annual Review of Pharmacology* 12.1, pp. 85–98.

Li, Yanan, Dong, Qingrong and Cui, Yukun (2019). "Synergistic inhibition of MEK and reciprocal feedback networks for targeted intervention in malignancy". In: *Cancer Biology & Medicine* 16.3, p. 415.

Hu-Lieskovan, Siwen, Lisberg, Aaron, Zaretsky, Jesse M, Grogan, Tristan R, Rizvi, Hira, Wells, Daniel K, Carroll, James, Cummings, Amy, Madrigal, John, Jones, Benjamin et al. (2019). "Tumor characteristics associated with benefit from pembrolizumab in advanced non–small cell lung cancer". In: *Clinical cancer research* 25.16, pp. 5061–5068.

Lin, Yuxin, Xu, Jianxin and Lan, Huiyin (2019). "Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications". In: *Journal of hematology & oncology* 12.1, p. 76.

Lindauer, A, Valiathan, CR, Mehta, K, Sriram, V, De Greef, R, Elassaiss-Schaap, J and De Alwis, DP (2017). "Translational pharmacokinetic/pharmacodynamic modeling of tumor growth inhibition supports dose-range selection of the anti–PD-1 antibody pembrolizumab". In: *CPT: pharmacometrics & systems pharmacology* 6.1, pp. 11–20.

Liu, Shu, Sun, Xiangdong, Luo, Jinhua, Zhu, Hongcheng, Yang, Xi, Guo, Qing, Song, Yaqi and Sun, Xinchen (2015). "Effects of radiation on T regulatory cells in normal states and cancer: mechanisms and clinical implications". In: *American journal of cancer research* 5.11, p. 3276.

Ma, Huilin, Wang, Hanwen, Sové, Richard J, Wang, Jun, Giragossian, Craig and Popel, Aleksander S (2020). "Combination therapy with T cell engager and PD-L1 blockade enhances the antitumor potency of T cells as predicted by a QSP model". In: *Journal for immunotherapy of cancer* 8.2.

Macal, C and North, M (2010). *Tutorial on agent-based modelling and simulation. J Simul 4: 151–162.*

Martínez-Lostao, Luis, Anel, Alberto and Pardo, Julián (2015). *How do cytotoxic lymphocytes kill cancer cells?*

McComb, Mason, Bies, Robert and Ramanathan, Murali (2022). "Machine learning in pharmacometrics: Opportunities and challenges". In: *British Journal of Clinical Pharmacology* 88.4, pp. 1482–1499.

Melo, Felipe De Sousa E, Vermeulen, Louis, Fessler, Evelyn and Medema, Jan Paul (2013). "Cancer heterogeneity—a multifaceted view". In: *EMBO reports* 14.8, pp. 686–695.

Metzcar, John, Wang, Yafei, Heiland, Randy and Macklin, Paul (2019). "A review of cell-based computational modeling in cancer biology". In: *JCO clinical cancer informatics* 2, pp. 1–13.

Michaelis, Leonor, Menten, Maud L et al. (1913). "Die kinetik der invertinwirkung". In: *Biochem. z* 49.333-369, p. 352.

Milberg, Oleg, Gong, Chang, Jafarnejad, Mohammad, Bartelink, Imke H, Wang, Bing, Vicini, Paolo, Narwal, Rajesh, Roskos, Lorin and Popel, Aleksander S (2019). "A QSP model for predicting clinical responses to monotherapy, combination and sequential therapy following CTLA-4, PD-1, and PD-L1 checkpoint blockade". In: *Scientific reports* 9.1, pp. 1–17.

Milling, Peter and Schieritz, Nadine (2003). "Modeling the forest or modeling the trees-a comparison of system dynamics and agent-based simulation". In.

Mokhtari, Reza Bayat, Homayouni, Tina S, Baluch, Narges, Morgatskaya, Evgeniya, Kumar, Sushil, Das, Bikul and Yeger, Herman (2017). "Combination therapy in combating cancer". In: *Oncotarget* 8.23, p. 38022.

Moodie, Erica EM, Chakraborty, Bibhas and Kramer, Michael S (2012). "Q-learning for estimating optimal dynamic treatment rules from observational data". In: *Canadian Journal of Statistics* 40.4, pp. 629–645.

Moore, Brett L, Pyeatt, Larry D, Kulkarni, Vivekanand, Panousis, Periklis, Padrez, Kevin and Doufas, Anthony G (2014). "Reinforcement learning for closed-loop propofol anesthesia: a study in human volunteers". In: *The journal of machine learning research* 15.1, pp. 655–696.

Murphy, Hope, Jaafari, Hana and Dobrovolny, Hana M (2016). "Differences in predictions of ODE models of tumor growth: a cautionary example". In: *BMC cancer* 16.1, pp. 1–10.

Nehme, A, Varadarajan, P, Sellakumar, G, Gerhold, M, Niedner, H, Zhang, Q, Lin, X and Christen, RD (2001). "Modulation of docetaxel-induced apoptosis and cell cycle arrest by all-trans retinoic acid in prostate cancer cells". In: *British journal of cancer* 84.11, pp. 1571–1576.

Ng, Hoi Yan, Li, Jian, Tao, Lihua, Lam, Alfred King-Yin, Chan, Kwok Wah, Ko, Josephine Mun Yee, Yu, Valen Zhuoyou, Wong, Michael, Li, Benjamin and Lung, Maria Li (2018). "Chemotherapeutic treatments increase PD-L1 expression in esophageal squamous cell carcinoma through EGFR/ERK activation". In: *Translational oncology* 11.6, pp. 1323–1333.

Norton, Larry (1988). "A Gompertzian model of human breast cancer growth". In: *Cancer research* 48.24_Part_1, pp. 7067–7071.

Obeid, Joseph M, Erdag, Gulsun, Smolkin, Mark E, Deacon, Donna H, Patterson, James W, Chen, Leiping, Bullock, Timothy N and Slingluff, Craig L (2016). "PD-L1, PD-L2 and PD-1 expression in metastatic melanoma: Correlation with tumor-infiltrating immune cells and clinical outcome". In: *Oncoimmunology* 5.11, e1235107.

Oduola, Wasiu Opeyemi and Li, Xiangfang (2018). "Multiscale tumor modeling with drug pharmacokinetic and pharmacodynamic profile using stochastic hybrid system". In: *Cancer informatics* 17, p. 1176935118790262.

Olsen, Megan M and Siegelmann, Hava T (2013). "Multiscale agent-based model of tumor angiogenesis". In: *Procedia Computer Science* 18, pp. 1016–1025.

Ott, Patrick A, Bang, Yung-Jue, Piha-Paul, Sarina A, Razak, Albiruni R Abdul, Bennouna, Jaafar, Soria, Jean-Charles, Rugo, Hope S, Cohen, Roger B, O'Neil, Bert H, Mehnert, Janice M et al. (2019). "T-cell–inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028". In: *Journal of Clinical Oncology* 37.4, pp. 318–327.

Özişik, M Necati, Orlande, Helcio RB, Colaço, Marcelo J and Cotta, Renato M (2017). *Finite difference methods in heat transfer*. CRC press.

Paalzow, Lennart K (1995). "Torsten Teorell, the father of pharmacokinetics". In: *Upsala journal of medical sciences* 100.1, pp. 41–46.

Paz-Ares, Luis, Luft, Alexander, Vicente, David, Tafreshi, Ali, Gümüş, Mahmut, Mazières, Julien, Hermes, Barbara, Çay Şenler, Filiz, Csőszi, Tibor, Fülöp, Andrea et al. (2018).

"Pembrolizumab plus chemotherapy for squamous non–small-cell lung cancer". In: *New England Journal of Medicine* 379.21, pp. 2040–2051.

Pinsky, Mark and Karlin, Samuel (2010). *An introduction to stochastic modeling*. Academic press.

Platt, Donovan (2020). "A comparison of economic agent-based model calibration methods". In: *Journal of Economic Dynamics and Control* 113, p. 103859.

Powathil, Gibin G, Adamson, Douglas JA and Chaplain, Mark AJ (2013). "Towards predicting the response of a solid tumour to chemotherapy and radiotherapy treatments: clinical insights from a computational model". In: *PLoS computational biology* 9.7, e1003120.

Recht, Abram, Silen, William, Schnitt, Stuart J, Connolly, James L, Gelman, Rebecca S, Rose, Mary Ann, Silver, Barbara and Harris, Jay R (1988). "Time-course of local recurrence following conservative surgery and radiotherapy for early stage breast cancer". In: *International Journal of Radiation Oncology* Biology* Physics* 15.2, pp. 255–261.

Rejniak, Katarzyna A and Anderson, Alexander RA (2011). "Hybrid models of tumor growth". In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 3.1, pp. 115–125.

Renshaw, Eric (2011). *Stochastic population processes: analysis, approximations, simulations*. Oxford University Press Oxford, UK.

Ribba, Benjamin, Bräm, Dominic Stefan, Baverel, Paul Gabriel and Peck, Richard Wilson (2022). "Model enhanced reinforcement learning to enable precision dosing: A theoretical case study with dosing of propofol". In: *CPT: Pharmacometrics & Systems Pharmacology* 11.11, pp. 1497–1510.

Rowland, Malcolm and Tozer, Thomas N (2005). *Clinical pharmacokinetics/pharmacodynamics*. Lippincott Williams and Wilkins Philadelphia.

Ryman, Josiah T and Meibohm, Bernd (2017). "Pharmacokinetics of monoclonal antibodies". In: *CPT: pharmacometrics & systems pharmacology* 6.9, pp. 576–588.

Sato, Hiro, Okonogi, Noriyuki, Yoshimoto, Yuya, Tamaki, Tomoaki and Suzuki, Yoshiyuki (2019). "Radiotherapy and PD-L1 expression". In: *Gan to Kagaku ryoho. Cancer & Chemotherapy* 46.5, pp. 845–849.

Sayama, Hiroki (2015). *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks.

Seel, Norbert M (2011). *Encyclopedia of the Sciences of Learning*. Springer Science & Business Media.

Sheiner, Lewis B and Beal, Stuart L (1983). "Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data". In: *Journal of pharmacokinetics and biopharmaceutics* 11.3, pp. 303–319.

Sinaga, Kristina P and Yang, Miin-Shen (2020). "Unsupervised K-means clustering algorithm". In: *IEEE access* 8, pp. 80716–80727.

Slone, Helen B, Peters, Lester J and Milas, Luka (1979). "Effect of host immune capability on radiocurability and subsequent transplantability of a murine fibrosarcoma". In: *Journal of the National Cancer Institute* 63.5, pp. 1229–1235.

Solovyev, Alexey, Mi, Qi, Tzen, Yi-Ting, Brienza, David and Vodovotz, Yoram (2013). "Hybrid equation/agent-based model of ischemia-induced hyperemia and pressure ulcer formation predicts greater propensity to ulcerate in subjects with spinal cord injury". In: *PLoS computational biology* 9.5.

Song, Yanlin, Bi, Zhenfei, Liu, Yu, Qin, Furong, Wei, Yuquan and Wei, Xiawei (2023). "Targeting RAS–RAF–MEK–ERK signaling pathway in human cancer: Current status in clinical trials". In: *Genes & Diseases* 10.1, pp. 76–88.

Sorger, Peter K, Allerheiligen, Sandra RB, Abernethy, Darrell R, Altman, Russ B, Brouwer, Kim LR, Califano, Andrea, D'Argenio, David Z, Iyengar, Ravi, Jusko, William J, Lalonde, Richard et al. (2011). "Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms". In: *An NIH white paper by the QSP workshop group*. Vol. 48. NIH Bethesda Bethesda, MD, pp. 1–47.

Srikrishnan, Vivek and Keller, Klaus (2021). "Small increases in agent-based model complexity can result in large increases in required calibration data". In: *Environmental Modelling & Software* 138, p. 104978.

St Clair, E William (2009). "Novel targeted therapies for autoimmunity". In: *Current opinion in immunology* 21.6, pp. 648–657.

Sutton, Richard S and Barto, Andrew G (2018). *Reinforcement learning: An introduction*. MIT press.

Syakur, MA, Khotimah, BK, Rochman, EMS and Satoto, Budi Dwi (2018). "Integration k-means clustering method and elbow method for identification of the best customer profile cluster". In: *IOP conference series: materials science and engineering*. Vol. 336. IOP Publishing, p. 012017.

Tabassum, Shabana, Rosli, Norhayati Binti and Binti Mazalan, Mazma Sayahidatul Ayuni (2019). "Mathematical modeling of cancer growth process: a review". In: *Journal of Physics: Conference Series*. Vol. 1366. 1. IOP Publishing, p. 012018.

Teorel, T (1937). "Kinetics of distribution of substances administered to the body II. The intravascular modes of administration". In: *Arch. Int. Pharmacodyn.* 57, p. 226.

Teorell, Torsten (1937). "Kinetics of distribution of substances administered to the body, I: The extravascular modes of administration". In: *Archives internationales de pharmacodynamie et de therapie* 57, pp. 205–225.

Terranova, Nadia, Jansen, Mendel, Falk, Martin and Hendriks, Bart S (2021). "Population pharmacokinetics of ATR inhibitor berzosertib in phase I studies for different cancer types". In: *Cancer Chemotherapy and Pharmacology* 87, pp. 185–196.

Topalian, Suzanne L, Taube, Janis M, Anders, Robert A and Pardoll, Drew M (2016). "Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy". In: *Nature Reviews Cancer* 16.5, pp. 275–287.

Tornøe, Christoffer W, Agersø, Henrik, Jonsson, E Niclas, Madsen, Henrik and Nielsen, Henrik A (2004). "Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in NLME using differential equations". In: *Computer Methods and Programs in Biomedicine* 76.1, pp. 31–40.

Truong, Van Thuy, Baverel, Paul G, Lythe, Grant, Vicini, Paolo, Yates, James WT and Dubois, Vincent FS (2022). "Step-by-step comparison of ordinary differential equation and agent-based approaches to pharmacokinetic-pharmacodynamic models". In: *CPT: Pharmacometrics and Systems Pharmacology* 11.2, pp. 133–148.

Tsimberidou, Apostolia-Maria (2015). "Targeted therapy in cancer". In: *Cancer chemotherapy and pharmacology* 76, pp. 1113–1132.

Usman, Muhammad, Khadka, Sitaram, Saleem, Mohammad, Rasheed, Huma, Kunwar, Bimal, Ali, Moshin et al. (2023). "Pharmacometrics: A New Era of Pharmacotherapy and Drug Development in Low-and Middle-Income Countries". In: *Advances in Pharmacological and Pharmaceutical Sciences* 2023.

Vanneman, Matthew and Dranoff, Glenn (2012). "Combining immunotherapy and targeted therapies in cancer treatment". In: *Nature reviews cancer* 12.4, pp. 237–251.

Vignali, Dario AA, Collison, Lauren W and Workman, Creg J (2008). "How regulatory T cells work". In: *Nature Reviews Immunology* 8.7, pp. 523–532.

Wang, Rong, Jin, Chengmeng and Hu, Xun (2017). "Evidence of drug-response heterogeneity rapidly generated from a single cancer cell". In: *Oncotarget* 8.25, p. 41113.

Wang, Zhihui, Butner, Joseph D, Kerketta, Romica, Cristini, Vittorio and Deisboeck, Thomas S (2015). "Simulating cancer growth with multiscale agent-based modeling". In: *Seminars in cancer biology*. Vol. 30. Elsevier, pp. 70–78.

Wang, Ziyan, Li, Hongyang and Guan, Yuanfang (2020). "Machine learning for cancer drug combination". In: *Clinical pharmacology and therapeutics* 107.4, p. 749.

Weichselbaum, Ralph R, Liang, Hua, Deng, Liufu and Fu, Yang-Xin (2017). "Radiotherapy and immunotherapy: a beneficial liaison?" In: *Nature reviews Clinical oncology* 14.6, pp. 365–379.

Weinstein, John N, Collisson, Eric A, Mills, Gordon B, Shaw, Kenna R, Ozenberger, Brad A, Ellrott, Kyle, Shmulevich, Ilya, Sander, Chris and Stuart, Joshua M (2013). "The cancer genome atlas pan-cancer analysis project". In: *Nature genetics* 45.10, pp. 1113–1120.

Wen, Pengbo, Gao, Yang, Chen, Bin, Qi, Xiaojing, Hu, Guanshuo, Xu, An, Xia, Junfeng, Wu, Lijun, Lu, Huayi and Zhao, Guoping (2020). "Pan-cancer analysis of radiotherapy benefits and immune infiltration in multiple human cancers". In: *Cancers* 12.4, p. 957.

Widmark, E and Tandberg, J (1924). "Uber die bedingungen fur die akkumulation indifferenter narkoliken theoretische bereckerunger". In: *Biochem Z* 147, pp. 358–369.

Wong, Harvey, Vernillet, Laurent, Peterson, Amy, Ware, Joseph A, Lee, Lillian, Martini, Jean-Francois, Yu, Peiwen, Li, Congfen, Rosario, Geoffrey Del, Choo, Edna F et al. (2012). "Bridging the gap between preclinical and clinical studies using pharmacokinetic–pharmacodynamic modeling: an analysis of GDC-0973, a MEK inhibitor". In: *Clinical Cancer Research* 18.11, pp. 3090–3099.

Xing, Xiaofang, Guo, Jianping, Ding, Guangyu, Li, Bo, Dong, Bin, Feng, Qin, Li, Shen, Zhang, Jian, Ying, Xiaomin, Cheng, Xiaojing et al. (2018). "Analysis of PD1, PDL1, PDL2 expression and T cells infiltration in 1014 gastric cancer patients". In: *Oncoimmunology* 7.3, e1356144.

Xing, Yingying, Ruan, Guojing, Ni, Haiwei, Qin, Hai, Chen, Simiao, Gu, Xinyue, Shang, Jiamin, Zhou, Yantong, Tao, Xi and Zheng, Lufeng (2021). "Tumor immune microenvironment and its related miRNAs in tumor progression". In: *Frontiers in Immunology* 12, p. 624725.

Yang, Jianchen, Virostko, Jack, Hormuth, David A, Liu, Junyan, Brock, Amy, Kowalski, Jeanne and Yankeelov, Thomas E (2021). "An experimental-mathematical approach to predict tumor cell growth as a function of glucose availability in breast cancer cell lines". In: *PloS one* 16.7, e0240765.

Yeh, Jen Jen, Routh, Elizabeth D, Rubinas, Tara, Peacock, Janie, Martin, Timothy D, Shen, Xiang Jun, Sandler, Robert S, Kim, Hong Jin, Keku, Temitope O and Der, Channing J (2009). "KRAS/BRAF mutation status and ERK1/2 activation as biomarkers for MEK1/2 inhibitor therapy in colorectal cancer". In: *Molecular cancer therapeutics* 8.4, pp. 834–843.

Zhao, Yufan, Kosorok, Michael R and Zeng, Donglin (2009). "Reinforcement learning design for cancer clinical trials". In: *Statistics in medicine* 28.26, pp. 3294–3315.

Appendix A

Model parameters of the tumour immune interaction hybrid PKPD agent based model

Parameter	**Value**	**Units**	**Source**
Reaction rates (Gillespie algorithm)			
Cancer cell division rate	0.01		assumed
Cancer cell mutation rate	0.01		assumed
Cancer cell kill rate	0.01		assumed
Cancer cell death rate	0.01		assumed
Suppressor cell moving rate	0.01		assumed
Suppressor cell division rate	0.01		assumed
Suppressor cell infiltration rate	0.01		assumed
Effector cell exhaustion rate	0.01		assumed
Effector cell moving rate	0.01		assumed
Effector cell division rate	0.01		assumed
Effector cell infiltration rate	0.01		assumed
Cell attributes			
Cancer cells			
Initial number	100		assumed
Time to division	24	h	assumed
Suppressor cells			

Initial number	1		assumed
Time to division	8	h	assumed
Division count	8		assumed
Lifespan	3*24	h	assumed
Effector cells			
Initial number	1		assumed
Time to division	8	h	(Gong, Milberg et al. 2017)
Division count	8		(Gong, Milberg et al. 2017)
Lifespan	3*24	h	(Gong, Milberg et al. 2017)
Oxygen PDE			
Diffusion coefficient D	1	$\mu m^2/d$	assumed
Division range	100-50	%	assumed
Quiescence range	50-30	%	assumed
Death range	30-0	%	assumed
Oxygen consumption rate for dividing cancer cells ko	0.001	$\%/d$	assumed
Factor for oxygen consumption of quiescent cells qc	0.01		assumed
PD1 antibody treatment			
Molecular weight	149000	g/mol	(Lindauer et al. 2017)
V1	2877/1000	l	(Lindauer et al. 2017)
V2	2854/1000	l	(Lindauer et al. 2017)
W0=V_tot	170/1000000	l	(Lindauer et al. 2017)
V_max	114/(MW)*1e3	nmol/h	(Lindauer et al. 2017)
V_es	0.005*V_tot	l	(Lindauer et al. 2017)
V_is	0.55*V_tot	l	(Lindauer et al. 2017)
K_M	0.078/(MW)*1e6	nmol/l	(Lindauer et al. 2017)

PLQ	12.7*V_vs	l/h/l	(Lindauer et al. 2017)
L	0.002*PLQ*V_vs	l/h/l	(Lindauer et al. 2017)
Q	384/(1000*24)	l/h	(Lindauer et al. 2017)
K_12	Q/V_1	1/h	(Lindauer et al. 2017)
K_21	Q/V_2	1/h	(Lindauer et al. 2017)
K_off_PD1	0.144	1/h	(Lindauer et al. 2017)
Kon_PD1_iv	2880e6/1e9	1/M/h	(Lindauer et al. 2017)
K_IVIV	1		(Lindauer et al. 2017)
K_on_PD1= Kon_PD1_iv/K_IVIV		1/nM/h	(Lindauer et al. 2017)
CL	167/(1000*24)	l/h	(Lindauer et al. 2017)
K	CL/V_1	1/h	(Lindauer et al. 2017)
N_Tcell	1000	number of T cells per μl blood	(Lindauer et al. 2017)
T_multi	4.3	initial ratio target concentration in tumour vs blood	(Lindauer et al. 2017)
V_blood	1400	μl	(Lindauer et al. 2017)
N_PD1_TC	10000	number of PD1 receptor per T cell	(Lindauer et al. 2017)
N_PD1_b	N_PD1_TC * N_Tcell * V_blood	number of PD1 in blood	(Lindauer et al. 2017)
M_PD1_b	N_PD1_b/AV*1e9	PD1 amount in nmoles in blood	(Lindauer et al. 2017)
C_PD1b	M_PD1_b/V_1	nmol/l	(Lindauer et al. 2017)

C_PD1t	T_multi*C_PD1b	nmol/l	(Lindauer et al. 2017)
vref	0.842		(Lindauer et al. 2017)
vref_is	0.2		(Lindauer et al. 2017)
CL_up	0.0366*V_es	l/h/l	(Lindauer et al. 2017)
FR	0.715		(Lindauer et al. 2017)
K_on_FcRn	792e6/(1e9)	1/nM/h	(Lindauer et al. 2017)
K_off_FcRn	23.9	1/h	(Lindauer et al. 2017)
K_deg	42.9	1/h	(Lindauer et al. 2017)
K_deg_PD1	0.00249	1/h	(Lindauer et al. 2017)
BW (human)	80	kg	(Lindauer et al. 2017)
Dose	2	mg/kg	(Lindauer et al. 2017)
Initial conditions			
C_1_0	D[nmol]/V_1	nmol/l	(Lindauer et al. 2017)
C_2_0	0	nmol/l	(Lindauer et al. 2017)
C_vs_0	0	nmol/l	(Lindauer et al. 2017)
C_ub_0	0	nmol/l	(Lindauer et al. 2017)
C_b_0	0	nmol/l	(Lindauer et al. 2017)
FcRn_0	49.8*1000	nmol/l	(Lindauer et al. 2017)
C_is_0	0	nmol/l	(Lindauer et al. 2017)
PD1_t_0	0	nmol/l	(Lindauer et al. 2017)
PD1_b_0	0	nmol/l	(Lindauer et al. 2017)
R0t_0	0	%	(Lindauer et al. 2017)

M_PD1_t_0	C_PD1b*T_multi*V_is_in	nmol	(Lindauer et al. 2017)
EMAXTP	94.7	%	(Lindauer et al. 2017)
EC50TP	1.46	nM	(Lindauer et al. 2017)
kout	K_deg_PD1	1/h	(Lindauer et al. 2017)
kin	M_PD1_t_0 * kout	nmol/h	(Lindauer et al. 2017)
Radiotherapy			
OER_m	3		(Powathil et al. 2013)
Km	3		(Powathil et al. 2013)
d	2.5	Gy	(Powathil et al. 2013)
α	0.3		(Powathil et al. 2013)
β	0.3		(Powathil et al. 2013)
Immune cell infiltration			
δ_E	0.05	1/h	(Alfonso et al. 2021)
δ_S	0.01	1/h	(Alfonso et al. 2021)
c	0.05	1/h	(Alfonso et al. 2021)
Chemotherapy			
k_d	0.0285	1/day	(Frances et al. 2011)
$D(0)$	0.132	g	(Frances et al. 2011)
p_D	0.047714		(Frances et al. 2011)
r_D	0.01251		(Frances et al. 2011)
DNA damage response inhibitor			
q	295	l/h	(Terranova et al. 2021)
cl	65	l/h	(Terranova et al. 2021)

v1	118	l	(Terranova et al. 2021)
v2	1030	l	(Terranova et al. 2021)
Emax	1		assumed
h	1		assumed
EC50	100.84	mg	(Terranova et al. 2021)

Table A.1: Model parameters of the tumour immune interaction hybrid PKPD agent based model