

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas

Programa de Pós-graduação em Matemática

Yuri Garcia Vilela

POPULATION MODELS IN CELL BIOLOGY:

exploring cancer adaptive therapy and

hematopoietic stem cell dynamics

Belo Horizonte

2025

Yuri Garcia Vilela

**POPULATION MODELS IN CELL BIOLOGY:
exploring cancer adaptive therapy and
hematopoietic stem cell dynamics**

Thesis submitted to the Graduate Program in
Mathematics of the Institute of Exact Sciences
at the Federal University of Minas Gerais as a
partial requirement for the degree of Doctor in
Mathematics.

Supervisor: Armando Gil Magalhães Neves

Co-supervisor: Artur César Fassoni

Co-supervisor: Ingmar Ruediger Glauche

Belo Horizonte

2025

	Vilela, Yuri Garcia.
V699p	Population models in cell biology: exploring cancer adaptive therapy and hematopoietic stem cell dynamics [recurso eletrônico] / Yuri Garcia Vilela — 2025. 1 recurso online (135 f. il.): pdf. Orientador: Armando Gil Magalhães Neves. Coorientador: Artur César Fassoni. Coorientador: Ingmar Ruediger Glauche. Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Matemática. Referências: f. 125-135 1. Matemática – Teses. 2. Modelagem matemática – Teses. 3. Biologia – População - Matemática – Teses. 4. Modelos estocásticos -Teses. 5. Células tronco hematopoiéticas – Teses. I. Neves, Armando Gil Magalhães. II. Fassoni, Artur César. III. Glauche, Ingmar Ruediger. IV. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Matemática. V. Título. CDU 51(043)



FOLHA DE APROVAÇÃO

*Population models in cell biology: Exploring cancer
adaptive therapy and hematopoietic stem cell dynamics*

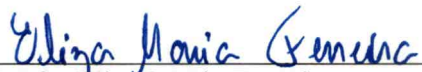
YURI GARCIA VILELA

Tese defendida e aprovada pela banca examinadora constituída por:

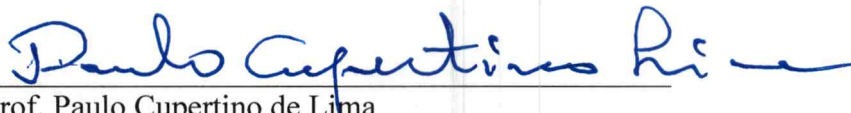

Prof. Armando G. M. Magalhães Neves
Orientador - UFMG


Prof. Artur César Fassoni
Coorientador - Unifei


Prof. Ingmar Ruediger Glauche
Coorientador - Technische Universität Dresden


Profa. Eliza Maria Ferreira
UFLA


Prof. Geraldo Lúcio Diniz
UFMT


Prof. Paulo Cupertino de Lima
UFMG


Prof. Renato Soares dos Santos
UFMG

Belo Horizonte, 31 de outubro de 2025.

Ao meu pai.

ACKNOWLEDGMENTS

This thesis is the result of a collective effort.

First, I would like to thank my family, especially my parents, Marina and Gilberto, and my brother, Patrick. They stood by my side through both the easiest and the hardest periods of my life and made me the person I am today. Without your love and support, this would never have been possible. Thank you.

Second, my deepest gratitude to my three wonderful supervisors, Armando, Artur, and Ingmar, for guidance you offered me all this time we worked together. I will always remember your kind support, and I look forward to our continued collaboration and friendship in the years to come.

I am also extremely grateful for the support of the entire Mathematics Department of UFMG and the Institute for Medical Informatics and Biometry at TU Dresden. Doing research is not an easy task, it requires outstanding commitment and perseverance to complete a PhD. Without the passionate support of all the professors and staff, it would have been impossible to carry on and remain motivated throughout this journey.

I also wish to thank CAPES for the financial support that funded most of my PhD, both in Brazil and during my time in Dresden. Funding public research brings immense social value, allowing the knowledge to flow through society rather than remain confined to a select few. Please continue this important work.

Finally, but by no means less important, I want to thank all my friends and colleagues who supported me in countless ways and at different moments over the years. During these last four years my life had many twists and without the presence of each one of you, in your own unique way, I would not have remained standing. Thank you, and I look forward to sharing my next adventures with you.

RESUMO

A modelagem matemática é uma poderosa ferramenta para descrever a dinâmica de diversos sistemas biológicos. Nesta tese são utilizados modelos populacionais para abordar dois temas da biologia celular: o planejamento de rotinas cíclicas na terapia adaptativa do câncer e a exploração de mecanismos que levam à dinâmica clonal de longo prazo das células-tronco hematopoiéticas (CTHs).

No que se refere à terapia adaptativa, modelos baseados nas equações de Lotka–Volterra e na dinâmica do replicador são analisados, sendo formalmente definidas e derivadas condições para o planejamento de rotinas de tratamento cíclicas nesses modelos. Um algoritmo para sua construção é também apresentado. Em seguida, a estabilidade do ciclo é definida, e uma conjectura é formulada relacionando-a à estabilidade do tratamento médio. Finalmente, a análise é estendida a versões estocásticas de ambos os modelos, discutindo-se a conexão entre a estabilidade de ciclos determinísticos e a robustez de rotinas estocásticas.

No segundo problema, a dinâmica das células-tronco hematopoiéticas é investigada sob duas hipóteses: se as CTHs constituem uma população de células em divisão contínua ou se um estado inativo (*quiescente*) também deve ser considerado. Esses dois conceitos são avaliados a partir da análise de diferentes aspectos de um conjunto de dados clonais recentemente publicado. O desempenho de cada modelo é comparado na reprodução dos padrões observados, e as implicações biológicas decorrentes dessa análise são discutidas.

Palavras-chave: modelagem matemática; dinâmica populacional; terapia adaptativa; tratamento do câncer; células-tronco hematopoiéticas; quiescência; dinâmica clonal; Lotka-Volterra; dinâmica do replicador; processo de Moran.

ABSTRACT

Mathematical modeling provides a powerful framework to describe and predict the dynamics of biological systems. This thesis applies population models to address two different problems in cell biology: designing cyclic routines in cancer adaptive therapy, and the mechanisms underlying long-term hematopoietic stem cell (HSC) clonal dynamics.

Adaptive therapy schemes are analyzed using models based on Lotka–Volterra and Replicator dynamics, with formal definitions and conditions provided for the design of cyclic treatment routines. An algorithm for constructing such routines is also introduced. Cycle stability is defined, and a conjecture is posed regarding the relationship between cycle and average treatment stability. The analysis is then extended to stochastic versions of both models, where connections between the stability of deterministic cycles and the robustness of stochastic routines are examined.

For the second problem, hematopoietic stem cell dynamics are analyzed under two competing hypotheses: whether HSCs constitute a population of continuously dividing cells, or whether a dormant (*quiescent*) state should also be considered. These mechanistic assumptions are evaluated using recent clonal data, analyzing multiple features of the dataset. The performance of different models is compared to explain the observed patterns, and the biological implications derived from this analysis are discussed.

Keywords: mathematical modeling; population dynamics; adaptive therapy; cancer treatment; hematopoietic stem cells; quiescence; clonal dynamics; Lotka–Volterra; replicator dynamics; Moran process.

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
HSC	Hematopoietic Stem Cells
LV	Lotka-Volterra
ODE	Ordinary Differential Equation
RSS	Residual Sum of Squares
SO	Single-time Occurring
WBC	White Blood Cells

LIST OF SYMBOLS

$\mathbf{e}_i$	i -th vector of canonical base
$\lfloor x \rfloor$	Floor of x (Largest integer smaller or equal to x)
$\overset{\circ}{S}$	Interior of set S
∂S	Boundary of set S
$\searrow$	Decreases to

CONTENTS

Introduction.....	13
I. Preamble.....	18
1. Notation and Basic Results	18
1.1. Deterministic Models	18
1.2. Stochastic Models.....	21
2. Classical Population Models.....	25
2.1. Malthus and Logistic Models for Population Growth	25
2.2. The Competitive Lotka-Volterra Model.....	27
2.3. The Replicator Dynamics	33
2.4. The Moran Process.....	41
II. Population Models in Cancer Treatment.....	49
3. Introduction	50
4. Deterministic Routines	54
4.1. Cyclic Routines in the LV Model	56
4.2. Cyclic Routines in the Adjusted Replicator Model.....	60
4.3. Cycle stability	65
5. Stochastic Routines	70
5.1. Stochastic Replicator Routines	71
5.2. Stochastic Lotka-Volterra Routines	74
5.3. Robustness of Stochastic Routines	76
6. Discussion	78
III. Population Models For HSC Dynamics.....	81
7. Introduction	82
8. Mathematical Models	86
8.1. Deterministic Formulation	87
8.2. Stochastic Formulation	91
8.3. Mean Time to Events.....	95

8.4. Multiclonal Dynamics	98
9. Model Fitting	102
9.1. Clonality Data	102
9.2. Simulation Routines and Optimization	106
9.3. Model Selection Criteria	107
10. Analysis of the Model Fit	109
10.1. Model based description of the data using a one-compartment model	109
10.2. Model based description of the data using a two-compartment model	110
10.3. Different Mechanisms of HSC Proliferation	113
11. Discussion	116
A. Supplemental Figures and Tables	119
Conclusion and Future Works	123
Bibliography	135

INTRODUCTION

“*Il libro della natura è scritto in linguaggio matematico*”¹. This famous reflection by Galileo Galilei captures the innate potential of mathematics to describe fundamental patterns underlying natural phenomena. From celestial bodies to microscopic life, mathematics provides a language through which complex processes can be translated into structures that may be analyzed, predicted, and reasoned about: the *mathematical models*.

Specially over the past century, the application of mathematics to biological problems, or *biomathematics*, has transformed our ability to study life quantitatively, bridging empirical observations with formal theoretical frameworks. In this context, mathematical modeling has proven to be an invaluable ally across diverse domains: in agriculture, to optimize crop growth and pest control; in epidemiology, to anticipate and contain infectious disease outbreaks; in ecology, to understand the balance of species and ecosystems; and, most importantly for the purposes of this thesis, in the biomedical sciences, where models help to understand the intricate dynamics of physiological systems and guide therapeutic strategies.

A central topic in modern biomathematics is *population modeling*, where mathematical tools are used to describe the growth, regulation, and interactions of individuals over time. The origins of this field can be traced back to the late 18th and 19th centuries, with the pioneering works of Thomas Malthus and Pierre-François Verhulst, who introduced models for population growth in single-species systems. In the early 20th century, Alfred Lotka [1925] and Vito Volterra [1926] independently advanced the field by developing systems of differential equations to describe predator–prey interactions, laying the foundations of modern mathematical ecology. Later on, contributions such as those of Patrick Moran [1958] and Peter Taylor and Leo Jonker [1978], further expanded the scope of population dynamics by incorporating ideas of game theory into their models.

¹ In free translation: “Mathematics is the language with which God wrote the universe.”.

An important “relative” of biomathematics is the field of *computational biology*. With its origins in studies of protein sequencing in the 1950s, computational biology has since evolved into an indispensable discipline in modern biology. It provides computational tools to generate model data, offering researchers initial insights into the general behavior of a system, as well as methodologies to handle, analyze, and interpret this model output alongside biological data. In doing so, computational biology enables efficient comparison between theoretical predictions and experimental observations, thus helping to bridge the gap between models and real-world applications.

As shown in this thesis, the integration of mathematical formalism with computational methods, both grounded by biologically meaningful ideas and interpretations, can bring valuable insights into complex biological problems. Here, these tools are combined to investigate population models in two areas of cell biology: (i) the design and stability of cyclic routines for cancer adaptive therapy, and (ii) the mechanisms underlying long-term hematopoietic stem cell (HSC) clonal dynamics.

In both cases, the term *clone* is used to describe the different types of individuals in the models. In this context, a clone is simply a group of cells that are alike in some essential way: either because they share a key property that affects how they respond to their surroundings (for example, drug resistance) or because they all come from the same ancestor cell.

Adaptive therapy is a promising paradigm for cancer treatment that seeks to exploit the inherent competition between tumor subpopulations to delay the emergence of drug resistance [Gatenby et al., 2009]. It builds on the idea that a prior presence of both drug-sensitive and drug-resistant clones in the tumor can be leveraged to design *cyclic routines*, i.e., treatment schedules that restore the tumor to the same composition before and after therapy is applied. Although adaptive therapy has already begun being translated into clinical studies [Gedye and Navani, 2022, West et al., 2023, Zhang et al., 2023] and has been the subject of extensive theoretical investigation [Zhang et al., 2017, West et al., 2018, 2019, 2020, Dua et al., 2021, Wang et al., 2021], the formal conditions under which such cycles can be systematically designed remain unclear. In this thesis, biologically motivated criteria are introduced for constructing these cycles based on two different models: the Lotka-Volterra model and replicator dynamics. The relationship between their stability and the robustness of their stochastic counterparts is discussed, highlighting the potential implications of these properties for clinical translation.

Hematopoietic stem cells are responsible for sustaining blood production throughout life,

balancing self-renewal with differentiation into multiple blood lineages [Morrison et al., 1995]. Over the past decades, it has been increasingly recognized that, in addition to these two processes, a fundamental aspect of HSC dynamics is the presence of a population of *quiescent* stem cells that can become activated over time, thereby generating additional layers of regulation [Ema et al., 2014, Cullen et al., 2014, Pinho and Frenette, 2019, Pereira et al., 2024]. The inclusion of transitions between active and quiescent states has gained substantial support, with notable contributions to the interpretation of label-dilution experiments [Glauche et al., 2009]. In this thesis, this mechanism is investigated in the context of *clonal HSC dynamics*, i.e., the evolution of the progeny derived from individual stem cells. A recently published dataset is used to analyze different metrics arising from this dynamics, systematically comparing models that include quiescence with those that do not, bringing forward new evidence to support this concept.

All models discussed here can be generally divided into two broad classes: *deterministic* and *stochastic* models. In deterministic frameworks, one assumes that the system's behavior is fully determined by initial conditions and parameters, producing the same outcome for every repetition of the same experiment. They are particularly useful when describing trends in large populations, where small perturbations are negligible in face of the magnitude of the system. In contrast, stochastic models incorporate elements of randomness, explicitly acknowledging fluctuations inherent to any natural process. This perspective is essential when considering small systems or longer time-scales, when the accumulated small perturbations significantly affect the observed outcomes. As will be demonstrated throughout this text, both approaches are valuable modeling tools, offering complementary perspectives for analyzing complex biological phenomena.

This thesis is organized in four parts. Part I (preamble) sets notation and reviews important mathematical tools, presenting classical population models together with their key dynamical properties. Part II develops and analyses cancer adaptive therapy models. A formal definition of cyclic routines is provided, together with an algorithm for their construction using deterministic models and a discussion of their stability. The corresponding stochastic routines are then introduced and explored with particular attention to how robustness may affect their translational potential in clinical settings. Part III addresses HSC dynamics. After an introduction of two deterministic models under different mechanistic assumptions, the focus shifts to stochastic, multiclonal formulations, presenting parameter estimation procedures to compare model predictions with experimentally-derived metrics. The best model fits are used to discuss how the different

biological assumptions explain the data. Finally, Part IV summarizes the main contributions, discusses limitations, and outlines directions for future work.

PART I

PREAMBLE

1 NOTATION AND BASIC RESULTS

1.1 Deterministic Models

Deterministic models are those in which the initial conditions of a system *uniquely determine* its state at future times. In other words, in these models there is no randomness: repeated runs under the same conditions always yield the same outcome. These models appear widely across several fields such as biology, chemistry, physics, and economics, and are often the first step in studying several phenomena.

One common way of formulating these models is using systems of *ordinary differential equations* (ODEs). Generally speaking, an ODE system is written as

$$\begin{cases} \mathbf{x}'(t) = F(\mathbf{x}(t), t), \end{cases}$$

where $\mathbf{x}(t) = (x_1(t), \dots, x_n(t)) \in D \subset \mathbb{R}^n$ represents the state of a system (e.g., population or frequency of a particular clone) in some region of biological interest D at time t . The function $F(\mathbf{x}, t)$ (often called the *vector field function*) determines the rate of change of the system's state $\mathbf{x} \in D$ at time t . Throughout the text the dependency of $\mathbf{x}(t)$ on t will often be omitted, writing simply $\mathbf{x}$.

This thesis focuses specially on *autonomous* ODE systems, that is, systems where the rates of change remain unchanged over time. In that context, the system simplifies to

$$\begin{cases} \mathbf{x}'(t) = F(\mathbf{x}(t)). \end{cases} \quad (1.1)$$

For such a system, the following general result holds (see, e.g., Section 17.1 of Hirsch et al. [2013]):

Theorem 1 (Existence and Uniqueness of ODE Solutions). *Let (1.1) be a system of ODEs with $F \in C^1$ in some open region D . Then for every $\mathbf{x}_0 \in D$ there exists a maximal (open) interval $I(x_0) \ni 0$ such that there exists a unique solution $\mathbf{x}(t)$ of the ODE in $I(x_0)$ with $\mathbf{x}(0) = \mathbf{x}_0$.*

An immediate observation from this theorem is that if $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are solutions of the same ODE with $\mathbf{x}(t_x) = \mathbf{y}(t_y)$ for some t_x and t_y then, in fact, these solutions coincide in their respective intervals of existence. In other words, two solutions of an (autonomous) ODE that coincide in some point coincide entirely.

For all systems studied here, unique solutions will exist for all initial conditions and all *future* time, i.e., for all $\mathbf{x}_0$ in the region of biological interest D , a unique solution with $\mathbf{x}(0) = \mathbf{x}_0$ will exist in a maximal interval of existence of the form $I(\mathbf{x}_0) = (-T_0, \infty)$ for some positive, possibly infinite, T_0 .

This leads to the definition of the *flow* of an ODE system. Formally, the flow is a function that associates each state $\mathbf{x}_0 \in D$ to the solution $\mathbf{x}(t)$ of the ODE with initial condition $\mathbf{x}(0) = \mathbf{x}_0$ on its maximal interval of existence $I(\mathbf{x}_0)$. In this thesis the flow is denoted as the two variable function $\Psi(t, \mathbf{x}_0)$, implicitly assuming that t is in the interval of existence of the ODE solution.

From this definition, two direct properties of the flow follow: first, the flow evaluated at time $t = 0$ returns the initial condition, i.e., $\Psi(0, \mathbf{x}_0) = \mathbf{x}_0$ for all $\mathbf{x}_0 \in D$. Second, the time derivative of the flow recovers the vector field F , meaning that

$$\frac{\partial}{\partial t} \Psi(t, \mathbf{x}_0) = \Psi'(t, \mathbf{x}_0) = F(\Psi(t, \mathbf{x}_0)).$$

The *orbit* of a point $\mathbf{x}_0$ (under an ODE) is the set of all points in D that are reached by the flow $\Psi(t, \mathbf{x}_0)$ as t varies over its interval of definition. Since the map $t \mapsto \Psi(t, \mathbf{x}_0)$ is continuous, these orbits can be identified with continuous *curves* in D . Moreover, since we are working with autonomous systems, maximal intervals and uniqueness of solutions holds, two *distinct* orbits cannot intersect, i.e., if two orbits intersect at a point, then they coincide entirely. Throughout this text, the *future orbit* and the *past orbit* of a point $\mathbf{x}_0$ will also be referred to, meaning the restriction of the flow $\Psi(t, \mathbf{x}_0)$ to $t \geq 0$ or $t \leq 0$, respectively.

With this nomenclature, one important result to establish the extension of ODE solutions for future times can be enunciated (see, e.g., Section 17.4 of Hirsch et al. [2013]):

Theorem 2. *Let (1.1) be a system of ODEs with $F \in C^1$ in some open region D and $\mathbf{x}_0 \in D$. If the future orbit of $\mathbf{x}_0$ lies entirely in some compact set $C \subset D$, then the maximum interval of existence of $I(\mathbf{x}_0)$ has the form $(-T_0, \infty)$ for some positive, possibly infinite, T_0 .*

A very important concept in the study of differential equations is that of *equilibrium solutions*. A solution $\mathbf{x}(t)$ of the system (1.1) is said to be an *equilibrium solution* (or an *equilibrium point*) if it remains constant over time, that is, $\mathbf{x}(t) = \mathbf{x}_0$ for all $t \in \mathbb{R}$. By the

uniqueness of the solutions, it trivially follows that $\mathbf{x}_0$ is an equilibrium point if, and only if, $F(\mathbf{x}_0) = \mathbf{0}$.

Often, as will be the case in this thesis, the dynamical analysis of an ODE system requires identifying and classifying the equilibrium points and understanding how other trajectories behave with respect to them. Since equilibrium points are those where the derivatives vanish, it follows that, in order to determine the system's equilibria, one must solve the equation $F(\mathbf{x}) = \mathbf{0}$. Naturally, there is no standard procedure to do this in general, since the vector field F may take any form. However, a useful technique is to study the system's *nullclines*.

The *nullcline* corresponding to the i -th coordinate (i.e., to the variable x_i) is the subset of D defined by the equation $F_i(\mathbf{x}) = 0$. These sets often provide valuable insight into the structure of the system. Nullclines partition the dynamical region D into areas where certain components of the state vector monotonically increase or decrease. Any point lying in the intersection of all nullclines is necessarily an equilibrium point.

Following their identification, equilibrium points can be generally classified according to their *stability* properties:

- **Stable equilibrium:** An equilibrium point $\mathbf{x}_0$ is said to be *stable* if solutions starting sufficiently close to $\mathbf{x}_0$ remain close for all future times.
- **Asymptotically stable equilibrium:** A stable equilibrium $\mathbf{x}_0$ is said to be *asymptotically stable* (also called a *sink* or an *attractor*) if, in addition, all nearby trajectories converge to it as $t \rightarrow \infty$.
- **Unstable equilibrium:** An equilibrium $\mathbf{x}_0$ is said to be *unstable* if it is not stable, that is, there exist solutions arbitrarily close to $\mathbf{x}_0$ that eventually move away.
- **Repellor:** An equilibrium point $\mathbf{x}_0$ is a *repellor* (or a *source*) if all nearby solutions eventually leave some neighborhood as time progresses. Notably, repellors can also be regarded as attractors for the backwards time dynamics.
- **Saddle point:** An equilibrium $\mathbf{x}_0$ is a *saddle* if it is unstable and there exist trajectories approaching it (stable directions) and others diverging from it (unstable directions).

Once again, there is no general technique to classify equilibrium points. Throughout this thesis, stability properties of equilibrium points will often be described without providing full formal proofs. Naturally, in these cases, appropriate references in the literature where such proofs

can be found will be indicated. The main tools employed here for this purpose are classical results from the theory of dynamical systems: the *Hartman-Grobman theorem* (see, e.g., Section 2.8 of Perko [1991]), which allows the study of local dynamics of *hyperbolic equilibria* via linearization, and the *Poincaré-Bendixson theorem* (see, e.g., Section 3.7 of Perko [1991]), which establishes convergence behaviors in two-dimensional systems.

1.2 Stochastic Models

A *stochastic model* is a mathematical model in which *randomness* plays a fundamental role in determining the evolution of the system. That is, even under identical initial conditions, different realizations of a stochastic process may evolve differently, purely due to chance. These models are crucial in capturing variability and uncertainty inherent in real-world systems, particularly in biological settings where such fluctuations are not only present but often influential.

In mathematical biology, stochastic models are especially relevant when studying *small population sizes* or *long-term generational effects* where random events, such as reproduction, mutation, or death, can significantly alter the system's behavior and lead to outcomes that deterministic models would not predict. This chapter introduces notation and enunciates important results to be used along the thesis.

This thesis focuses on *discrete-time*, *finite state space*, *Markovian* processes. Formally, a process with n clones will be represented as a sequence of vectors

$$\{\mathbf{X}^k\}_{k \in \mathbb{N}}, \quad \mathbf{X}^k = (X_1^k, \dots, X_n^k) \in D \subset \mathbb{N}^n,$$

where the component $X_i^k \in \mathbb{N}$ denotes the number of individuals of clone i at time step k . The evolution of the process in discrete steps $k \in \mathbb{N}$ is precisely what it means to be a *discrete-time* process.

To have a *finite state space* means that the set of possible states D is a finite subset of $\mathbb{N}^n$. Finally, a process is *Markovian* (or a *Markov chain/process*) if its next step depends only on its current state, not on its history. Precisely, being a Markov process means that for any $k \in \mathbb{N}$ and any sequence of states $\mathbf{Y}_0, \dots, \mathbf{Y}_{k+1} \in D$ it holds that

$$P(\mathbf{X}^{k+1} = \mathbf{Y}_{k+1} \mid \mathbf{X}^0 = \mathbf{Y}_0, \dots, \mathbf{X}^k = \mathbf{Y}_k) = P(\mathbf{X}^{k+1} = \mathbf{Y}_{k+1} \mid \mathbf{X}^k = \mathbf{Y}_k).$$

This *memoryless* property simplifies the analysis of such models and naturally reflects many biological contexts where transitions are governed by local or time-independent rules.

Unlike deterministic models, where the system state is uniquely determined by the initial condition, stochastic models incorporate randomness, so the system's evolution cannot be predicted exactly. Instead, stochastic processes have *realizations* (or *trajectories*): individual sequences of states that occur in particular runs of the system.

The evolution of a stochastic process is fully determined by its *transition probabilities*, which specify the likelihood of moving from one state to another. Formally, the probability of transitioning from state $\mathbf{Y}_k$ to state $\mathbf{Y}_{k+1}$ on step $k+1$ is

$$P(\mathbf{X}^{k+1} = \mathbf{Y}_{k+1} \mid \mathbf{X}^k = \mathbf{Y}_k) = \pi_{k+1}(\mathbf{Y}_k, \mathbf{Y}_{k+1}), \quad \mathbf{Y}_k, \mathbf{Y}_{k+1} \in D, k \in \mathbb{N}.$$

When working with Markov chains, these transition probabilities completely characterize the process. If they are independent of the step k , the chain is called *time-homogeneous* (or *time-invariant*). Unless otherwise stated, all processes presented in this thesis are time-homogeneous.

For a time-homogeneous Markov chain with finite state space D , the transition probabilities can be arranged into a *transition matrix*

$$\Pi = (\Pi_{\Lambda(\mathbf{Y}_0), \Lambda(\mathbf{Y}_1)} := \pi(\mathbf{Y}_0, \mathbf{Y}_1))_{\mathbf{Y}_0, \mathbf{Y}_1 \in D},$$

where $\Lambda : D \rightarrow \{1, \dots, \#D\}$ is an enumeration of the state space. Notice that the element on row $\Lambda(\mathbf{Y}_0)$ and column $\Lambda(\mathbf{Y}_1)$ of this matrix is precisely the probability of transitioning from state $\mathbf{Y}_0$ to state $\mathbf{Y}_1$ in one step. Consequently, the transition matrix Π is a *row stochastic matrix*, i.e., a non-negative matrix on which each row sums to 1. By abuse of notation, the enumeration Λ will be omitted, and the elements of the transition matrix will be written as $\Pi_{\mathbf{Y}_0 \mathbf{Y}_1} = \pi(\mathbf{Y}_0, \mathbf{Y}_1)$.

This matrix formulation provides a convenient way to compute multi-step transition probabilities. Namely, the probability of being in state $\mathbf{Y}_k$ at step k given that the system started in state $\mathbf{Y}_0$ at step 0 is

$$P(\mathbf{X}^k = \mathbf{Y}_k \mid \mathbf{X}^0 = \mathbf{Y}_0) = (\Pi^k)_{\mathbf{Y}_0 \mathbf{Y}_k},$$

where (Π^k) is the multiplication of matrix Π k times.

Despite being the core of stochastic processes, in some contexts those transition probabilities and transition matrices need not be given *in closed form*. Instead, they can be defined implicitly through rules or mechanisms that govern individual events from which the transition probabilities can be derived when necessary. For example, one might specify the probability that a particular species reproduces or dies at a given step, and then compute the resulting transition probabilities between full population states if needed.

Indeed, in the case of this thesis, the explicit transition probabilities will not be required. From a theoretical perspective, it will be enough to understand which transitions are *possible* (i.e., have positive probability) rather than to compute their exact values, while from a practical perspective, the description of these underlying mechanisms will be sufficient to simulate realizations of the processes.

Two important concepts in the context of Markov chains are those of *absorbing* and *transient* states [Durrett, 2019, Neves, 2014]. A state $\mathbf{Y} \in D$ is said to be *absorbing* if, once the process enters this state, it remains there forever with probability one. That is,

$$P(\mathbf{X}^{k+1} = \mathbf{Y} \mid \mathbf{X}^k = \mathbf{Y}) = 1 \quad \forall k \in \mathbb{N}.$$

Absorbing states represent final configurations from which the system cannot escape.

In contrast, a state is said to be *transient* if, with positive probability, a process starting from that state never returns. An immediate consequence of that definition is that *with probability one*, a Markov chain visits transient states only a finite number of times. In other words, although the process may revisit a transient state a few times, *almost surely* it will eventually leave and never return.

A state that is *not* transient is called *recurrent*, meaning that the process returns to it infinitely often with probability one. Absorbing states form a special class of recurrent states where, after return, the process never leaves again.

The Fixation Effect

A fundamental effect that differentiates the deterministic and stochastic models presented here is the so-called *fixation effect*. Whereas multiple-clone coexistence is a feasible outcome in all deterministic models explored here, all stochastic models ultimately share a common fate: the fixation of a single clone. Particularly in models where all cells are functionally equivalent, this result is not intuitive. In such systems, one might expect perpetual coexistence of all clones initially present in the population. Yet, the stochastic dynamics drive these systems toward monoclonality due to *random drift*.

Before proceeding to formal results, the introduction of some final notation is in order. Let $\rho_{\mathbf{X}\mathbf{Y}}^k$ denote the probability of reaching state $\mathbf{Y} \in D$ from an initial state $\mathbf{X} \in D$ in *exactly* k steps, with no visits to $\mathbf{Y}$ in the intermediate steps. For a time-homogeneous Markov chain, this

probability can be written as

$$\rho_{\mathbf{X}\mathbf{Y}}^k = P\left(\mathbf{X}^1 \neq \mathbf{Y}, \dots, \mathbf{X}^{k-1} \neq \mathbf{Y}, \mathbf{X}^k = \mathbf{Y} \mid \mathbf{X}^0 = \mathbf{X}\right).$$

With that notation, the probability that the process *eventually* reaches $\mathbf{Y}$ starting from $\mathbf{X}$ is given by

$$\rho_{\mathbf{X}\mathbf{Y}} := \sum_{k=1}^{\infty} \rho_{\mathbf{X}\mathbf{Y}}^k.$$

The following results summarize the central ideas to explain the fixation behavior in the models presented here:

Theorem 3 (Theorem 1.5 of Durrett [2019]). *If $\rho_{\mathbf{X}\mathbf{Y}} > 0$ but $\rho_{\mathbf{Y}\mathbf{X}} < 1$, then state $\mathbf{X}$ is transient.*

Theorem 4 (Adapted from Theorem 1.7 of Neves [2014]). *Let $\{\mathbf{X}^k\}_{k \in \mathbb{N}}$ be a Markov chain with a finite state space D . Suppose there exists a non-empty subset $S \subset D$ of absorbing states and that all other states in $D \setminus S$ are transient. Then, with probability one, all trajectories will eventually be absorbed into some state in S .*

Besides the one presented in [Neves, 2014], another elegant demonstration of Theorem 4 can be found in [Grinstead and Snell, 2003], using the *transition matrix* of the Markov chain.

Intuitively, Theorem 3 states that if it is possible to move from state $\mathbf{X}$ to state $\mathbf{Y}$, but there exists a possibility of never returning from $\mathbf{Y}$ to $\mathbf{X}$, then eventually the process will not return. Theorem 4 states that if only a finite set of states exists, and either these states are visited only a finite number of times, or are absorbing, then, with probability 1, the process will eventually reach a state that cannot be exited. In the context of this thesis, these results lead to an important theoretical distinction between deterministic and stochastic models: whereas deterministic dynamics may predict stable coexistence between clones, the incorporation of stochasticity implies that coexistence is only temporary, and eventual fixation becomes unavoidable, even if the time to fixation is exceedingly long.

It is worth noting, however, that this absorption phenomena does not always occur in timescales that are relevant in practice. As will be shown in later chapters, when the deterministic model predicts stable coexistence, the underlying stability structure can hold the system in a balanced, *metastable* state for a very long time. In such cases, the shift to monoclonality happens so slowly that it may never be observed in realistic biological settings, with fixation occurring only in small, simplified systems, whereas in larger and more realistic ones, long-term coexistence remains the practical outcome.

2 CLASSICAL POPULATION MODELS

2.1 Malthus and Logistic Models for Population Growth

One of the simplest, and often first introduced, population models is the *Malthusian population model*, also known as the *exponential growth model* [Hofbauer and Sigmund, 1998, Murray, 2002]. This model describes a population that grows (or declines) at a constant rate r over time. More precisely, if $x(t)$ is the population size at time t , then

$$x(t + \Delta t) = x(t) + rx(t)\Delta t + \mathcal{O}(\Delta t^2)$$

Taking the limit as $\Delta t \rightarrow 0$, this formulation leads to the ordinary differential equation model

$$x'(t) = rx(t).$$

This ODE can also be solved generally by separation of variables, yielding the solution

$$x(t) = x(0)e^{rt}.$$

As in most population models, the Malthus model exhibits an extinction equilibrium $x_0 \equiv 0$. However, the main drawback of this model is that, when $r > 0$, it predicts unlimited population growth starting from any non-zero population. This is unrealistic, as resource constraints and environmental limitations naturally inhibit reproduction at large population sizes.

A more realistic model accounting for such constraints was proposed by Verhulst between 1838 and 1845 [Murray, 2002]. He introduced a linear reduction in the growth rate as the population increases, resulting in the *Verhulst population model*, or *logistic growth model*, given by

$$x'(t) = rx(t) \left(1 - \frac{x(t)}{K} \right)$$

Here, r remains the intrinsic growth rate, and $K > 0$ is the *carrying capacity*, i.e., the maximum population size that the environment can sustain.

This ODE also admits an explicit solution via separation of variables

$$x(t) = \frac{Kx(0)}{Ke^{-rt} + x(0)(1 - e^{-rt})}.$$

From this expression, it is clear that for any initial population $x(0) > 0$ and $r > 0$, the population $x(t)$ tends to the carrying capacity K as $t \rightarrow \infty$. In other words, $x(t) \equiv K$ is a *globally asymptotically stable* equilibrium of the dynamics in $\mathbb{R}_+$.

Moreover, when $0 < x(0) \ll K$, we have the approximation $x(t) \approx x(0)e^{rt}$ for small t , showing that the logistic model reduces to the Malthusian model during the early phase of growth (Figure 1).

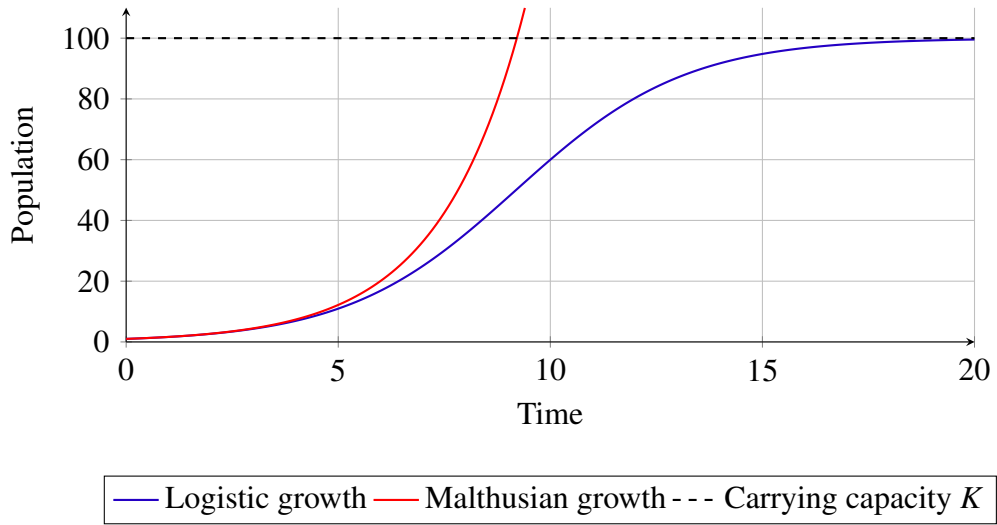


Figure 1 – Comparison of Malthusian and Logistic growth models. While the Malthusian model predicts unbounded growth, the Logistic model saturates at the carrying capacity K . Both models show similar behavior for small populations on early time points.

Although the solution $x(t)$ is defined for every initial condition $x(0) \in \mathbb{R}$, it is not necessarily defined for all values of t . Indeed, notice that the solution becomes ill-defined when the denominator of the solution expression vanishes, that is, for

$$t^* = -\frac{1}{r} \ln \left(\frac{x(0)}{x(0) - K} \right).$$

When $0 < x(0) < K$, the argument of the logarithm is negative, and hence t^* is not defined in $\mathbb{R}$. In this case, we have that $x(t) \rightarrow 0$ as $t \rightarrow -\infty$, and the solution is well-defined for all $t \in \mathbb{R}$.

However, if $x(0) > K$, the expression for t^* becomes real and negative. That is, there exists a finite time $t^* < 0$ for which the denominator of the solution becomes zero. As $t \searrow t^*$, the denominator decreases to zero, and thus $x(t) \rightarrow \infty$. Therefore, evolving the system backwards in time, the population $x(t)$ would “blow up” in finite time t^* . Although $x(0) < 0$ has no biological meaning, for such initial conditions we have $t^* > 0$ and the solution of the initial value problem is defined for $t \in (-\infty, t^*)$.

Despite this blow up for a finite negative time, the idea of *logistic adjustment* modifying rates to account for environmental limitations is widely used in population modeling and will naturally reappear in some of the models discussed throughout this thesis. In particular, the solution of the logistic model for any positive initial population is well defined for all future time, which is a satisfactory feature for biologic use.

2.2 The Competitive Lotka-Volterra Model

The *Lotka-Volterra* (LV) equations [Hofbauer and Sigmund, 1998, Murray, 2002] are a system of ordinary differential equations (ODEs) independently introduced in the 1920s by mathematician Vito Volterra and biophysicist Alfred J. Lotka. In Volterra’s formulation, the model considers a prey population that grows at a constant rate in the absence of predators, while predators die at a constant rate in the absence of prey. When both species are present, their interaction is modeled through a constant-rate consumption of prey by predators, who in turn require the presence of prey to survive.

A generalized version of this framework for n interacting species is given by the system

$$x'_i = x_i f_i(\mathbf{x}), \quad i = 1, \dots, n,$$

where each f_i is a polynomial of degree 1 in the coordinates of the vector $\mathbf{x} = (x_1, \dots, x_n)$. In this model, the functions $x_i(t)$ denotes the population size of species i at time t , and the functions f_i encode the nature of the interactions between species.

A particularly important case is that of *competitive systems*, where all individuals compete for a common pool of resources. In this setting, the presence of any individual, regardless of species, reduces the ability of others to thrive. A standard way to model such interactions is through the system

$$x'_i = r_i x_i \left(1 - \frac{\sum_{j=1}^n a_{ij} x_j}{k_i} \right), \quad i = 1, \dots, n. \quad (2.1)$$

Here, $r_i > 0$ is the *intrinsic growth rate* of species i , and the interaction parameter a_{ij} measures the effect of individuals of species j on the growth of species i . Competitiveness lies in the hypothesis that

$$a_{ij} > 0 ,$$

i.e., the presence of any species decreases the growth potential of all others, including itself. These coefficients will be referred to as the *aggressiveness* of species j toward species i . Without loss of generality, the intraspecific aggressiveness will be set to 1, i.e.,

$$a_{ii} = 1 .$$

Under this convention, the constants $k_i > 0$ correspond precisely to the *carrying capacity* of species i , i.e., the maximum population of species i sustainable by the environment in the absence of competitors. Notably, if $x_j = 0$ for all $j \neq i$, the LV equations reduce to the logistic model of species i with carrying capacity k_i .

In a competitive n -species Lotka-Volterra system, there are always at least $n + 1$ equilibrium points in the biologically relevant region $\mathbb{R}_+^n$, namely, the *extinction equilibrium* $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^n$, and the n *single-species equilibria* given by $\mathbf{E}_i = k_i \mathbf{e}_i$, where $\mathbf{e}_i$ is the i -th canonical basis vector and $i = 1, \dots, n$.

The extinction equilibrium $\mathbf{0}$ is always a *repellor*. Indeed, when $x_1, \dots, x_n \approx 0$, the inequality $\sum_j a_{ij} x_j < k_i$ holds for all i , and the derivative $x'_i = r_i x_i \left(1 - \frac{\sum_j a_{ij} x_j}{k_i} \right)$ is positive, indicating that small populations tend to move away from the origin.

In contrast, the single-species equilibria $\mathbf{E}_i = k_i \mathbf{e}_i$ are never repelling along their respective coordinate axes. Indeed, the system dynamics restrained to the i -th axis is precisely a *logistic population model*, so any positive initial condition on that axis tends to $\mathbf{E}_i$. However, whether this equilibrium is an *attractor* or a *saddle point* in the full phase space depends on the interspecific interaction terms. Namely, if other species are sufficiently aggressive toward species i , then $\mathbf{E}_i$ become unstable in directions outside the i -th axis.

Similar to the logistic model, solutions of the competitive Lotka–Volterra system are defined for all future times. To see this, consider the compact rectangle

$$L = [0, L_1] \times \dots \times [0, L_n], \quad L_i > k_i \text{ for } i = 1, \dots, n.$$

If $\mathbf{x} \in \mathbb{R}_+^n$ satisfies $x_i = L_i$, then

$$\sum_{j=1}^n a_{ij} x_j \geq a_{ii} x_i = x_i = L_i > k_i,$$

therefore $x'_i < 0$. Thus, on each face $\{\mathbf{x} : x_i = L_i\}$ of L , the vector field points strictly inward. On the other hand, in the faces $\{\mathbf{x} : x_i = 0\}$ the derivative x'_i is null, so trajectories starting on these faces remain there. Together, these observations show that L is positively invariant: if a trajectory starts in L , it remains in L for all $t \geq 0$.

Since for any initial condition it can be defined such a compact L containing it, Theorem 2 implies

Proposition 1. *Non-negative solutions of the competitive Lotka-Volterra system are defined for all $t \geq 0$.*

On the other hand, as in the logistic model, some solutions of the LV system also exhibit finite-time blow-up for negative time. Indeed, consider an initial condition $\mathbf{x}(0)$ such that $x_i(0) > k_i$ for all i . Naturally, it follows that

$$\sum_{j=1}^n a_{ij}x_j(0) > k_i \quad \text{for all } i = 1, \dots, n.$$

As already noted this condition implies $x'_i(0) < 0$ for all i , and hence, moving backwards in time ($t < 0$), the populations $x_i(t)$ monotonically increase. Consequently the inequality $\sum_j a_{ij}x_j(t) > k_i$ still holds for all $t < 0$, and it follows that all populations increase backwards in time.

In particular, such growth can be bound from below:

$$\begin{aligned} -x'_i(t) &= r_i x_i(t) \left(\frac{\sum_{j=1}^n a_{ij}x_j(t)}{k_i} - 1 \right) \\ &> r_i x_i(t) \left(\frac{a_{ii}x_i(t)}{k_i} - 1 \right) \\ &= -r_i x_i(t) \left(1 - \frac{x_i(t)}{k_i} \right), \end{aligned}$$

which coincides with the negative of the logistic growth derivative. Therefore, for $t < 0$, the population $x_i(t)$ grows *faster* than the logistic model. Since the logistic model exhibits finite-time blow-up for negative time when starting above the carrying capacity, it follows that the same holds for the Lotka–Volterra system in this regime.

Despite not being defined for all negative times, this construction still tells us that prior to the blow-up these populations are monotonically increasing backwards in time. For the purposes of this thesis, this is enough to derive important properties for this model.

2.2.1 Two Species Competitive Lotka-Volterra

This thesis is particularly focused in the Lotka-Volterra model in the context of *two competing species*, i.e., $n = 2$.

In this scenario, the qualitative behavior of the system can be easily analyzed by examining the relative positions of the nullclines. Consider the system

$$\begin{cases} x_1' = r_1 x_1 \left(1 - \frac{x_1 + a_{12} x_2}{k_1} \right), \\ x_2' = r_2 x_2 \left(1 - \frac{a_{21} x_1 + x_2}{k_2} \right), \end{cases}$$

where $r_1, r_2, k_1, k_2, a_{12}$ and a_{21} are all positive.

In the interior of the positive quadrant $\mathbb{R}_+^2$, the nullcline of x_1 is the line segment connecting the points $(k_1, 0)$ and $(0, \frac{k_1}{a_{12}})$, while the nullcline of x_2 connects $(0, k_2)$ and $(\frac{k_2}{a_{21}}, 0)$. These nullclines divide the phase space and determine the direction of the flow (Fig. 2).

By analyzing the signs of the derivatives, it follows that in the region *below* a given nullcline, the corresponding population *increases*, while in the region *above* that nullcline, it *decreases*. This leads to two key dynamical regions that always exist in the phase space: the region *below both nullclines*, where both populations increase, and the region *above both nullclines*, where both populations decrease. Similar to the construction made in the previous sections, the solutions starting above both nullclines present monotonic increase until a finite-time blow-up backwards in time.

In general, two distinct scenarios may occur: either the nullclines intersect in $\mathbb{R}_+^2$, yielding a unique interior equilibrium point corresponding to the coexistence of both species (Figure 2A and C), or they do not intersect in $\mathbb{R}_+^2$, in which case the only biologically relevant equilibria are the extinction point and the two single-species equilibria, $(k_1, 0)$ and $(0, k_2)$ (Figure 2B and D).

When the nullclines intersect, an interior coexistence equilibrium occurs at the point $\mathbf{x}^* = \left(\frac{k_1 - a_{12}k_2}{1 - a_{12}a_{21}}, \frac{k_2 - a_{21}k_1}{1 - a_{12}a_{21}} \right)$. Regarding its stability, it can arise under two different regimes:

- If $k_1 < \frac{k_2}{a_{21}}$ and $k_2 < \frac{k_1}{a_{12}}$ (Figure 2A), neither species is sufficiently aggressive. Each single-species equilibrium lies *below* the other species' nullcline, meaning that, even if introduced with few individuals, the new species population increases. In this case, both single-species equilibria are saddle points and the interior coexistence equilibrium is globally asymptotically stable in the interior of $\mathbb{R}_+^2$.

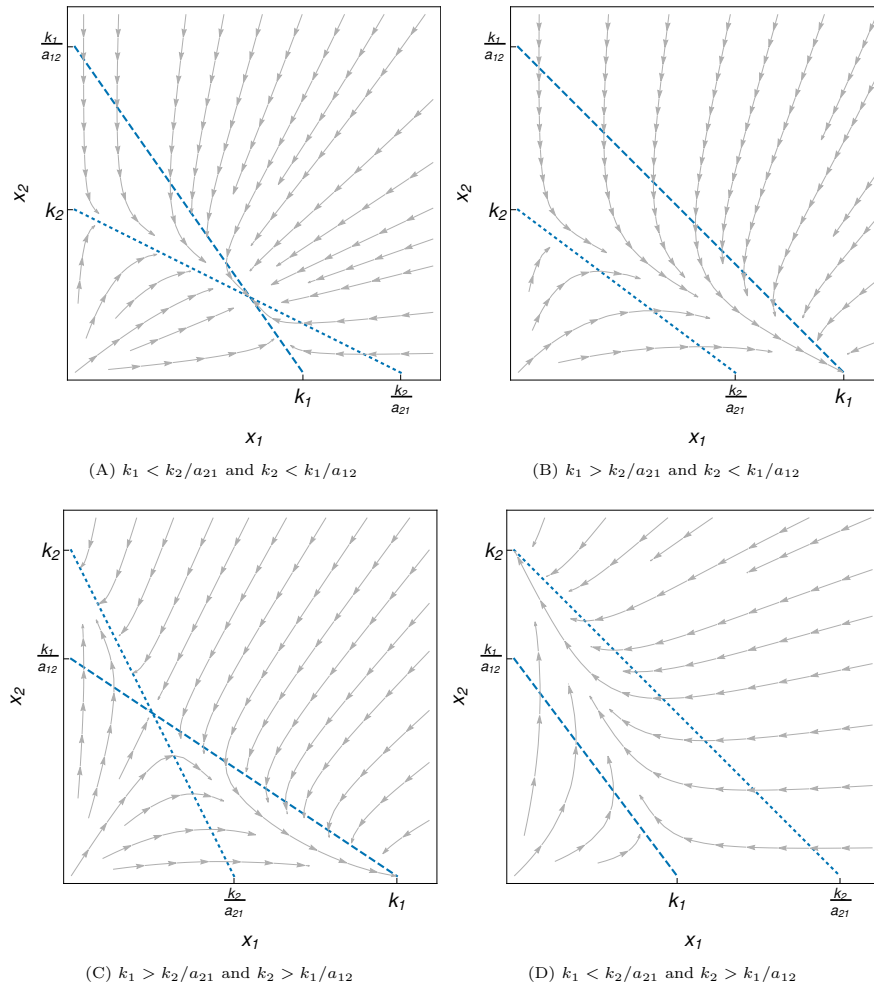


Figure 2 – Flow diagrams of Lotka-Volterra systems with stable coexistence (A), sole survival of x_1 (B), unstable coexistence (C) and sole survival of x_2 (D). In all four cases, the dashed line is the x_1 nullcline and the dotted line is the x_2 nullcline.

- If $k_1 > \frac{k_2}{a_{21}}$ and $k_2 > \frac{k_1}{a_{12}}$ (Figure 2C), both species are said to be *aggressive*. In this case, each single-species equilibrium lies *above* the other species' nullcline, meaning the absent species would decrease if introduced at low density. This implies that both single-species equilibria are locally stable attractors and the interior coexistence equilibrium is a saddle point.

On the other hand, if the nullclines do not intersect in $\mathbb{R}_+^2$, no coexistence equilibrium exists in the biologically relevant region. This situation arises when only one species is sufficiently aggressive, leading to the competitive exclusion of the less aggressive species. Once again, two distinct configurations can be identified:

- If $k_1 > \frac{k_2}{a_{21}}$ and $k_2 < \frac{k_1}{a_{12}}$ (Figure 2B), species 1 is the more aggressive competitor. In this

case, the nullcline of x_2 lies entirely below that of x_1 , meaning that species 1 is still growing even after species 2 is already out of resources. It can be shown that the only locally stable equilibrium is the single-species equilibrium $\mathbf{E}_1 = (k_1, 0)$, and all trajectories in the interior of $\mathbb{R}_+^2$ converge to it. Species 1 survives, and species 2 is driven to extinction.

- Conversely, if $k_1 < \frac{k_2}{a_{21}}$ and $k_2 > \frac{k_1}{a_{12}}$ (Figure 2D), species 2 is the more aggressive competitor. The nullcline of x_1 lies entirely below that of x_2 , and species 1 cannot persist. The only locally stable equilibrium is the single-species equilibrium $\mathbf{E}_2 = (0, k_2)$, with all interior trajectories converging to it. In this case, species 2 survives while species 1 goes extinct.

The complete classification of equilibria and their stability in the two-species Lotka-Volterra system is summarized in Table 1. The complete stability study of these equilibria through the analysis of the Jacobian matrix $DF(\mathbf{x})$ can be found in Section 3.3 of Hofbauer and Sigmund [1998]. A more general result for two competing species can also be found in section 11.3 of Hirsch et al. [2013].

	$\mathbf{E}_1$	$\mathbf{E}_2$	$\mathbf{x}^*$	Asymptotic Dynamics
$k_1 > \frac{k_2}{a_{21}}, k_2 < \frac{k_1}{a_{12}}$	Stable	Unstable	–	Sole survival of species 1.
$k_1 < \frac{k_2}{a_{21}}, k_2 > \frac{k_1}{a_{12}}$	Unstable	Stable	–	Sole survival of species 2.
$k_1 < \frac{k_2}{a_{21}}, k_2 < \frac{k_1}{a_{12}}$	Unstable	Unstable	Stable	Stable coexistence.
$k_1 > \frac{k_2}{a_{21}}, k_2 > \frac{k_1}{a_{12}}$	Stable	Stable	Unstable	Unstable coexistence. (Only one species survives)

Table 1 – Classification of equilibria in the two-species competitive Lotka-Volterra system, based on the carrying capacities (k_1, k_2) and interaction coefficients (a_{12}, a_{21}) . The coexistence equilibrium $\mathbf{x}^* = \left(\frac{k_1 - a_{12}k_2}{1 - a_{12}a_{21}}, \frac{k_2 - a_{21}k_1}{1 - a_{12}a_{21}} \right)$ exists if and only if the nullclines intersect in $\mathbb{R}_+^2$.

2.3 The Replicator Dynamics

This section is dedicated to another important class of population models: the *replicator dynamics* [Hofbauer and Sigmund, 1998, Nowak, 2006]. Replicator dynamics was first introduced by Taylor and Jonker in 1978 within the emerging field of *evolutionary game theory*, as a mathematical framework to study how populations composed of different types or species evolve over time, based on the fitness advantages (or disadvantages) that arise from their interactions.

More precisely, consider a population composed of n distinct species and a corresponding $n \times n$ *payoff matrix* $A = (a_{ij})$, where each entry a_{ij} represents the benefit obtained by an individual of species i when interacting with an individual of species j . As in the Lotka-Volterra framework, these interaction coefficients drive the dynamics, but here the vector $\mathbf{x} = (x_1, \dots, x_n)$ denotes the *frequency distribution* of the population, that is, now $x_i \in [0, 1]$ is the proportion of individuals of species i , and the sum of all coordinates satisfies $\sum_{i=1}^n x_i = 1$.

The replicator dynamics are governed by the system of ordinary differential equations

$$\begin{cases} x'_i = x_i (f_i(\mathbf{x}) - \Phi(\mathbf{x})), & i = 1, \dots, n \end{cases} \quad (2.2)$$

where

$$f_i(\mathbf{x}) = \sum_{j=1}^n a_{ij}x_j = (A\mathbf{x})_i \quad (2.3)$$

is the *fitness* of species i , and

$$\Phi(\mathbf{x}) = \sum_{j=1}^n x_j f_j(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} \quad (2.4)$$

is the *average fitness* of the population.

Biologically, the fitness $f_i(\mathbf{x}) = \sum_{j=1}^n a_{ij}x_j$ measures the average benefit that an individual of species i receives when randomly interacting with members of the population.

Intuitively, equation (2.2) formalizes the principle of *natural selection*: species whose fitness exceeds the average fitness of the population tend to increase in frequency, while less fit species decline. This is a mathematical embodiment of the idea of “survival of the fittest.”

The next proposition presents the natural space for the study of the replicator dynamics:

Proposition 2. *The simplex*

$$S_{n-1} = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1, x_i \geq 0 \right\}$$

is invariant through the replicator dynamics, i.e., if $\mathbf{x}(0) \in S_{n-1}$, then $\mathbf{x}(t) \in S_{n-1}$ for all t .

Proof. If $\mathbf{x}(t)$ is a solution of the replicator equation (2.2), then the sum of its components $S(t) = \sum_{i=1}^n x_i(t)$ satisfies the ODE:

$$\begin{aligned}
 S'(t) &= \left(\sum_{i=1}^n x_i(t) \right)' \\
 &= \sum_{i=1}^n x_i'(t) \\
 &= \sum_{i=1}^n x_i(t) (f_i(\mathbf{x}) - \Phi(\mathbf{x})) \\
 &= \sum_{i=1}^n x_i(t) f_i(\mathbf{x}) - \Phi(\mathbf{x}) \sum_{i=1}^n x_i(t) \\
 &= \Phi(\mathbf{x}) - \Phi(\mathbf{x}) \cdot S(t) \\
 &= \Phi(\mathbf{x}) (1 - S(t))
 \end{aligned}$$

This implies that if $S(0) = 1$, then by the uniqueness of solutions to ordinary differential equations, it holds $S(t) \equiv 1$ for all t .

Moreover, notice that if $x_i = 0$, then $x_i' = 0$, so if $x_i(0) = 0$, then $x_i(t) \equiv 0$. This further implies that if $x_i(0) > 0$ then, by continuity, $x_i(t) > 0$ for all t .

It follows that if the initial condition satisfies $\mathbf{x}(0) \in S_{n-1}$, then $\mathbf{x}(t) \in S_{n-1}$ for all t . $\square$

Proposition 2 has important dynamical consequences. First, this proposition shows that, although this system is defined on an n -dimensional space, the model-relevant dynamics is restricted to a $(n - 1)$ -dimensional region of the space. That is particularly relevant for the three species replicator (Subsection 2.3.3) since, in that case, the proposition reduces the dynamics to a planar system and, as different orbits can not intersect (recall Section 1.1), the orbits of the replicator dynamics can define regions in the state space (Figure 3).

Moreover, with this result one can easily derive an important property of the replicator dynamics solutions:

Corollary 1. *The solutions of the system (2.2) are defined for all future and past times.*

Proof. The proof follows directly from the fact that both future and past orbits of the replicator dynamics are entirely contained in the compact set $S_{n-1} \subset \mathbb{R}^n$ so, by Theorem 2, it follows that they are defined for all time. $\square$

The equilibria of the replicator equations are characterized by the condition that all *coexisting* species, i.e., all species with positive frequency, must have equal fitness. This follows

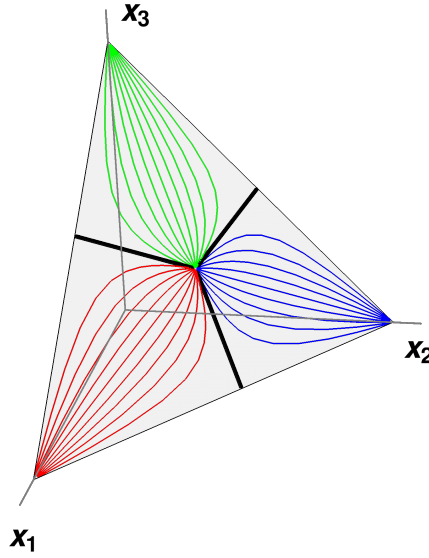


Figure 3 – Illustration of the replicator dynamics in $S_2 \subset \mathbb{R}^3$. The simplex is invariant over the dynamics. For three species, orbits of the replicator dynamics can define regions in the state space. In the illustration, the black orbits split S_2 , dividing the space into regions of attraction for each vertex equilibrium.

directly from the replicator equation (2.2) which implies that if $x_i > 0$ and $x'_i = 0$, then $f_i(\mathbf{x}) = \Phi(\mathbf{x})$. Consequently, if $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ is an equilibrium point then, for any species i with $x_i^* > 0$, the fitness must equal the population average fitness, and thus all such coexisting species must have the same fitness.

In particular, it follows directly from this observation that all vertices of S_{n-1} , that is, the n *single-species states* given by $\mathbf{E}_i = \mathbf{e}_i$ are always equilibrium points of the system. In fact, since at such a point the species i is the *only* population with positive (in fact, total) frequency, the average fitness on vertex $\mathbf{E}_i$ is precisely f_i , and it follows that it is, indeed, an equilibrium.

The following proposition describes a method to adjust the payoff matrix while preserving the underlying dynamics:

Proposition 3. *The replicator dynamics is invariant under the addition of constants to the columns of the payoff matrix A .*

Proof. Let $\tilde{A} = (a_{ij} + c_j)$ be a modified payoff matrix obtained by adding constants $c_1, \dots, c_n \in \mathbb{R}$

to the columns of A . Then, the new fitness functions satisfy:

$$\begin{aligned}\tilde{f}_i(\mathbf{x}) &= (\tilde{A}\mathbf{x})_i \\ &= \sum_{j=1}^n (a_{ij} + c_j) x_j \\ &= \sum_{j=1}^n a_{ij} x_j + \sum_{j=1}^n c_j x_j \\ &= f_i(\mathbf{x}) + \sum_{j=1}^n c_j x_j.\end{aligned}$$

Similarly, the average fitness becomes:

$$\begin{aligned}\tilde{\Phi}(\mathbf{x}) &= \sum_{i=1}^n x_i \tilde{f}_i(\mathbf{x}) \\ &= \sum_{i=1}^n x_i \left(f_i(\mathbf{x}) + \sum_{j=1}^n c_j x_j \right) \\ &= \sum_{i=1}^n x_i f_i(\mathbf{x}) + \sum_{i=1}^n x_i \left(\sum_{j=1}^n c_j x_j \right) \\ &= \Phi(\mathbf{x}) + \left(\sum_{j=1}^n c_j x_j \right) \underbrace{\left(\sum_{i=1}^n x_i \right)}_{=1} \\ &= \Phi(\mathbf{x}) + \sum_{j=1}^n c_j x_j.\end{aligned}$$

Therefore, the fitness difference that drives the dynamics remains invariant:

$$\tilde{f}_i(\mathbf{x}) - \tilde{\Phi}(\mathbf{x}) = \left(f_i(\mathbf{x}) + \sum_{j=1}^n c_j x_j \right) - \left(\Phi(\mathbf{x}) + \sum_{j=1}^n c_j x_j \right) = f_i(\mathbf{x}) - \Phi(\mathbf{x})$$

This implies that the replicator systems defined by A and $\tilde{A}$ are precisely the same. $\square$

2.3.1 The Adjusted Replicator Dynamics

One important variation of the replicator dynamics commonly used in evolutionary game theory is the *adjusted replicator dynamics* [Traulsen et al., 2006], which also appears as *Moran replicator dynamics* on Ferreira and Neves [2020]. In this model, the selection process is still driven by differences between individual fitness and the population's average fitness, but this difference is now normalized by the average fitness itself. More precisely, the adjusted replicator dynamics is defined by the ODE system

$$\begin{cases} x'_i = x_i \frac{f_i(\mathbf{x}) - \Phi(\mathbf{x})}{\Phi(\mathbf{x})} \end{cases} \quad (2.5)$$

This new system can be interpreted as a time reparametrization of the standard replicator equation (2.2), that is, both systems generate the same orbits in the phase space, but trajectories are traversed at different speeds.

Indeed, if $\mathbf{x}(t)$ is a solution of the standard replicator equation (2.2), then $\mathbf{y}(t) = \mathbf{x}(\tau(t))$ is a solution of the adjusted system (2.5), where the reparametrization function $\tau(t)$ satisfies the differential equation

$$\tau'(t) = \frac{1}{\Phi(\mathbf{x}(\tau(t)))} \quad (2.6)$$

Notice, however, both from the definition in (2.5) and the time reparametrization equation (2.6), that the adjusted replicator dynamics is not well-defined when the average fitness $\Phi(\mathbf{x})$ is null. This limitation does not restrict the applicability of the model, since, as shown in Proposition 3, one can always add suitable constants to the columns of the payoff matrix without altering the dynamics. By choosing these constants appropriately, one can ensure that all fitness values $f_i(\mathbf{x})$ are strictly positive throughout the simplex S_{n-1} , thereby guaranteeing that the average fitness $\Phi(\mathbf{x})$ remains strictly positive as well.

2.3.2 Equivalence of Lotka-Volterra and Replicator Dynamics

This subsection presents an important result that connects the two population models introduced so far. Specifically, the *dynamical equivalence* between the general Lotka–Volterra system for $n - 1$ interacting species and the replicator dynamics for n species. Equivalence here means that the orbits in $\mathbb{R}_+^{n-1}$ of a Lotka–Volterra system can be mapped to orbits in S_{n-1} of a replicator equation and vice-versa. The mapping between the two systems is described below, together with the payoff matrix that defines the replicator equations associated with a given Lotka–Volterra system.

Consider a population composed of $n - 1$ species given by the vector $\mathbf{x} = (x_1, \dots, x_{n-1}) \in \mathbb{R}_+^{n-1}$ and add a “fictitious species” with constant population, conventionally chosen as $x_n = 1$.

Define, then, the vector $\mathbf{y} \in \mathbb{R}^n$ by its coordinates as

$$y_i = \frac{x_i}{\sum_{j=1}^n x_j}, \quad (2.7)$$

for $i = 1, \dots, n$. Note that $\sum_{i=1}^n y_i = 1$ and, since, $x_1, \dots, x_n \geq 0$, it follows that $y_i \in [0, 1]$ and $\mathbf{y} \in S_{n-1}$.

If the function $\mathbf{x}(t)$ is a solution of an $(n-1)$ -species Lotka–Volterra system, then the frequency vector $\mathbf{y}(t)$, defined by (2.7) is a solution of the n -species replicator equation (2.2) for a suitable payoff matrix B .

To facilitate the calculations, one can rewrite the standard Lotka–Volterra system in a more convenient form. Recall that the standard form for the i -th species equation presented here is

$$x'_i = r_i x_i \left(1 - \frac{\sum_{j=1}^{n-1} a_{ij} x_j}{k_i} \right).$$

which can be rewritten as

$$x'_i = x_i \left(r_i \underbrace{x_n}_{=1} + \sum_{j=1}^{n-1} \left(-\frac{a_{ij} r_i}{k_i} \right) x_j \right),$$

for $i = 1, \dots, n-1$. This motivates defining the $n \times n$ matrix $\tilde{B} = (b_{ij})$ with the entries

$$\tilde{b}_{ij} = \begin{cases} -\frac{a_{ij} r_i}{k_i}, & \text{for } i, j = 1, \dots, n-1, \\ r_i, & \text{for } i = 1, \dots, n-1 \text{ and } j = n, \\ 0, & \text{for } i = n \text{ and } j = 1, \dots, n. \end{cases}$$

Using this definition, the system becomes

$$x'_i = x_i \left(\sum_{j=1}^n \tilde{b}_{ij} x_j \right) = x_i (\tilde{B}\mathbf{x})_i, \quad (2.8)$$

which is now defined for $i = 1, \dots, n$. Notably, as $b_{nj} = 0$ for $j = 1, \dots, n$ it follows that $x'_n = 0$, which is consistent with the definition of $x_n \equiv 1$.

Computing the time derivative of the function y_i yields

$$\begin{aligned} y'_i &= \left(\frac{x_i}{\sum_{j=1}^n x_j} \right)' \\ &= \frac{x'_i \sum_{j=1}^n x_j - x_i \left(\sum_{j=1}^n x_j \right)'}{\left(\sum_{j=1}^n x_j \right)^2} \\ &= \frac{x_i (\tilde{B}\mathbf{x})_i \sum_{j=1}^n x_j - x_i \sum_{j=1}^n x_j (\tilde{B}\mathbf{x})_j}{\left(\sum_{j=1}^n x_j \right)^2} \\ &= \frac{x_i}{\sum_{j=1}^n x_j} \left((\tilde{B}\mathbf{x})_i - \sum_{j=1}^n \frac{x_j}{\sum_{l=1}^n x_l} (\tilde{B}\mathbf{x})_j \right) \\ &= y_i \left((\tilde{B}\mathbf{x})_i - \sum_{j=1}^n y_j (\tilde{B}\mathbf{x})_j \right). \end{aligned}$$

Now, multiplying and dividing the right-hand side by $\sum_{l=1}^n x_l$, it follows that

$$\begin{aligned} y'_i &= y_i \left(\left(\tilde{B} \frac{\mathbf{x}}{\sum_{l=1}^n x_l} \right)_i - \sum_{j=1}^n y_j \left(\tilde{B} \frac{\mathbf{x}}{\sum_{l=1}^n x_l} \right)_j \right) \left(\sum_{l=1}^n x_l \right) \\ &= y_i \left((\tilde{B}\mathbf{y})_i - \sum_{j=1}^n y_j (\tilde{B}\mathbf{y})_j \right) \left(\sum_{l=1}^n x_l \right). \end{aligned}$$

This expression corresponds to the replicator equation (2.2) with payoff matrix $\tilde{B}$, scaled by the total population size $\sum_{i=1}^n x_i$. Thus, via a time reparametrization, it follows that the curves $\mathbf{y}(t)$ are, indeed, orbits of the replicator dynamics. More precisely, the curve $\mathbf{z}(t) = \mathbf{y}(\tau(t))$ satisfies the replicator equation with payoff matrix $\tilde{B}$, where $\tau(t)$ is the solution of the ODE:

$$\tau'(t) = \frac{1}{\sum_{l=1}^n x_l(\tau(t))}. \quad (2.9)$$

The proof of the converse, i.e, that a solution for the n -species replicator equation induces a solution for the $(n-1)$ -species Lotka-Volterra system, can be found in Hofbauer and Sigmund [1998] (Theorem 7.5.1).

Before proceeding, a small remark on the payoff matrix $\tilde{B}$ is in order. While the matrix $\tilde{B}$ defined above is a natural choice given the structure of the Lotka–Volterra system, for a *competitive* system its first $n-1$ columns are always negative, while the n -th column is positive. In the context of this thesis, however, it is more convenient to work with a *positive* payoff matrix (see Subsection 2.3.1 and Section 2.4). To this end, one can use the modified matrix $B = (b_{ij})$ defined by

$$b_{ij} = \tilde{b}_{ij} + \kappa_j, \quad \kappa_j = 1 - \min_{i=1, \dots, n} \tilde{b}_{ij}. \quad (2.10)$$

By Theorem 3, the replicator dynamics induced by B and $\tilde{B}$ are precisely the same. On the other hand, this construction ensures that the payoff matrix B is strictly positive. Throughout the remainder of this thesis, whenever a Lotka–Volterra system is mapped to a replicator system, this positive payoff matrix must be considered.

The above considerations prove the following result:

Proposition 4 (Equivalence of Lotka-Volterra and Replicator Dynamics).

- If $\mathbf{x}(t)$ is a non-negative solution of the Lotka-Volterra equations with $n-1$ species, then $\mathbf{y}(\tau(t))$ is a solution of the replicator equations with n species and payoff matrix $B = (b_{ij})$ given by (2.10), where $\mathbf{y}$ is given by (2.7) and $\tau(t)$ is a solution of (2.9).

- If $\mathbf{y}(t)$ is a solution of the replicator equations with n species, $y_n(t) \neq 0$ and payoff matrix $B = (b_{ij})$, then $\mathbf{x}(\tau(t))$ is a solution of the Lotka-Volterra equations with $n - 1$ species with $r_i = k_i = b_{in} - b_{nn}$ and interaction parameters $a_{ij} = b_{nj} - b_{ij}$, where $\mathbf{x}$ is given by

$$x_i = \frac{y_i}{y_n}, \quad (2.11)$$

and $\tau(t)$ is a solution of

$$\tau' = \frac{1}{y_n(\tau(t))}.$$

2.3.3 Three Species Replicator Dynamics

To conclude, a phase space representation for the *three-species replicator dynamics* is now presented.

In this particular case, the dynamics occurs in the two-dimensional simplex S_2 , an equilateral triangle in three-dimensional space (see Figure 3). Several geometric representations can be used for this set. Here it will be represented as an equilateral triangle in the *plane* with sides equal to 1 length unit. The three vertices of the triangle are labelled by V_1 , V_2 and V_3 and represent the population states in which all cells are of the same clone as the vertex label. The sides of the triangle represent the states in which one of the clones is absent. For example, the states in which only clones 1 and 2 are present in the population are represented by points in the $V_1 V_2$ side.

To locate the points in this set, this thesis adopts trilinear coordinates, in which the point (x_1, x_2, x_3) of S_2 is represented by the point at the intersection of lines parallel to the sides of the triangle, as shown in Fig. 4. Notably, this parametrization can be regarded of a rotation applied to the x_2 axis in the projection of the simplex in the $x_1 x_2$ plane. Precisely, given $(x_1, x_2, x_3) \in S_2$, its trilinear coordinates are given by the map

$$(x_1, x_2, x_3) \mapsto (x_1 + \cos(\pi/3)x_2, \sin(\pi/3)x_2) \quad (2.12)$$

The *interior* of S_2 , i.e. the set of points in S_2 , but not in its boundary, will be denoted by $\overset{\circ}{S}_2$.

Unlike in the case of *competitive Lotka-Volterra* presented in the last chapter, even this planar setup of the replicator dynamics can present several different dynamical behaviors. In fact, there are 47 distinct phase portraits that may arise depending on the magnitudes of the entries in the payoff matrix [Bomze, 1983]. However, the multitude of global dynamics of the

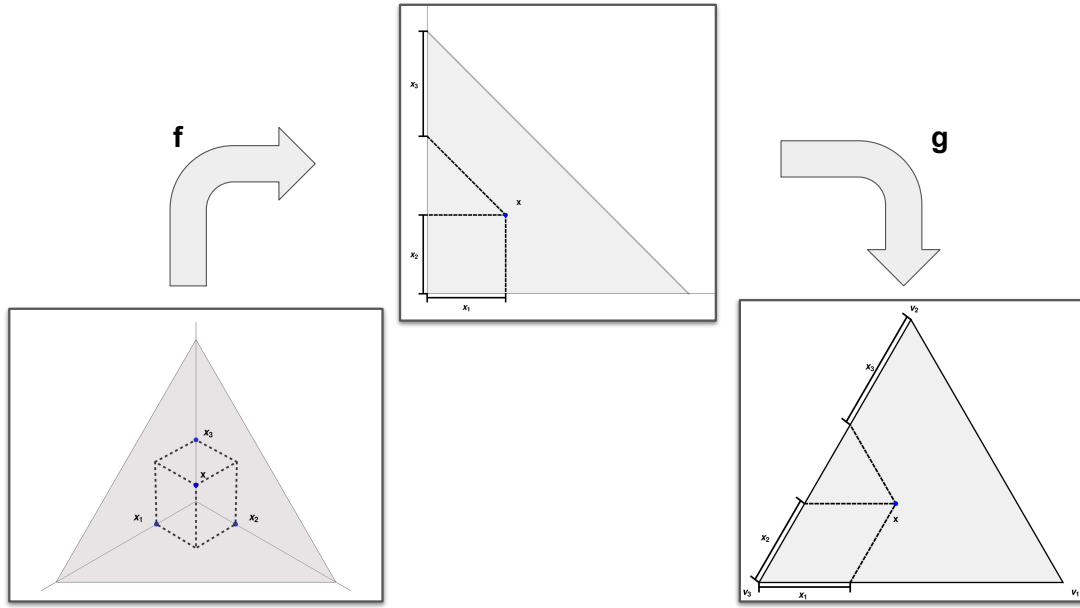


Figure 4 – Parametrization of a point of the simplex in trilinear coordinates. Here f represents the projection in the x_1x_2 plane and g is a suitable rotation (see (2.12)).

three-species model will not play any role in the context of this thesis. Instead, the core idea that will be crucial throughout the text is that *clones with above-average fitness increase in frequency, while clones with below-average fitness decrease in frequency.*

This fundamental mechanism implies that, unless the system is already at equilibrium, there will always exist at least one clone whose frequency is increasing and at least one whose frequency is decreasing, since, unless all fitness are equal, there must always be (at least) one clone above and one below the population's average fitness.

2.4 The Moran Process

A well-suited stochastic model to use in conjunction with the adjusted replicator dynamics is the *Moran process*. Originally devised by P.A.P. Moran in 1958 to describe populations with two types of individuals [Moran, 1958, Nowak, 2006], the Moran process models the dynamics of a *finite population* composed of $n \geq 2$ species. The state of the population is defined as

$$\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}_+^n,$$

where X_i denotes the number of individuals of species i . As in the replicator dynamics, the interactions between individuals are governed by a payoff matrix $A = (a_{ij})$, whose entries describe the benefits obtained from interactions among different individuals depending on their species [Ferreira and Neves, 2020].

The fitness of an individual of type i is defined as

$$f_i(\mathbf{X}) = \frac{a_{ii}(X_i - 1) + \sum_{j \neq i} a_{ij}X_j}{N - 1},$$

where $N = \sum_{i=1}^n X_i$ is the population size. Unlike in the replicator dynamics case, this model requires that the payoff matrix elements are positive. This is necessary, because fitnesses will have a probabilistic interpretation below and, therefore, they must be positive. Notice that the definition of fitness for the replicator equations (2.3) is recovered when $N \rightarrow \infty$. As will be shown shortly, these two processes are tightly related.

The Moran process is a discrete-time Markov chain where at each time step one individual reproduces and one individual dies. The newly born individual is of the same species of its parent (i.e., no mutation occurs) and replaces the individual that died. The population size N remains thus constant in time, although the numbers of individuals of each type may vary. Both reproducing and dying individuals are randomly drawn by independent lotteries and, in the formulation used in the thesis, a uniform death lottery is considered, while the reproduction lottery is such that the probability of an individual being drawn is proportional to its fitness. Furthermore, the lotteries considered here are drawn simultaneously, so it is possible that the same individual reproduces and dies in the same time step.

Precisely, the probability that an individual of type i is drawn for reproduction is

$$\alpha_i(\mathbf{X}) = \frac{X_i f_i(\mathbf{X})}{\sum_{j=1}^n X_j f_j(\mathbf{X})}.$$

On the other hand, every individual has the same probability of dying. Accordingly,

$$\beta_i(\mathbf{X}) = \frac{X_i}{N}$$

is the probability that a type i individual dies.

Both from the definition of the process and from the structure of the reproduction and death probabilities, one can observe that the Moran process possesses exactly n *absorbing states*, namely the single-species configurations $\mathbf{E}_i = N\mathbf{e}_i$, where the entire population consists of individuals of species i . Indeed, if $\mathbf{X}^k$ denotes the population state on step k and $\mathbf{X}^k = N\mathbf{e}_i$, then it holds

$$\alpha_j(\mathbf{X}^k) = \frac{0 \cdot f_j(\mathbf{X}^k)}{N \cdot f_i(\mathbf{X}^k)} = 0, \quad \beta_j(\mathbf{X}^k) = \frac{0}{N} = 0,$$

for all $j \neq i$ and

$$\alpha_i(\mathbf{X}^k) = \frac{N \cdot f_i(\mathbf{X}^k)}{N \cdot f_i(\mathbf{X}^k)} = 1, \quad \beta_i(\mathbf{X}^k) = \frac{N}{N} = 1,$$

for $j = i$. Thus, from this state it must occur both the reproduction and the death of a single individual of species i , which implies that $\mathbf{X}^{k+1} = \mathbf{X}^k$. Hence, the state is absorbing.

For all other states, the process has both a positive probability of increasing and decreasing the population of any extant and non-fixated species.

Proposition 5. *Let $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{N}^n$ with $\sum X_i = N \geq 2$. Then for any i such that $0 < X_i < N$, it follows $\alpha_i(\mathbf{X}) \in (0, 1)$, and $\beta_i(\mathbf{X}) \in (0, 1)$.*

Proof. Since $0 < X_i < N$, one immediately obtains:

$$0 < \beta_i(\mathbf{X}) = \frac{X_i}{N} < 1.$$

Also, as $0 < X_i < N$, there exists some $j \neq i$ with $0 < X_j < N$. Since all entries of A are positive, the fitness of species i satisfies:

$$f_i(\mathbf{X}) = \frac{a_{ii}(X_i - 1) + \sum_{l \neq i} a_{il}X_l}{N - 1} \geq \frac{a_{ii}(X_i - 1) + a_{ij}X_j}{N - 1} > 0.$$

Similarly, $f_j(\mathbf{X}) > 0$.

Then, the reproduction probability satisfies:

$$0 < \alpha_i(\mathbf{X}) = \frac{X_i f_i(\mathbf{X})}{\sum_{l=1}^n X_l f_l(\mathbf{X})} \leq \frac{X_i f_i(\mathbf{X})}{X_i f_i(\mathbf{X}) + X_j f_j(\mathbf{X})} < 1.$$

□

This result immediately implies that, for any species i such that $0 < X_i^k < N$, the process has strictly positive probability both of increasing and decreasing X_i^k . Precisely,

$$P(X_i^{k+1} = X_i^k + 1 \mid 0 < X_i^k < N) = \alpha_i(\mathbf{X}^k) \cdot (1 - \beta_i(\mathbf{X}^k)) > 0,$$

$$P(X_i^{k+1} = X_i^k - 1 \mid 0 < X_i^k < N) = (1 - \alpha_i(\mathbf{X}^k)) \cdot \beta_i(\mathbf{X}^k) > 0.$$

Thus, as long as a species is neither extinct nor fixated, the process has a positive probability of changing its population size in either direction.

2.4.1 The Moran Process and the Replicator Dynamics

Although the Moran process is a *discrete-time stochastic process*, it exhibits a strong relationship with the adjusted replicator dynamics, a *continuous-time deterministic process* [Benaïm and Weibull, 2003, Traulsen et al., 2006, Ferreira and Neves, 2020]. In a broader context, Lemma 1 in [Benaïm and Weibull, 2003] shows that for any *finite* time span $T > 0$, the

probability that a realization of the Moran process at step $k < NT$ “is far” from the deterministic adjusted replicator solution (defined with the *same payoff matrix*) evaluated at time k/N tends to zero as $N \rightarrow \infty$.

More precisely, let $\mathbf{X}^k$ denote the state at step k of a realization of the Moran process, and let $\Psi(t, \mathbf{x}_0)$ be the flow of the adjusted replicator dynamics, both with n species and defined by the same payoff matrix and initial condition $\mathbf{x}_0$. Define the distance function $D : \mathbb{R}_+ \times S_{n-1} \rightarrow \mathbb{R}_+$ as the maximum distance between the Moran process and the flow during the time interval $[0, T]$:

$$D(T, \mathbf{x}_0) = \max_{0 \leq k \leq NT} \left\| \frac{\mathbf{X}^k}{N} - \Psi\left(\frac{k}{N}, \mathbf{x}_0\right) \right\|_\infty,$$

where $\|\mathbf{x}\|_\infty = \max_i |x_i|$ is the L_∞ norm in $\mathbb{R}^n$.

Lemma 1 from [Benaïm and Weibull, 2003] then states the following bound:

Proposition 6 (Adapted from Lemma 1 in [Benaïm and Weibull, 2003]). *There exists a constant $c > 0$ such that for any $\varepsilon > 0$, $T > 0$, and N sufficiently large,*

$$P(D(T, \mathbf{x}_0) \geq \varepsilon \mid \mathbf{X}^0 = N\mathbf{x}_0) \leq 2ne^{-\varepsilon^2 c N}$$

holds for all $\mathbf{x}_0 \in S_{n-1}$.

Notice that, for a *fixed payoff matrix*, $\varepsilon > 0$ and $T > 0$, this result holds **uniformly** over all initial conditions in the simplex S_{n-1} , depending only on the total population size N . Precisely, given any $\varepsilon, \delta, T > 0$, there exists some N_0 such that if $N > N_0$ then the Moran and the adjusted replicator trajectories are, with probability *at least* $1 - \delta$, ε -close for at least NT steps, regardless of the initial condition $\mathbf{x}_0$. In this sense, the Moran process *converges* to the adjusted replicator dynamics as $N \rightarrow \infty$, or equivalently, that the Moran process is the stochastic counterpart of the adjusted replicator model.

Notice, however, that Proposition 6 *does not* imply that any particular trajectory of the Moran process remains close to the adjusted replicator dynamics counterpart *indefinitely*. Indeed, as it will be discussed on Subsection 2.4.3, the stochastic nature of the Moran process guarantees that, with probability one, all trajectories eventually reach *fixation*. That is, regardless of the initial state $\mathbf{X}^0 \in S_{n-1}$, for each realization of the process there will *almost surely* exist some $i \in \{1, \dots, n\}$ and some $K > 0$ such that $\mathbf{X}^k = N\mathbf{e}_i$ for all $k > K$. In other words, even if it takes a long time, the Moran process, with probability 1, ultimately fixates on a single species i even if the adjusted replicator flow stays bounded away from the simplex vertices (e.g., it converges to an interior equilibrium).

2.4.2 The Stochastic Lotka-Volterra Process

As seen in Subsection 2.3.1, the adjusted replicator dynamics with n species is simply a time reparametrization of the classic replicator dynamics. Furthermore, it is shown in Subsection 2.3.2 that the replicator dynamics is equivalent to a Lotka–Volterra system with $n - 1$ species. Consequently, the Lotka–Volterra system is also equivalent to the adjusted replicator dynamics, under a suitable definition of the payoff matrix.

More precisely, given a solution $\mathbf{x}(t)$ of the $(n - 1)$ -species Lotka–Volterra system (2.1), one can define a fictitious constant population $x_n(t) \equiv 1$, and the frequency vector $\mathbf{y}$ as in (2.7).

Consider the time-reparametrized trajectory $\mathbf{z}(t) = \mathbf{y}(\tau(t))$, where $\tau(t)$ satisfies the differential equation

$$\tau'(t) = \frac{1}{\Phi(\mathbf{y}(\tau(t))) \left(\sum_{j=1}^n x_j(\tau(t)) \right)}, \quad (2.13)$$

it follows that $\mathbf{z}(t)$ is a solution of the adjusted replicator dynamics (2.5) with payoff matrix $B = (b_{ij})$ as defined in (2.10).

Indeed, using the fact that $\mathbf{y}$ satisfies the replicator equation multiplied by the sum $\sum_{i=1}^n x_i$, one obtains

$$\begin{aligned} z_i'(t) &= [y_i(\tau(t))]' \\ &= y_i'(\tau) \cdot \tau'(t) \\ &= y_i(\tau) [f_i(\mathbf{y}(\tau)) - \Phi(\mathbf{y}(\tau))] \cdot \left(\sum_{j=1}^n x_j(\tau) \right) \cdot \frac{1}{\Phi(\mathbf{y}(\tau)) \left(\sum_{j=1}^n x_j(\tau) \right)} \\ &= y_i(\tau) \cdot \frac{f_i(\mathbf{y}(\tau)) - \Phi(\mathbf{y}(\tau))}{\Phi(\mathbf{y}(\tau))} \\ &= z_i(t) \cdot \frac{f_i(\mathbf{z}(t)) - \Phi(\mathbf{z}(t))}{\Phi(\mathbf{z}(t))}, \end{aligned}$$

which is precisely the adjusted replicator equation.

With this relationship in hand, a straightforward definition of a stochastic counterpart to the Lotka–Volterra dynamics using the Moran process follows. Namely, consider the Moran process $\mathbf{Z}^k$ governed by the payoff matrix B defined in Equation (2.10) with a population of size $N = \sum_{i=1}^n Z_i^0$. Define the stochastic process

$$\mathbf{X}^k = \frac{\mathbf{Z}^k}{Z_n^k}, \quad (2.14)$$

where Z_n^k is the n -th coordinate of $\mathbf{Z}^k$. Naturally, this is not well defined when $Z_n^k = 0$.

From Proposition 6, it follows that for a fixed time span $T > 0$, if the total population N is large enough, the trajectory of $\mathbf{Z}^k$ is likely to closely follow that of the adjusted replicator dynamics solution $\mathbf{z}(k/N)$ for all $k < NT$, given that $\mathbf{Z}^0 = \mathbf{z}(0)$.

Recalling that $\mathbf{z}(t) = \mathbf{y}(\tau(t))$, $\mathbf{y}(t) = \frac{\mathbf{x}(t)}{\sum_{i=1}^n x_i(t)}$ and $x_n \equiv 1$, compute

$$\frac{\mathbf{z}(k/N)}{z_n(k/N)} = \frac{\mathbf{y}(\tau(k/N))}{y_n(\tau(k/N))} = \frac{\frac{\mathbf{x}(\tau(k/N))}{\sum_{i=1}^n x_i(\tau(k/N))}}{\frac{x_n(\tau(k/N))}{\sum_{i=1}^n x_i(\tau(k/N))}} = \mathbf{x}(\tau(k/N)). \quad (2.15)$$

Thus, the process $\mathbf{X}^k$ closely approximates, with large probability, the Lotka–Volterra trajectory $\mathbf{x}(\tau(k/N))$ for sufficiently large N and finite time T , establishing it as a stochastic analogue to the deterministic Lotka–Volterra system.

Some important comments are in order. First, it is crucial to notice that the total population N of the induced Moran process $\mathbf{Z}^k$ *does not* correspond to the total population size of the original Lotka–Volterra system $\mathbf{x}(t)$. In fact, unlike in the Moran process, it is generally not true that the total population size of a Lotka–Volterra system, $\sum_{i=1}^{n-1} x_i(t)$, remains constant, unless the system is at an equilibrium point. In that sense a suitable interpretation for the population size N of the stochastic counterpart to a certain Lotka–Volterra system is no more as a real population size, but as an *effective* population size related to the size of the stochastic fluctuations in the stochastic counterpart.

Next, the effect of choosing the fictitious population $x_n \equiv 1$ when mapping the LV system to the replicator dynamics must also be discussed. If the *real* total population $\sum_{j=1}^{n-1} x_j$ is much larger than 1 (e.g. thousands of cells in a tissue), then the frequency of the fictitious species, given by $y_n = 1 / \sum_{j=1}^n x_j$ is close to zero. In this case, the corresponding state $\mathbf{y} \in S_{n-1}$ lies near the boundary face $\{\mathbf{y} \in S_{n-1} \mid y_n = 0\}$.

Conversely, if $\sum_{j=1}^{n-1} x_j \ll 1$ (e.g., x_i is measured in millions of cells), then $y_n \approx 1$, and $\mathbf{y}$ is close to the vertex V_n of the simplex. In this regime, the fictitious population dominates, and the contributions from the real species are minimal.

Neither scenario is desirable, as the associated Moran process $\mathbf{Z}^k$ would, in both cases, lie dangerously close to an absorbing face or state. In particular, if the fictitious population is negligible ($y_n \approx 0$), the Moran process approaches absorption at the boundary face where $Z_n^k = 0$. This case is especially problematic: since the stochastic Lotka–Volterra process is defined by dividing the vector $\mathbf{Z}^k$ by Z_n^k , the extinction of species n leads to a division by zero, rendering the stochastic Lotka–Volterra process ill-defined.

In order to avoid these degeneracies, one possible solution, which was adopted in this thesis, is to rescale the original Lotka–Volterra populations by a factor $\alpha > 0$, simultaneously adjusting the carrying capacities by the same factor. More precisely, one can define new variables $\tilde{x}_i = \alpha x_i$ and $\tilde{k}_i = \alpha k_i$ for all $i = 1, \dots, n-1$. This change corresponds to a linear rescaling of the population units and leaves the dynamics of the system invariant, since the governing equations are still the same.

2.4.3 Fixation in the Moran Process

As discussed in Section 1.2 stochastic population models with a finite state space, in which all states are either transient or absorbing, share a common fate in the long run: random fluctuations eventually drive the system into an absorbing state where only one species remains. The result below formalizes this property for the n -species Moran process:

Corollary 2 (Fixation in the Moran Process). *Let $\{\mathbf{X}^k\}_{k \in \mathbb{N}}$ be a Moran process with n species and total population size N . Then, with probability one, every trajectory eventually reaches fixation of a single species; that is, the process is absorbed in one of the states $\mathbf{E}_i = N\mathbf{e}_i$ for some $i \in \{1, \dots, n\}$.*

Proof. As shown in Section 2.4, the only absorbing states of the n -species Moran process are the single-species states $\mathbf{E}_i = N\mathbf{e}_i$. Define the set of absorbing states as $S = \{\mathbf{E}_i \mid i = 1, \dots, n\}$.

From Proposition 5, it follows that if a species is neither extinct nor fixated, then it has a positive probability of both increasing and decreasing in number. Hence, from any state $\mathbf{X}^*$ with $0 < X_i^* < N$, there is a non-zero probability of reaching the absorbing state $\mathbf{E}_i$, namely, by increasing population X_i one unit at each step.

On the other hand, once a trajectory reaches $\mathbf{E}_i$, it cannot return to $\mathbf{X}^*$, since that is an absorbing state. That is, $\rho_{\mathbf{E}_i \mathbf{X}^*} = 0 < 1$ while $\rho_{\mathbf{X}^* \mathbf{E}_i} > 0$, and thus, by Theorem 3, $\mathbf{X}^*$ is transient.

Since all mixed-species states are transient and the only absorbing states are those in S , applying Theorem 4 yields that, with probability one, every trajectory is eventually absorbed in S . □

Corollary 2 highlights a fundamental distinction between stochastic and deterministic models: while deterministic systems such as Lotka–Volterra and replicator dynamics may admit stable coexistence equilibria, implying the indefinite persistence of multiple species, the Moran process ultimately leads to the fixation of a single species with probability one, even though,

given a sufficiently large population size for the Moran model, both dynamics follow, with large probability, a close course for arbitrarily long finite periods.

This apparent contradiction is, in fact, the key to understanding why some Moran processes fixate faster than others. Precisely, the agreement of the stochastic dynamics with the underlying deterministic flow can either delay or accelerate fixation, depending on the qualitative behavior of the deterministic model:

- When the deterministic system exhibits a *stable coexistence equilibrium*, the stochastic process tends to remain near this equilibrium for extended periods. The deterministic flow continually pulls the system back toward the coexistence state, mitigating the impact of random fluctuations and slowing down the fixation process.
- Conversely, if the deterministic system has no coexistence equilibrium or the equilibrium is *unstable*, then the deterministic flow itself drives the system toward the boundary of the state space. In such cases, stochastic fluctuations are enhanced by this trend, accelerating fixation.

In this way, the deterministic model not only shapes the expected transient behavior of the stochastic system, but also gives insight into the timescale over which fixation is likely to occur. Specifically for the Moran process with two species ($n = 2$), when the corresponding replicator dynamics exhibits an *unstable coexistence equilibrium*, or when *no coexistence equilibrium* exists, the mean time to fixation is generically of order $N \log(N)$. In contrast, when there exists a *stable coexistence equilibrium*, the mean time to fixation grows exponentially with the population size N . These results appeared in Antal and Scheuring [2006], but their proof is not entirely rigorous. The complete proof is given in Pires [2022]. In practical terms, this means that, if both species coexist, fixation occurs on timescales so long that it is effectively unobservable, both in empirical systems and in simulations, except possibly for very small populations.

As it will be shown in Chapter 5, the interplay between deterministic stability and stochastic robustness can be extremely relevant when applying models to clinical contexts.

PART II

**POPULATION MODELS IN CANCER
TREATMENT**

This part of the thesis is based on the paper “On the design and stability of cancer adaptive therapy cycles: deterministic and stochastic models”, published in *Mathematical Biosciences* in July, 2025.

3 INTRODUCTION

Despite continuous research efforts and significant advancements in its treatment over the years, cancer remains one of the major causes of premature death worldwide [Sung et al., 2021]. Drug resistance, in particular, has proven to be a formidable challenge in this field, affecting every known drug and being responsible for over 90% of cancer-related deaths among chemotherapy patients [Aktipis et al., 2011, Bukowski et al., 2020].

An evolutionary mechanism based on the concept of competitive release explains the evolution of resistance [Aktipis et al., 2011]. Due to the highly heterogeneous environment, resistant cancer cells are often already present in the tumor prior to any treatment, but are naturally suppressed by a fitter population of sensitive cells in the absence of the drugs. However, as treatment proceeds, strong selective pressure is applied to the tumor, creating an environment favoring the rise of a resistant population. Therefore, if the treatment does not weaken the tumor enough so that non-cancerous cells are able to control it by themselves, a new, predominantly resistant, tumor emerges. This is illustrated in the top row of Fig. 5.

Borrowing concepts from pest control in agriculture, *adaptive therapy* [Gatenby et al., 2009] proposes a “control over elimination” approach, exploiting the cost of resistance and the naturally competitive environment of the tumor to manage its growth, allowing patients to coexist with the disease for extended time periods. To achieve this, personalized treatment routines are designed to include “treatment holidays”, i.e. periods without any drugs, in order to allow the sensitive cell population to recover and suppress the emergence of resistant cells. An illustration for that can be seen in the bottom row of Fig. 5. Treatments with adaptive therapy have already begun to be translated into preclinical trials for various cancer types, showing promising results and retarding tumor progression [West et al., 2023, Gedye and Navani, 2022, Zhang et al., 2023].

From a mathematical standpoint, several *deterministic models* have been proposed over the years, encompassing both single-drug [Zhang et al., 2017, West et al., 2018, Wang et al.,

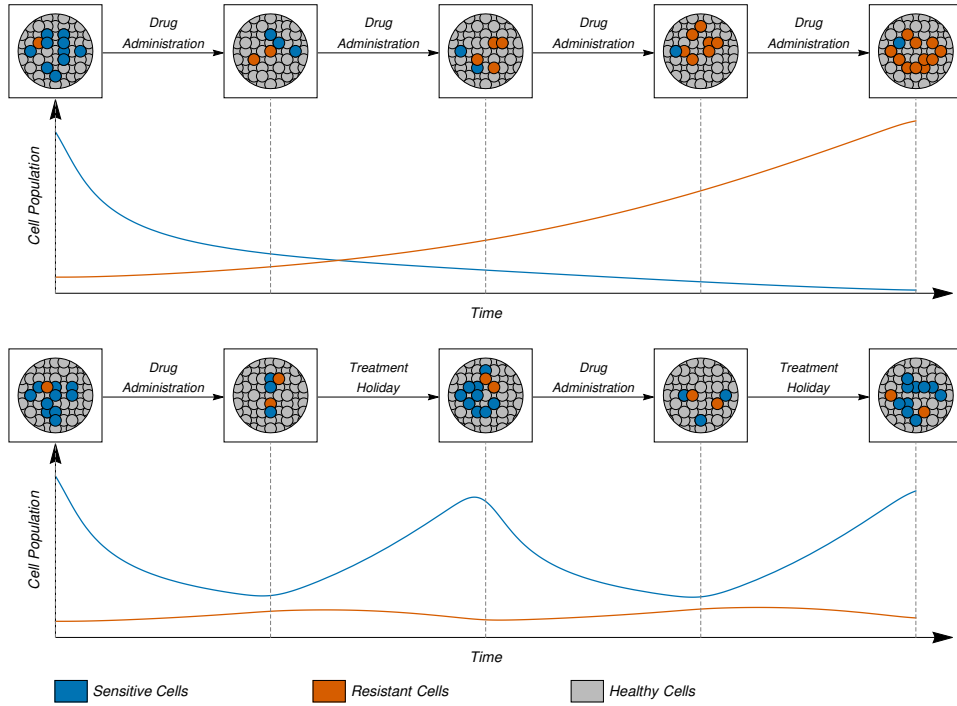


Figure 5 – Representation of a continuous treatment (above) and a treatment routine including treatment holidays (below). While continuous treatment leads to the evolution of resistant cells due to competitive release, intermittent/adaptive therapy with drug holidays allows the competition with sensitive cells to suppress the growth of resistant cells, aiming to keep the tumor under control instead of eradicating it.

2021] and multidrug [West et al., 2019, 2020, Dua et al., 2021] adaptive therapy schemes. While the simplicity of single-drug models greatly facilitates their analysis, most real-life treatments involve multiple drugs, making multidrug models more representative of reality. Furthermore, the use of multiple therapies allows for greater control over the tumor by targeting different cell populations at different times, facilitating the creation of *cyclic treatment routines*. In theory, these routines restore exactly the same tumor composition before and after their application, and can be repeated *ad infinitum*, arguably being translated into clinical treatments with a very long time before relapse [West et al., 2020].

Throughout this thesis, the deterministic dynamics of cell populations within a tumor are mathematically modeled by a system of ordinary differential equations containing several parameters, and the effects of different drugs are dealt with by different values of the parameters. The construction of a cyclic routine amounts to finding *exact* durations for the treatment with each drug, so that the solution of the ODE system returns to the same point where it was at the beginning of the cycle. One potential approach is to geometrically identify intersections of the orbits corresponding to each treatment phase. However, this method becomes infeasible for

systems with many cell types, as it requires finding intersections in high-dimensional spaces. Even in planar contexts, the existence of such intersection is not guaranteed without mathematical assumptions. Throughout this thesis we establish formal, yet biologically motivated, conditions that guarantee the existence of such cycles in two different deterministic models: Lotka-Volterra equations [Hofbauer and Sigmund, 1998, Murray, 2002] and adjusted replicator dynamics [Taylor and Jonker, 1978, Hofbauer and Sigmund, 1998, Nowak, 2006]. With the notable exception of a very recent work [Wang and Lei, 2025], it seems that mathematical conditions for existence of cyclic treatments have been overlooked.

Naturally, the theoretical treatment designed with these models is a simplification of what is attainable in practice. Indeed, the *precise* return to an *exact* tumor condition is *not*, and it should not be, a realistic and determinant factor for clinical success. Therefore, after successfully designing an ideal, theoretical cyclic routine, another critical consideration is its *stability*, i.e., how closely the actual solutions mirror the intended ones in case of small perturbations. In practice, models - even highly accurate ones - are approximations. In this context, a stable cycle implies that small deviations due to model limitations, parameter uncertainty, measurement noise, or intrinsic biological variability, will not accumulate over time. Instead, both model solutions and their real-world counterparts may remain close to the intended cyclic behavior, leading to sustained tumor control.

We show here that stability of deterministic routines also plays an important role when considering *stochastic models* for tumors. Recent studies [Dua et al., 2021, Park and Newton, 2023], motivated by the fact that cell populations are always finite and subject to random fluctuations, particularly relevant within the context of low tumor burden, considered robustness of cyclic deterministic routines produced within a deterministic model (adjusted replicator dynamics) when this model is replaced by its *stochastic counterpart*, the Moran process [Moran, 1958, Nowak, 2006]. They observed the collapse of deterministic cycles in the stochastic counterparts due to the stochastic fluctuations that quickly deviate the cycles.

Of course, due to the very stochastic nature of the Moran process, one should not expect that a stochastic routine derived from a deterministic cycle will return exactly to its starting point. We can expect it to return, with large probability, *close* to its starting point, if the population size N is large enough. If the deterministic cycle is stable, we will see that when it is realized in the stochastic context, the tumor composition remains close to the intended deterministic solution for many cycles. On the contrary, *if the cycle is unstable* as in [Dua et al., 2021, Park

and Newton, 2023], we expect that the stochastic tumor composition will break down after a few cycle repetitions. We will show that this breakdown, already observed in [Dua et al., 2021, Park and Newton, 2023], is here correlated with the instability of the underlying deterministic cycle. Besides the Moran process, which is a stochastic counterpart of the adjusted replicator dynamics, we also devise and study a stochastic counterpart for the Lotka-Volterra dynamics. These results highlight the idea that designing stable deterministic cycles is not merely a mathematical exercise, but a foundational step toward achieving long-term tumor control under realistic, noisy conditions. From a biological standpoint, it suggests that adaptive therapy protocols may be calibrated to promote sustained tumor containment even in the face of inevitable fluctuations in treatment and tumor behavior.

This part of the thesis is organized as follows. In Chapter 4 we formally define cyclic routines in light of deterministic models and provide, for the Lotka-Volterra and Replicator models, biologically relevant conditions that guarantee the existence of such routines starting at a given point. We show that, in fact, infinitely many of these cycles exist and exhibit a simple algorithm to design them. We introduce then (Section 4.3) the concept of stability and give examples of stable and unstable routines for both models. In Chapter 5, we present stochastic routines and compare their behavior with the associated deterministic routines, examining the roles of population size and cycle stability in the expected behavior when random fluctuations are considered. This part is closed by concluding remarks in Chapter 6.

4 DETERMINISTIC ROUTINES

Let us consider the deterministic dynamics of a tumor with n cell types subject to m different treatment phases. Each phase will be modelled by a different ODE system, so that we have m ODE systems

$$x' = F_1(x), \dots, x' = F_m(x)$$

with respective *flows* $\Psi_i(t, x)$, $i = 1, 2, \dots, m$. We will denote by $T_1, \dots, T_m$ the treatment phases durations.

A routine composed by the above treatments, starting with the tumor composition x^* , is *cyclic* if the sequential application of the m treatments with the corresponding durations returns to the original state, i.e., if

$$(\Psi_m(T_m, \cdot) \circ \dots \circ \Psi_1(T_1, \cdot))(x^*) = x^*.$$

Let the initial point x^* and durations $T_1, \dots, T_{m-2}$ for the $m-2$ first treatment phases be given. Define

$$y^* = (\Psi_{m-2}(T_{m-2}, \cdot) \circ \dots \circ \Psi_1(T_1, \cdot))(x^*)$$

as the tumor state after the first $m-2$ treatment phases. A putative algorithm for constructing a cyclic routine consists of determining the durations T_{m-1} and T_m of the last two phases such that the future orbit of y^* under the $(m-1)$ -th treatment intersects the past orbit of x^* under the m -th treatment, i.e.

$$\Psi_{m-1}(T_{m-1}, y^*) = \Psi_m(-T_m, x^*).$$

A pseudo-code for computing the times T_{m-1} and T_m that close the cycle is given in Algorithm 1.

Algorithm 1 Determine Cycle Times

1: **procedure** DETERMINEFINALPOINT($\mathbf{x}^*$, Params, Times)

```

2:   initialPhases  $\leftarrow$  Length(Times)
3:    $\mathbf{y}^* \leftarrow \mathbf{x}^*$ 
4:   for  $i = 1$  to initialPhases do
5:       sol  $\leftarrow$  SolveODE(ModelField(Params[ $i$ ]),  $\mathbf{y}^*$ , [0, Times[ $i$ ]])
6:        $\mathbf{y}^* \leftarrow$  sol(Times[ $i$ ])
7:   end for
8:   return  $\mathbf{y}^*$ 
9: end procedure

10: procedure DETERMINEFINALTIMES( $\mathbf{x}^*$ , Params, Times)
11:    $\mathbf{y}^* \leftarrow$  DetermineFinalPoint( $\mathbf{x}^*$ , Params, Times)
12:   totalPhases  $\leftarrow$  Length(Params)
13:   solFuture  $\leftarrow$  SolveODE(ModelField(Params[totalPhases - 1]),  $\mathbf{y}^*$ , [0,  $T_{\max}$ ])
14:   solPast  $\leftarrow$  SolveODE(ModelField(Params[totalPhases]),  $\mathbf{x}^*$ , [ $-T_{\max}$ , 0])
15:    $(T_{m-1}, T_m) \leftarrow$  Intersect(solFuture( $T_{m-1}$ ), solPast( $-T_m$ ))
16:   return  $(T_{m-1}, T_m)$ 
17: end procedure

```

Here, Params is a list of m parameter sets corresponding to the m treatment phases, and Times is a list of the known durations for the first $m - 2$ phases. The SolveODE method refers to a numerical ODE solver, and ModelField returns the model's vector field given its parameters. A large time parameter $T_{\max}$ is used as a practical substitute for ∞ and $-\infty$ in lines 12 and 13. As seen in Section 2.2, the ODE solution for backwards time may not be defined. This case should be handled carefully.

The routine Intersect denotes a curve intersection or root-finding procedure that returns the first parameter pair (T_{m-1}, T_m) such that $\text{solFuture}(T_{m-1}) = \text{solPast}(-T_m)$. Of course, in general these times and intersection may not exist, and in such cases it is impossible to build cyclic routines from the given point with the desired treatment durations. In the following subsections, we will present, in two planar ODE models with $m = 3$, conditions that guarantee the existence of cyclic routines determined by the above putative algorithm, transforming it in a *de facto* algorithm. In fact, we will show that whenever a cyclic routine starting at a given point exists, an infinite number of such routines can be designed, which enables the possibility of using optimization techniques to choose a best routine according to some criterion.

4.1 Cyclic Routines in the LV Model

We start our studies with a tumor composed of two cell types, type 1 (*sensitive* cells) and type 2 (*resistant* cells) treated with two different drugs, D_1 and D_2 . We assume that type 1 cells are affected by both drugs while type 2 cells are resistant to D_1 but sensitive to D_2 , and consider routines composed of three phases:

- **Phase 1** is a treatment holiday, so no drug is used;
- During **phase 2** drug D_1 is used;
- During **phase 3** drug D_2 is used.

Let x_1 and x_2 stand for the population sizes of cell types 1 and 2. We model this scenario using *competitive Lotka-Volterra* equations. The dynamics in any phase $i = 1, 2, 3$ is described by the ODE systems

$$\begin{cases} x'_1 = r_{1,i}x_1 \left(1 - \frac{x_1 + a_{12,i}x_2}{k_{1,i}} \right) \\ x'_2 = r_{2,i}x_2 \left(1 - \frac{a_{21,i}x_1 + x_2}{k_{2,i}} \right) \end{cases}, \quad (4.1)$$

where the parameters values depend on the phase.

As already observed on Section 2.2, there are four possible dynamical scenarios for these systems, depending on the relative position of their nullclines: stable and unstable coexistence if the nullclines intersect in the first quadrant, or, otherwise, the sole survival of a single (*dominant*) species regardless of the initial condition.

In all four cases populations of both cell types grow in the region below the two nullclines (Figure 2). We define R_1 as the region below the two nullclines in the first phase of the treatment, i.e. when no drugs are administered, see Figure 6A. Points in R_1 can be interpreted as tumors where populations of both cell types are still increasing, and are natural candidates to start a treatment routine.

If we start our routine with $x^* \in R_1$, then the orbit $\Psi_1(t, x^*)$ will generically escape region R_1 at a time T_1^* . Taking $T_1 < T_1^*$, we have $y^* = \Psi_1(T_1, x^*) \in R_1$, and y^* is above and to the right of x^* (Figs. 6B and 6C).

During the second phase, we administer drug D_1 , which is effective against type 1, but not against type 2 cells. It is reasonable, then, to assume that the flow diagram of the second phase is of the type as in Figure 2D, i.e. we would observe the *extinction of the sensitive cells* and

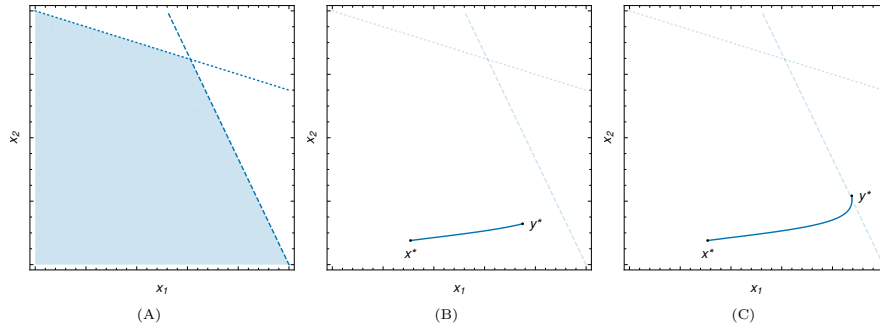


Figure 6 – Region R_1 (A) and the orbit of x^* during the first phase of the treatment (blue curve) with different duration times T_1 and their respective endpoints $y^* = \Psi_1(x^*, T_1)$ (B,C).

survival only of the resistant cells if this treatment were used for infinite time. Geometrically, this implies that the x_1 nullcline must be below the x_2 nullcline in the first quadrant during the second phase. We define R_2 as the region below the x_2 nullcline during second phase of the treatment. In other words, R_2 is the region in which population x_2 increases during the administration of D_1 , see Fig. 7A. We further assume that treatment with D_1 *weakens sensitive cells* by decreasing their carrying capacity and their aggressiveness, but *does not affect* the same quantities for the resistant ones. It follows that $R_1 \subset R_2$. We see that every tumor in R_2 would eventually become resistant to D_1 if treated only with this drug.

The above assumptions about the second phase of the treatment may be mathematically summarized as:

- **Extinction of the sensitive cell population:**

$$k_{1,2} < \frac{k_{2,2}}{a_{21,2}} \quad \text{and} \quad \frac{k_{1,2}}{a_{12,2}} < k_{2,2} . \quad (4.2a)$$

- **Weakening of sensitive cells:**

$$k_{1,2} < k_{1,1} \quad \text{and} \quad a_{21,2} \leq a_{21,1} . \quad (4.2b)$$

- **Resistant cells are not affected:**

$$k_{2,2} = k_{2,1} \quad \text{and} \quad a_{12,2} = a_{12,1} . \quad (4.2c)$$

If $y^* = \Psi_1(T_1, x^*)$ as before, the above assumptions imply that the future orbit of y^* , i.e. the curve $\Psi_2(t, y^*)$, $t > 0$, must cross a vertical line through x^* at some point w^* at some time $T_2^* > 0$. More yet, $w^* = \Psi_2(T_2^*, y^*)$ must be *above* x^* . We define R^* as the region limited by the curves $\gamma_1 = \Psi_1([0, T_1], x^*)$, $\gamma_2 = \Psi_2([0, T_2^*], y^*)$ and the line segment between the points x^* and w^* . The above curves and regions are illustrated at Figs. 7B and 7C.

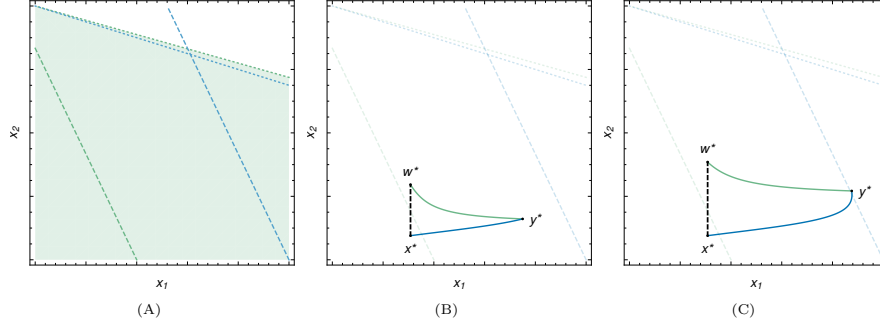


Figure 7 – Regions R_2 (A) and R^* limited by the curves connecting the points x^* , y^* and w^* , considering different choices for T_1 , the first phase duration (B, C). The orbits for the first and second phases are shown in blue and green, respectively.

Finally, during the third phase we administer drug D_2 , which is effective against both cell types. Naturally, such a treatment makes sense if tumors that are still developing do not become larger than they would without any treatment. In geometrical terms, this means that both nullclines of the third phase of the treatment are below the nullclines of the first phase of the treatment. In terms of parameters, see Figure 2, this imposes the following constraints:

• x_1 nullcline constraints:

$$k_{1,3} < \min \left\{ k_{1,1}, \frac{k_{2,1}}{a_{21,1}} \right\} \quad \text{and} \quad \frac{k_{1,3}}{a_{12,3}} < \min \left\{ k_{2,1}, \frac{k_{1,1}}{a_{12,1}} \right\}. \quad (4.3a)$$

• x_2 nullcline constraints:

$$k_{2,3} < \min \left\{ k_{2,1}, \frac{k_{1,1}}{a_{12,1}} \right\} \quad \text{and} \quad \frac{k_{2,3}}{a_{21,3}} < \min \left\{ k_{1,1}, \frac{k_{2,1}}{a_{21,1}} \right\}. \quad (4.3b)$$

We define R_3 as the region where both cell types *decrease* under the treatment with D_2 , i.e., the region above both nullclines in phase 3, and also $R = R_3 \cap R_1$. For a tumor composition $x^* \in R$, populations for both cell types increase in treatment holidays and both decrease when the tumor is treated with drug D_2 .

If $x^* \in R_3$ and we move *backward* in time, then both cell types *increase* indefinitely. This suggests a steepness condition necessary to guarantee existence of a cyclic routine: the past orbit of x^* for the third phase is *steeper* than the future orbit of x^* for the first phase (i.e., the orange curve is, in the region close to x^* , above the blue curve as in Fig. 8).

In terms of derivatives, the steepness condition may be written as

$$\frac{\Psi'_1(0, x^*)_2}{\Psi'_1(0, x^*)_1} < \frac{-\Psi'_3(0, x^*)_2}{-\Psi'_3(0, x^*)_1}. \quad (4.4)$$

Although providing valuable geometric information, the inequality above may be rewritten in order to convey more biological information. In fact, $\Psi'_1(0, x^*)_j$ is the growth rate of cell type j in the absence of drugs, whereas $-\Psi'_3(0, x^*)_j$ represents the death rate of cell type j under treatment with D_2 . Rewriting the above inequality as

$$\frac{\Psi'_1(0, x^*)_2}{-\Psi'_3(0, x^*)_2} < \frac{\Psi'_1(0, x^*)_1}{-\Psi'_3(0, x^*)_1} \quad (4.5)$$

reveals that it becomes a *cost of resistance* condition stating that *resistance of type 2 cells to drug D_1 comes at a cost either in cell proliferation in the absence of treatment, or in sensitivity to drug D_2 .*

If the cost of resistance condition (4.5) holds, we choose $x^* \in R$, and parameters satisfy biological conditions (4.2) and (4.3), then we can clearly design a cycle starting from x^* with duration T_1 for the first phase, subject only to $T_1 < T_1^*$. Indeed, for small values of $t > 0$, (4.5) implies that the curve $\Psi_3(-t, x^*)$ enters region R^* . As both coordinates of $\Psi_3(-t, x^*)$ increase and tend to infinity as t increases (see Subsection 2.2), the past orbit $\Psi_3(-t, x^*)$, $t > 0$, will eventually leave R^* . Thus it must intersect the boundary of R^* either at γ_1 (Fig. 8C) or γ_2 (Fig. 8B), forming a cyclic treatment. Durations T_2 and T_3 are such that $\Psi_3(-T_3, x^*) = \Psi_2(T_2, y^*)$.

If we want to avoid the atypical situation depicted in Fig. 8C, in which $T_2 = 0$ and only drug D_2 is used, instead of taking $T_1 < T_1^*$, it suffices to take $T_1 < \tilde{T}_1$, where $\tilde{T}_1$ is such that $\Psi_3(-t, x^*)$, $t > 0$, is strictly above $\Psi_1(t, x^*)$ for $t \in [0, \tilde{T}_1]$.

With the above considerations, we have proved the following theorem:

Theorem 5. *Assume that a tumor with two cell types under the three treatment phases as specified at the beginning of this subsection can be modeled by equations (4.1) with the treatments imposing constraints (4.2) and (4.3) on the parameters. If $x^* \in R$ and the cost of resistance condition (4.5) holds, then there exists $\tilde{T}_1 > 0$ such that for every $T_1 \in (0, \tilde{T}_1)$ we can find unique $T_2, T_3 > 0$ such that*

$$\Psi_3(T_3, \Psi_2(T_2, \Psi_1(T_1, x^*))) = x^*.$$

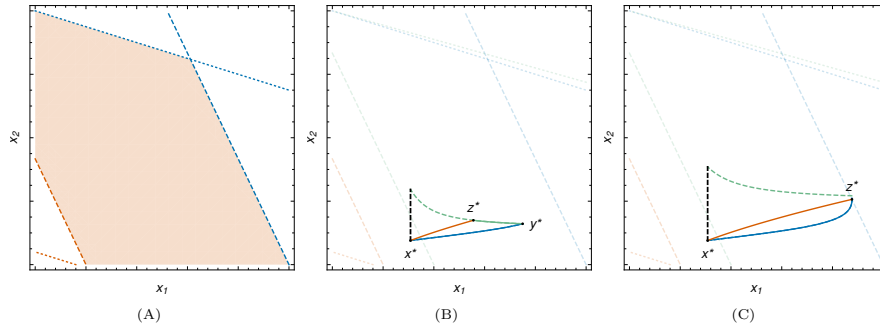


Figure 8 – Region $R = R_1 \cap R_3$ (A) and cycles designed between the points x^* , y^* and z^* considering different times for the first phase (B, C). Notice (C) that a cycle not using drug D_1 is possible. The orbits for the first, second and third phases are shown in blue, green and orange, respectively.

From a biological perspective, Theorem 5 states that if two *effective* drugs are available to treat a tumor, one of which the tumor is partially resistant to, then adaptive therapy cycles can indeed be constructed for tumors that are actively growing. This holds as long as resistance carries a cost, either by reducing proliferative capacity in the absence of treatment or by increasing sensitivity to the more effective drug. Moreover, infinitely many cycles, with different phase durations, can be generated for the same initial tumor composition, opening the door to optimization across multiple criteria, such as duration under specific treatments, total drug usage, or average tumor burden. The above proof also shows that it is possible to construct a cycle using only drug D_2 in one phase and no drug in the other phase (on/off treatment), as illustrated in Fig. 8C.

4.2 Cyclic Routines in the Adjusted Replicator Model

Notice that if we consider the Lotka-Volterra model proposed on Subsection 4.1, the tumor could be controlled using only drug D_2 . On the other hand, as a continuous use of this drug would maintain lower populations of the type 1 and type 2 cells, it would possibly lead to the release of a third cell type, resistant to this treatment.

We take this into account by proposing a tumor model with three kinds of cancer cells: besides cells of types 1 and 2 as before, we assume here existence of type 3 cells, resistant to D_2 , but sensitive to D_1 . We can still search for a routine composed by a treatment holiday (duration T_1), followed by administration of D_1 (duration T_2) and administration of D_2 (duration T_3).

Since we are now on a three-dimensional context, building a cycle using our geometric

approach in the Lotka-Volterra model is no longer feasible. We will proceed by modeling this problem with the *adjusted replicator dynamics* (see Section 2.3):

$$\begin{cases} x'_i = x_i \frac{f_i(x) - \Phi(x)}{\Phi(x)}, & i = 1, 2, 3. \end{cases} \quad (4.6)$$

As previously observed, as now $\mathbf{x} \in S_2$ represent *population fractions*, we find ourselves in a *planar context*. Similar to [Dua et al., 2021, Park and Newton, 2023], we decompose the elements of the pay-off matrix as

$$1 + w_i a_{ij},$$

where 1 is a baseline fitness for all cell types, the *interaction coefficient* w_i is a non-negative parameter representing the intensity of the interactions for cell type i , and the a_{ij} are positive values representing the benefit that a cell of type i experiences when interacting with a cell of type j . In terms of matrix $A = (a_{ij})$ and of the interaction coefficients, the fitness of cell type i is

$$f_i(x) = 1 + w_i (Ax)_i. \quad (4.7)$$

With assumptions $a_{ij} > 0$ and $w_i \geq 0$, the expression above implies that the smallest fitness value is equal to 1 if $w_i = 0$. As proposed in [Dua et al., 2021], we consider a common interaction coefficient $w_0 > 0$ for all three cell types during the treatment holiday, when no drugs are used. The effects of the drugs on the cells will be introduced by decreasing the interaction coefficient of a cell type whenever this type is sensitive to the drug. Concretely, we assume that when treated with the largest concentration $C_i = 1$ of a drug to which it is sensitive, a cell type will have its interaction coefficient equal 0. Taking into account that cells of type 1 are sensitive to both drugs D_1 and D_2 , that cells of type 2 are sensitive to D_2 , but not to D_1 , and cells of type 3 are sensitive to D_1 , but not to D_2 , we have

$$w_1 = w_0(1 - C_1 - C_2) \quad (4.8a)$$

$$w_2 = w_0(1 - C_2) \quad (4.8b)$$

$$w_3 = w_0(1 - C_1). \quad (4.8c)$$

To complete the specification of the dynamics, we only have to specify the values of C_1 and C_2 at each phase:

- **Phase 1** is a treatment holiday: $C_1 = C_2 = 0$.
- During **phase 2** only drug D_1 is used: $C_1 = 1, C_2 = 0$.
- During **phase 3** only drug D_2 is used: $C_1 = 0, C_2 = 1$.

With the above described dynamics and hypotheses to be described, we may state a theorem analogous to Theorem 5 on the existence of cyclic treatments. The proof follows ideas similar to the proof of the analogous result, but is a bit more technical.

First, similar to hypothesis (4.5) in the Lotka-Volterra model, it is natural to assume a cost of resistance and expect that sensitive cells, in the absence of any drugs, are fitter than any resistant cells. We suppose that this is the case, at least when we start the treatment. Without loss of generality, we may suppose also that cells of type 2 are fitter than cells of type 3. Our hypothesis on the pay-off matrix is thus that there exists a region $R_1 \subset S_2$ such that, on phase 1 ($w_1 = w_2 = w_3 = w_0$),

$$f_1(x) > f_2(x) \geq f_3(x) \quad (4.9)$$

for all $x \in R_1$ and that the initial state x^* of the tumor is such that $x^* \in R_1$.

In terms of matrix A , this condition imposes that $(Ax)_1 > (Ax)_2 \geq (Ax)_3$ which implies, from a biological perspective, that during the treatment holiday sensitive cells experience, on average, the greatest benefit from the tumor environment, followed by type 2 and type 3 cells. We need not to impose any further hypothesis on the relative sizes of the pay-off matrix entries.

In Fig. 9 we illustrate an example of a cyclic treatment for the adjusted replicator dynamics described, along with some concepts to be used in the proof of existence for such cycles.

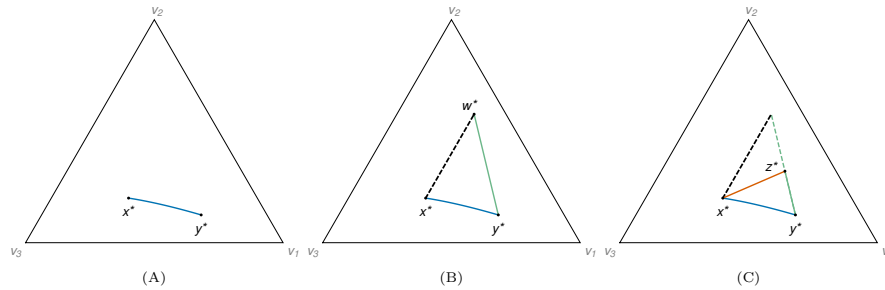


Figure 9 – Cyclic routine designed between the points x^* , y^* and z^* , with treatment orbits for the first (blue), second (green) and third (orange) phases. In panel (B) we can see the region R^* limited by the curves connecting x^* , y^* and w^* .

Our result is

Theorem 6. Consider a tumor with three cell types under the three treatment phases as specified at this subsection. Assume that the tumor can be modeled by equations (4.6), that the pay-off matrix elements obey $a_{ij} > 0$, and the interaction coefficients are subject to (4.8). Finally, suppose that the initial tumor composition x^* is in a region R_1 where (4.9) holds. Then there exists $\tilde{T}_1 > 0$ (possibly ∞) such that for every $T_1 \in (0, \tilde{T}_1)$ we can find unique $T_2, T_3 > 0$ with

$$\Psi_3(T_3, \Psi_2(T_2, \Psi_1(T_1, x^*))) = x^*.$$

Proof. Let x^* be any initial condition in the region R_1 such that (4.9) holds, and $\tilde{T}_1$ be the time it takes for the future orbit $\Psi_1(t, x^*)$ to reach the border of R_1 . If the orbit never reaches the border in finite time, we define $\tilde{T}_1$ to be infinite.

We choose the duration $T_1 > 0$ of the first treatment phase to be arbitrary, but smaller than $\tilde{T}_1$. Ordering in (4.9) is thus preserved up to time T_1 . This implies that during the first phase, the fraction x_1 of sensitive cells increases, while the fraction x_3 of the less fit cells decreases within the tumor.

In fact, by (4.9),

$$\Phi(x^*) = x_1^* f_1(x^*) + x_2^* f_2(x^*) + x_3^* f_3(x^*) < (x_1^* + x_3^* + x_3^*) f_1(x^*) = f_1(x^*).$$

Therefore

$$\Psi'_1(0, x^*)_1 = x_1^* \frac{f_1(x^*) - \Phi(x^*)}{\Phi(x^*)} > 0,$$

showing that the sensitive cell proportion is increasing when we start the treatment. An analog argument proves that the x_3 proportion is decreasing at the first phase.

We define $y^* = \Psi_1(T_1, x^*)$, i.e. the tumor composition at the end of the first phase. By the above arguments,

$$y_1^* > x_1^* \quad \text{and} \quad y_3^* < x_3^*,$$

i.e., the points y^* is closer to the $V_1 V_2$ side and further from the $V_2 V_3$ side of S_2 when compared to the initial condition x^* , as illustrated in Fig. 9A.

During the second phase of the treatment, drug concentrations are $C_1 = 1$ and $C_2 = 0$. Therefore, by (4.8), we have $w_1 = w_3 = 0$ and, consequently, $f_1 \equiv f_3 \equiv 1$. On the other hand, as x_2 is resistant to the applied drug D_1 , we also have $w_2 = w_0 > 0$ so

$$f_2(x) = 1 + w_0(Ax)_2 > 1$$

for all $x \in \overset{\circ}{S}_2$. Then, as shown for the first phase, the x_2 fraction increases while the x_1 and x_3 fractions decrease along the second phase. In particular, this implies that, as $t \rightarrow \infty$, $\Psi_2(t, y^*)_2 \rightarrow 1$ and both $\Psi_2(t, y^*)_1 \rightarrow 0$ and $\Psi_2(t, y^*)_3 \rightarrow 0$.

It follows that the future orbit of y^* under Ψ_2 must intercept the parallel to the V_2V_3 side through x^* at a finite time T_2^* . Let $w^* = \Psi_2(T_2^*, y^*)$ be this intersection point, see Fig. 9B. Since $w_1^* = x_1^*$ and, as the proportion of cells of type 3 decreases on phase two, $w_3^* < y_3^* < x_3^*$ it follows that

$$w_2^* = 1 - w_1^* - w_3^* > 1 - x_1^* - x_3^* = x_2^*.$$

Therefore, the point w^* lies between the initial condition x^* and the side V_1V_2 of S_2 . It follows that a region R^* , similar to the one defined in the context of the Lotka-Volterra model, can be defined, see Fig. 9B. Precisely, R^* is the region bounded by the orbit section between x^* and y^* , the orbit section between y^* and w^* and the line segment between x^* and w^* .

During the third phase, we administer the drug D_2 and, once again, it is natural to expect an increase and eventual fixation of type 3 cells, the only ones that are resistant to D_2 . This implies that the *past* orbit $\Psi_3(-t, x^*)$ tends to the face V_1V_2 side of the simplex as $t \rightarrow \infty$, with *strictly increasing* fractions x_1 and x_2 .

Again, in analogy with the construction of a cyclic treatment in the Lotka-Volterra model, all we need in order to build a cycle here is to guarantee that the past orbit $\Psi_3(-t, x^*)$ enters region R^* . In fact, since this curve tends, as $t \rightarrow \infty$, to the V_1V_2 side of S_2 , it must leave R^* , forming a cycle (Fig. 9C).

Contrary to the proof of Theorem 5, the curve $t \mapsto \Psi_3(-t, x^*)$ indeed enters the region R^* without any further assumptions.

To see that, notice that $\frac{\Psi_1'(0, x^*)_2}{\Psi_1'(0, x^*)_1}$ is the slope of the *projection* of the flow of the first phase at x^* onto the x_1x_2 plane. Analogously for $\frac{-\Psi_3'(0, x^*)_2}{-\Psi_3'(0, x^*)_1}$ and the third phase. The past orbit of x^* will enter region R^* if and only if its projection onto the x_1x_2 plane enters the projection of R^* onto the same plane. The steepness condition (4.4) is exactly the condition for the latter. If we prove it, the proof of Theorem 6 is completed.

As $f_1(x^*) > f_2(x^*)$ we have

$$\frac{\Psi_1'(0, x^*)_2}{\Psi_1'(0, x^*)_1} = \frac{x_2^*(f_2(x^*) - \Phi(x^*))}{x_1^*(f_1(x^*) - \Phi(x^*))} < \frac{x_2^*}{x_1^*}.$$

Meanwhile, on phase 3, we have $f_1 \equiv f_2 \equiv 1$, so

$$\frac{-\Psi_3'(0, x^*)_2}{-\Psi_3'(0, x^*)_1} = \frac{x_2^*(1 - \Phi(x^*))}{x_1^*(1 - \Phi(x^*))} = \frac{x_2^*}{x_1^*}.$$

We conclude then, as stated, that

$$\frac{\Psi'_1(0, x^*)_2}{\Psi'_1(0, x^*)_1} < \frac{-\Psi'_3(0, x^*)_2}{-\Psi'_3(0, x^*)_1},$$

proving the theorem. $\square$

From a biological perspective, Theorem 6 states that if two drugs are available to treat a tumor that is *partially resistant to both of them*, then adaptive therapy cycles can be constructed if sensitive cells are naturally fitter, but resistant ones dominate during the application of the drugs. Once again, infinitely many cycles can be designed for the same initial tumor composition, with different durations for the phases.

4.3 Cycle stability

In the previous subsections, we established conditions under which a cyclic treatment routine can be designed, i.e., one in which the tumor composition returns exactly to its initial state after each cycle. Naturally, such precision is an idealization. Any real-world therapy that depends on exact recurrence is bound to fail due to biological variability, experimental uncertainties and other reasons. However, existence of exact cycles means that return after each cycle to a tumor composition *close* to the initial could be feasible in practice. This leads to a key question: how closely can a realistic implementation of an adaptive treatment follow the idealized cycle, and for how long? In mathematical terms, these concerns are intimately tied to the stability of the cycles within the underlying dynamical system.

Roughly speaking, a cyclic routine will be *asymptotically stable* if orbits of perturbed initial conditions converge to the theoretical cycle, if the perturbation is sufficiently small. Conversely, a cyclic routine will be *unstable* if there exist solutions to the system starting as close as desired to the initial condition of the cycle which orbits flee away from the cycle. Precisely, similar to the stability definition in deterministic systems, we define:

Definition 1. A cyclic routine

$$\Psi(\mathbf{x}) = (\Psi_m(T_m, \cdot) \circ \Psi_1(T_1, \cdot))(\mathbf{x})$$

starting on $\mathbf{x}^*$ is

- *asymptotically stable* if there exists a neighborhood $V \ni \mathbf{x}^*$ such that $\Psi^n(\mathbf{x}) \rightarrow \mathbf{x}^*$ as $n \rightarrow \infty$;

- *unstable if there exists a neighborhood $V^* \ni \mathbf{x}^*$ such that for all neighborhoods $V_1 \subset V^*$ there exists at least one $\mathbf{x} \in V_1$ and $n \in \mathbb{N}$ such that $\Psi^n(\mathbf{x}) \notin V^*$;*

In our numerical explorations with the cycles of the previous subsections for Lotka-Volterra and adjusted replicator dynamics, we found both asymptotically stable and unstable cycles, for both dynamics. For examples of asymptotically stable cycles, see Figs. 10a and 10c, in which orbits for the cycles are shown, along with orbits of perturbed initial conditions that are attracted to them. For examples of unstable cycles, we refer to Figs. 10b and 10d. In these, we also show orbits of initial conditions very close to the cycle, but quickly going far from it. Parameter values producing the examples in Fig. 10 are given in Table 2 for the Lotka-Volterra model and in Table 3 for the adjusted replicator model.

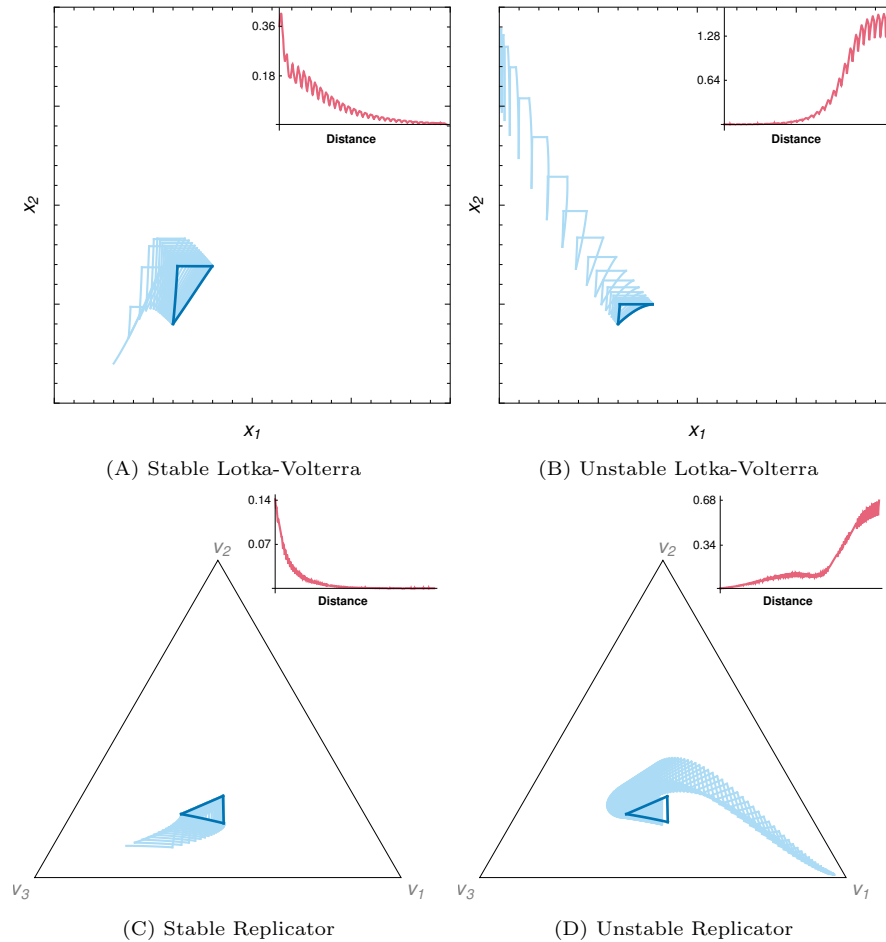


Figure 10 – In dark blue, plots of orbits of treatment cycles created with the parameter values in Tables 2 and 3. In light blue, the orbits of perturbed initial conditions. In the top-right corner, the distance between the designed and the disturbed plots through time. The distance between the treatment initial condition and the perturbed one was taken as ≈ 0.0014 for the unstable Lotka-Volterra routine and ≈ 0.024 for the unstable replicator routine.

Lotka-Volterra Cyclic Routines	
Common Parameters: $x^* = (0.6, 0.4)$ $r_{1,1} = 0.278; r_{1,2} = 0.278 \cdot 10^{-3}; r_{1,3} = 0.278 \cdot 10^{-3}$ $r_{2,1} = 0.665; r_{2,2} = 0.665 \cdot 10^{-3}; r_{2,3} = 0.665 \cdot 10^{-3}$ $k_{1,1} = 2; k_{1,2} = 2 \cdot 10^{-3}; k_{1,3} = 10^{-2}$ $k_{2,1} = 2; k_{2,2} = 2; k_{2,3} = 2 \cdot 10^{-3}$	
Stable Parameters: $T_1 = 2; T_2 \approx 1.706; T_3 \approx 1.535$ $a_{12,i} = 0.5; a_{21,i} = 0.9$	Unstable Parameters $T_1 = 3; T_2 \approx 1.207; T_3 \approx 0.419$ $a_{12,i} = 1.5; a_{21,i} = 1.9$

Table 2 – Specifications for Lotka-Volterra cyclic routines. Parameter values, except for the unstable aggressiveness, were based on the values given in [Zhang et al., 2017]. The carrying capacities were scaled for a simpler transition to the adjusted replicator model, see 2.4.2, and the growth rates were scaled so that T_1 , T_2 and T_3 have the same order of magnitude.

Stability of solutions for Lotka-Volterra systems with periodic coefficients is a well-studied topic, see e.g. [Xia and Han, 2009] and references therein. In fact, Theorem 3.1 in that paper states conditions that guarantee that a competitive periodic Lotka-Volterra system has a unique periodic solution which is also *globally* asymptotically stable. Roughly speaking, the conditions are that the coefficients in the Lotka-Volterra system are *continuous* periodic functions with the same period, that the related time-averaged Lotka-Volterra constant coefficients system has a globally asymptotic coexistence equilibrium, such as in Figure 2A, and that the minimum intraspecific competition is larger than the maximum interspecific competition. However Theorem 3.1 in [Xia and Han, 2009] does not strictly apply to the Lotka-Volterra cycles whose existence is described in Subsection 4.1, because the coefficients in our equations are functions with jump discontinuities at the times in which one phase is replaced by another.

We did not find in the literature any results similar to Theorem 3.1 in [Xia and Han, 2009] for the adjusted replicator dynamics, but in our numerical explorations on adjusted replicator cycles built as in the preceding section, we also found both asymptotically stable and unstable cycles.

Based on our numerical explorations (Fig. 10, Tables 2 and 3), both for Lotka-Volterra and adjusted replicator dynamics, we state the following:

Conjecture 1. *Let a tumor treatment dynamics be described by systems*

$$\left\{ \begin{array}{l} x' = F_1(x), \\ x' = F_2(x), \\ x' = F_3(x) \end{array} \right.$$

with $F_i(x)$ as in (4.1) or (4.6), and durations $T_1, T_2, T_3 > 0$ calculated in order to describe a cyclic treatment routine.

Consider the averaged system

$$\begin{cases} x' = \frac{T_1 F_1(x) + T_2 F_2(x) + T_3 F_3(x)}{T_1 + T_2 + T_3} \end{cases}.$$

Then:

- If the averaged system has a globally asymptotically stable coexistence equilibrium, the designed cycle is asymptotically stable;
- If the average system has an unstable coexistence equilibrium, the designed cycle is unstable.

The above conjecture proposes a link between the stability of the averaged system and that of the designed cycle. It suggests that even if a particular phase exhibits unstable behavior, the influence of subsequent phases promoting stability could stabilize the overall cycle. In terms of tumor biology, Conjecture 1 offers a crucial insight into designing stable cyclic protocols: the choice of drugs, dosages, and timing must be such that, while each treatment phase (drug 1, drug 2, holiday) induces a distinct tumor environment (selection of one type or another, coexistence), their time average should foster coexistence in the long-term average treatment to induce a stable cyclic behavior. Given that standard maximum tolerated dose (MTD) treatments often lead to the survival of resistant clones, and treatment holidays encourage the re-emergence of sensitive populations, achieving a balance that promotes average coexistence seems plausible.

Adjusted Replicator Cyclic Routines	
Common Parameters: $x^* = (0.3, 0.2, 0.5)$ $w_0 = 0.1$	
Stable Parameters: $T_1 = 10; T_2 \approx 5.759; T_3 \approx 7.010$ $A = \begin{bmatrix} 1 & 1.8 & 1.8 \\ 0.8 & 0.8 & 1.1 \\ 0.8 & 1.1 & 0.8 \end{bmatrix}$	Unstable Parameters $T_1 = 10; T_2 \approx 5.287; T_3 \approx 6.866$ $A = \begin{bmatrix} 1 & 1.8 & 1.8 \\ 0.5 & 1.1 & 1.3 \\ 0.5 & 0.8 & 1.2 \end{bmatrix}$

Table 3 – Specifications for adjusted replicator cyclic routines. All parameter values, except for the stable matrix, were based on the values given in [Dua et al., 2021]. Due to the difference of our equation (4.7) with their analog equation, all entries in our unstable matrix are 1 less than theirs. Given the same initial condition and durations, the dynamics are the same.

It should be noted that, if proven correct, Conjecture 1 would imply that slight changes not only in the initial conditions, but in the system as a whole would not change the observed

outcome of the therapy. We stress that this refers both to stable and unstable cycles but, naturally, with different implications: if a designed cycle is asymptotically stable, then therapies applied with similar conditions will have a result close to the designed, while if a designed cycle is unstable, even slight changes in the therapy can lead to dramatically different outcomes. Both behaviors are illustrated in Fig. 10.

In summary, we state that designing appropriate adaptive therapy courses using a deterministic model requires not only identifying cyclic routines, but also accounting for their stability regarding perturbations inherent to the translation to clinical practice. As we will see in the following, deterministic stability is also important if we choose a stochastic model.

5 STOCHASTIC ROUTINES

In this chapter, we define and discuss the role of *stochastic routines* in the design of adaptive therapy schemes. Analogously to deterministic routines, a stochastic routine is determined by the law that governs its dynamics at each time step. Formally, suppose we have m distinct treatments, each modeled by a time-homogeneous Markov chain with transition probabilities

$$P(X^{k+1} = Y \mid X^k = X) = \pi_i(X, Y), \quad i \in \{1, \dots, m\}.$$

Here, X^k denotes the state of the tumor at time step k and π_i is the transition probability for treatment i .

A stochastic routine that applies treatment i for K_i consecutive steps, before switching to the next treatment in a fixed cycle, can be described by a *time-inhomogeneous* Markov chain $\{X^k\}_{k \geq 0}$ with transition probabilities

$$P(X^{k+1} = Y \mid X^k = X) = \pi_{\lambda(k)}(X, Y),$$

where the *treatment schedule index* $\lambda(k)$ is given by

$$\lambda(k) = i \quad \text{if} \quad \left(\sum_{j=1}^{i-1} K_j \right) \leq (k \bmod K) < \left(\sum_{j=1}^i K_j \right),$$

and $K = \sum_{j=1}^m K_j$ is the total number of steps of the cycle.

By this construction, $\{X^k\}$ is generally *not* a time-homogeneous Markov chain, because the transition probabilities depend explicitly on the step index k in addition to the current state X^k . However, if we consider the induced process

$$\mathcal{X}^n := X^{nK}, \quad n \in \mathbb{N},$$

which models the state of the system at the *end* of n full treatment cycle, then $\{\mathcal{X}^n\}$ is a time-homogeneous Markov chain. This follows from the fact that the evolution over one full

cycle can be represented by a fixed transition matrix

$$\Pi_{\text{cycle}} = \Pi_m^{K_m} \cdots \Pi_2^{K_2} \Pi_1^{K_1},$$

where Π_i is the transition matrix associated with treatment i , and $\Pi_i^{K_i}$ denotes its K_i -step transition matrix. Thus,

$$P(\mathcal{X}^{n+1} = Y \mid \mathcal{X}^n = X) = [\Pi_{\text{cycle}}]_{X,Y}$$

is independent of the number of cycles n .

In the remainder of this chapter, we focus on stochastic routines in which each treatment's effect is modeled via a Moran process. Specifically, we will examine those Moran processes that serve as stochastic counterparts to deterministic dynamics of the adjusted replicator equation or the Lotka–Volterra (LV) model.

5.1 Stochastic Replicator Routines

As discussed on Section 2.4 well-suited stochastic model to use in conjunction with the *adjusted replicator dynamics* is the *Moran process*. A *stochastic routine* may be defined for the Moran process by simply specifying its pay-off matrix and selection coefficients at each time step. Furthermore, given a deterministic routine we can define an *associated* stochastic routine by stating that the parameter values at time step $\tau(k)$ correspond to the deterministic parameter values at time k/N . We may expect that, for large values of N , the associated stochastic and deterministic routines will remain *close* for extended periods. But even if a deterministic routine is a cycle, i.e. its orbit returns *exactly* to its initial point at the end of the cycle, we cannot expect that the corresponding stochastic routine does exactly the same. Fig. 11 illustrates closeness between a deterministic routine and its associated stochastic routine, both for the stable and the unstable cycles specified in Table 3.

The example realizations of the cycles in Fig. 11 were run with a very large population size $N = 100,000$ in order to showcase that the Moran process is an accurate stochastic counterpart for the adjusted replicator dynamics. In Fig. 12 we use instead smaller values of N and we realize the same process 1,000 times, as if we were using the same therapy in a cohort of 1,000 individuals. As expected, the final state after one cycle *spreads* around the initial state (marked with an $\times$) and the spread, measured by the standard deviation of the results, is larger in simulations with smaller population sizes. At this point, the stability of the deterministic cycle

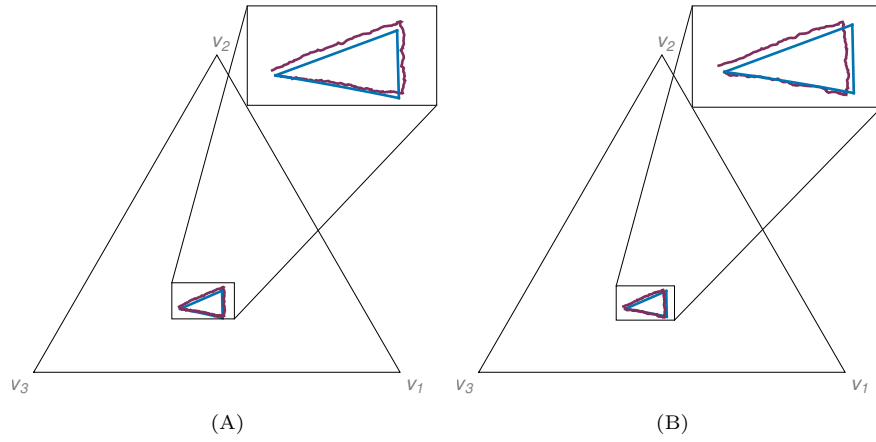


Figure 11 – Realization of stochastic routines (purple) associated with the stable (A) and unstable (B) adjusted replicator routines (blue) designed with the parameters presented on Table 3 and $N = 100,000$ cells. In the upper-right corner of each panel, we show magnified pictures of the most relevant part of the dynamics.

does not seem to play a significant role on the robustness of the stochastic routine. Despite the spread, the initial condition is recovered in the average (Fig. 12).

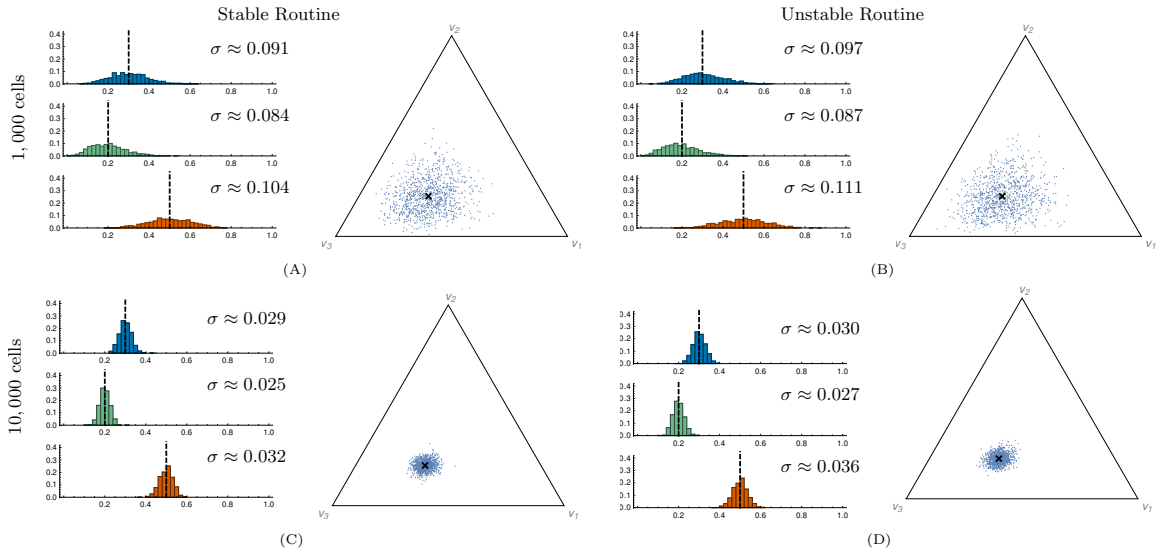


Figure 12 – Spread of 1,000 realizations of the stochastic routines associated with the deterministic cycles on Table 3 after one cycle and the distribution of each coordinate, x_1 in blue, x_2 in green and x_3 in orange. The initial state is marked with an $\times$. The standard deviation σ for each distribution is shown with the corresponding histogram. (A) Stable stochastic routine with $N = 1,000$ cells; (B) Unstable stochastic routine with $N = 1,000$ cells; (C) Stable stochastic routine with $N = 10,000$ cells; (D) Unstable stochastic routine with $N = 10,000$ cells.

A large difference between the unstable and the stable routines appears if we look at several repetitions of the same routine, as it is usually the case in cancer therapy. In the unstable case, the underlying effect of the deterministic dynamics tends to amplify the small distances due to the stochastic spread. On the contrary, in the stable case the deterministic dynamics tends to

decrease the stochastic spread, which is inevitably recreated as time goes by. As a consequence, the stochastic spread is expected to increase with each repetition of the cycle in the unstable case. In the stable case, we expect the stochastic spread to preserve some stability over long periods of time. Results in Fig. 13 confirm these expectations.

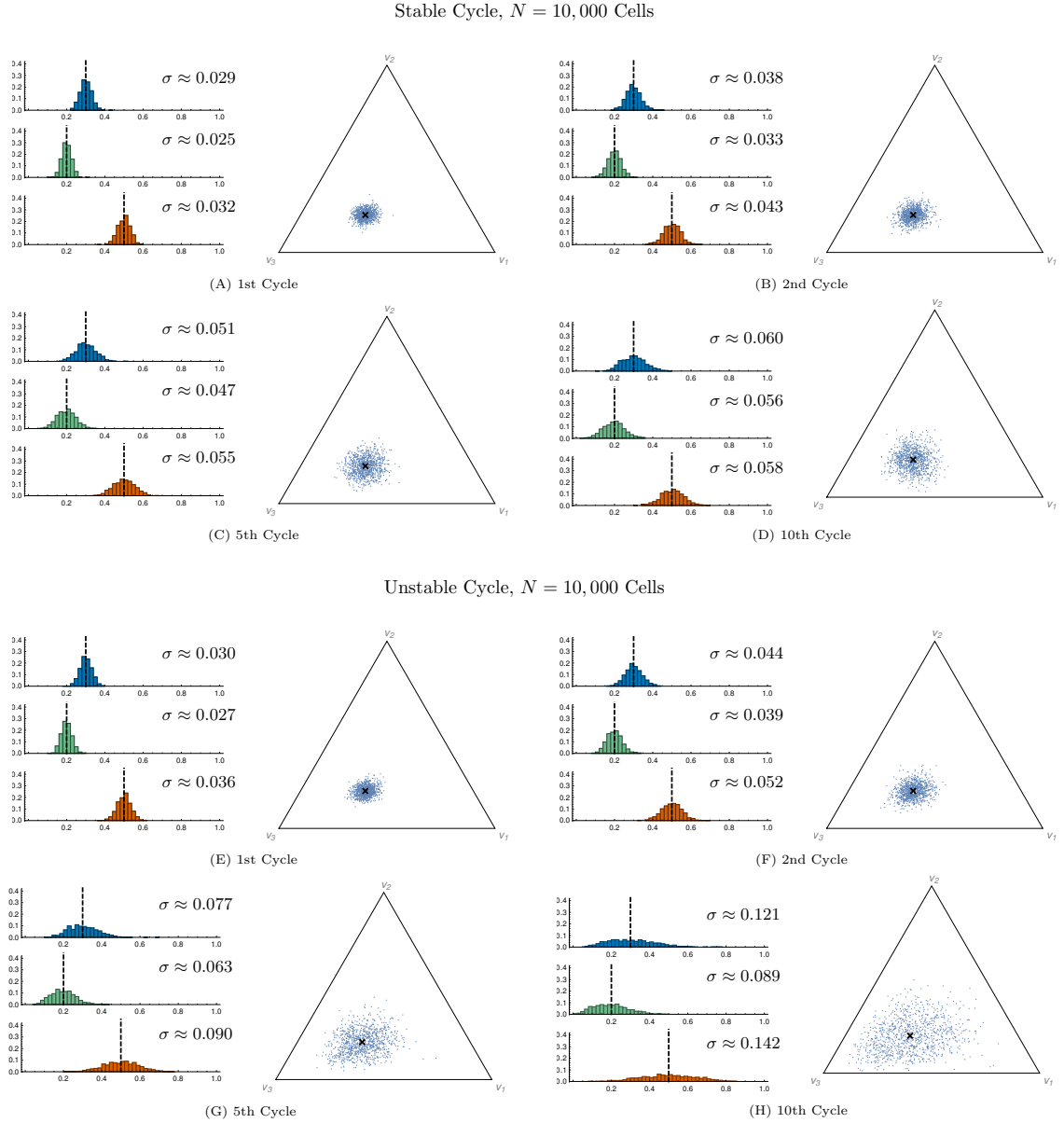


Figure 13 – Spread of 1,000 realizations of the stochastic routines associated with the deterministic cycles on Table 3 with $N = 10,000$ cells after successive cycles and the distribution of each coordinate, x_1 in blue, x_2 in green and x_3 in orange. The initial state is marked with an $\times$. The standard deviation σ for each distribution is shown with the corresponding histogram. (A) Stable stochastic routine after 1 cycle; (B) Stable stochastic routine after 2 cycles; (C) Stable stochastic routine after 5 cycles; (D) Stable stochastic routine after 10 cycles; (E) Unstable stochastic routine after 1 cycle; (F) Unstable stochastic routine after 2 cycles; (G) Unstable stochastic routine after 5 cycles; (H) Unstable stochastic routine after 10 cycles;

5.2 Stochastic Lotka-Volterra Routines

Similarly to what was done on Section 2.4.2, one can use the relationship between the Lotka-Volterra and the replicator dynamics to induce a stochastic counterpart to the cyclic routines for Lotka-Volterra dynamics. Precisely, one can use the time adjustments introduced on that Section to determine the number of steps for each phase to repeatedly execute an *stochastic Lotka-Volterra routine*.

Before translating a Lotka-Volterra system to an equivalent adjusted replicator system, a *scaling* procedure may be advisable, in order that the associated Moran process path is not too close to absorption in one of the faces of the simplex. This scaling was already mentioned in the caption of Table 2 and is explained in more depth on Section 2.4.2.

In Fig. 14, we show realizations of stochastic routines associated with Lotka-Volterra systems. As expected, due to the large value of the effective population size N used, we see that the stochastic paths remain close to the deterministic routines.

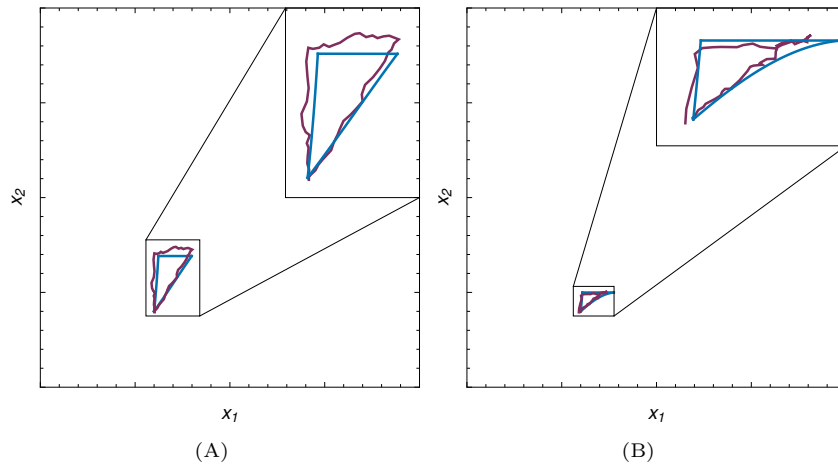


Figure 14 – Realization of stochastic routines (purple) associated with the stable (A) and unstable (B) Lotka-Volterra routines (blue) designed with the parameters presented on Table 2 with 100,000 cells. In the upper-right corner of each panel, we show magnified pictures of the most relevant part of the dynamics.

The overall impact of the effective population size N and the stability of the associated deterministic cycle are perceived in this model similarly to the adjusted replicator model. After one cycle (Fig. 15), we notice that the final positions of many realizations of the same stochastic process spread over the plane and recover, on average, the initial condition. The standard deviation at this point is heavily impacted by the effective population size, but not by the cycle stability.

As we repeat the routine over several iterations, the influence of the stability of the

deterministic cycle becomes evident. Indeed, Fig. 16 shows the final points of 1,000 realizations of the routines associated with the Lotka-Volterra cycles specified in Table 2 after 1, 2, 5 and 10 repetitions of the routines. Once again, the stochastic fluctuations are enhanced by unstable deterministic dynamics and suppressed in the stable, leading to a much slower spread.

As in the adjusted replicator model, both the effective population size and the stability or instability of the associated deterministic dynamics are crucial factors that impact the overall outcome of the stochastic Lotka-Volterra routine: after the first cycle, the spread of the final points is primarily driven by the population size (Fig. 15) and, as the number of cycles increases, the effects of stability become more evident (Fig. 16).

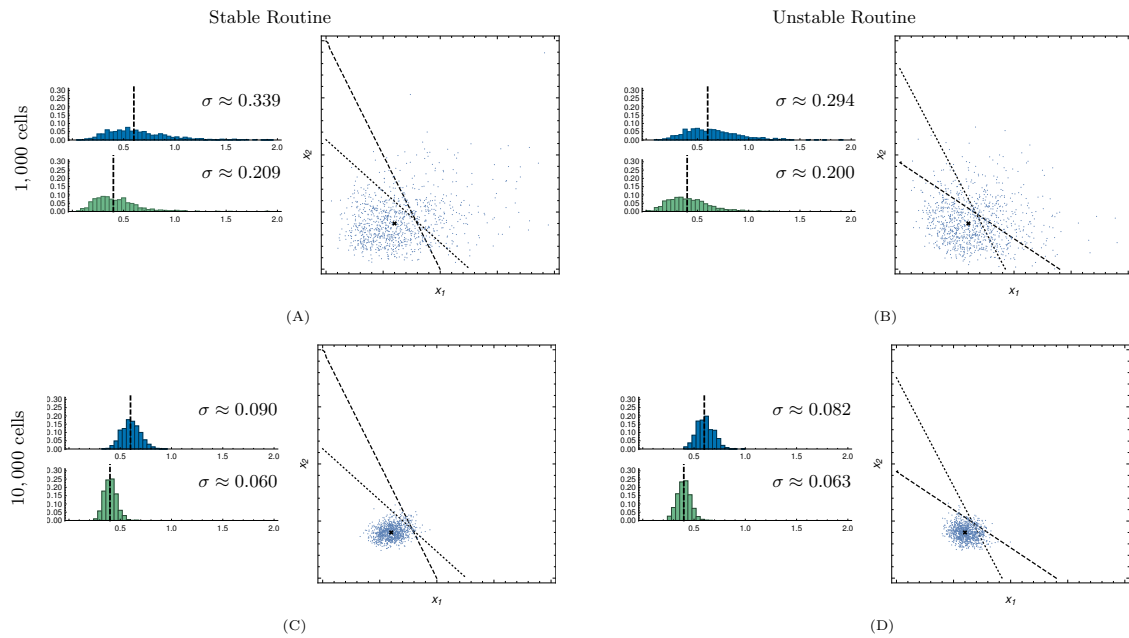


Figure 15 – Spread of 1,000 realizations of stochastic routines associated with the deterministic cycles on Table 2 after one cycle and the distribution of each coordinate, x_1 in blue and x_2 in green. Also in the pictures, the nullclines relative to x_1 (dashed) and x_2 (dotted) for the treatment's time-averaged ODE system. The initial state is marked with an $\times$. The standard deviation σ for each distribution is shown with the corresponding histogram. (A) Stable stochastic routine with $N = 1,000$ cells; (B) Unstable stochastic routine with $N = 1,000$ cells; (C) Stable stochastic routine with $N = 10,000$ cells; (D) Unstable stochastic routine with $N = 10,000$ cells;

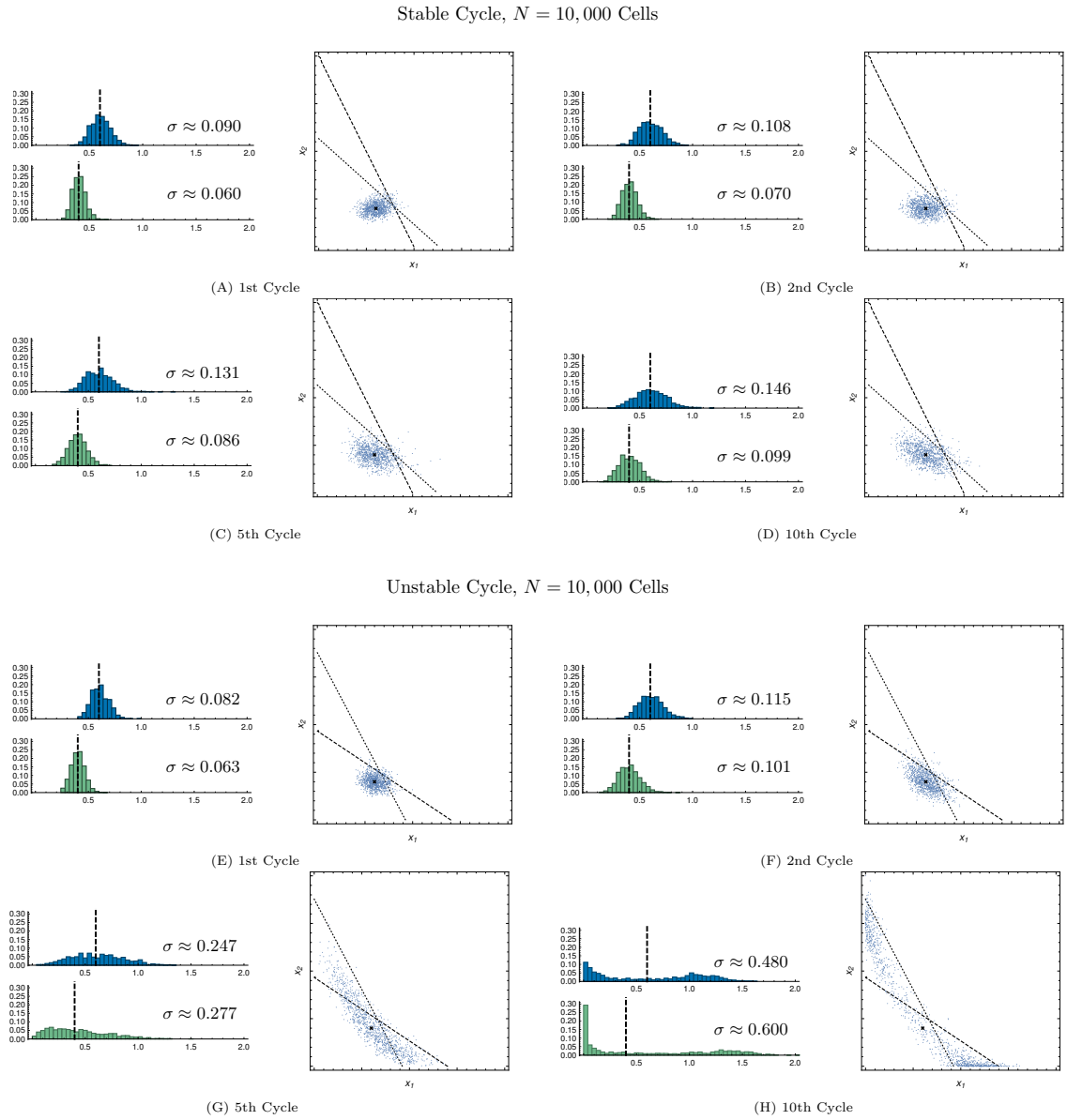


Figure 16 – Spread of 1,000 trials of stochastic routines associated with the deterministic cycles on Table 2 with $N = 10,000$ cells after successive cycles and the distribution of each coordinate, x_1 in blue and x_2 in green. Also in the pictures, the nullclines relative to x_1 (dashed) and x_2 (dotted) for the treatment's time-averaged ODE system. The initial state is marked with an $\times$. The standard deviation σ for each distribution is shown with the corresponding histogram. (A) Stable stochastic routine after 1 cycle; (B) Stable stochastic routine after 2 cycles; (C) Stable stochastic routine after 5 cycles; (D) Stable stochastic routine after 10 cycles; (E) Unstable stochastic routine after 1 cycle; (F) Unstable stochastic routine after 2 cycles; (G) Unstable stochastic routine after 5 cycles; (H) Unstable stochastic routine after 10 cycles;

5.3 Robustness of Stochastic Routines

As discussed in Section 2.4.3, the Moran process is eventually absorbed into a single-species state if iterated for a sufficiently long time. This property naturally extends to stochastic

routines constructed from the Moran process. Indeed, the same reasoning as in Corollary 2 applies: despite time-varying pay-off matrices, there remains a strictly positive probability of increasing the population of any non-extinct cell type at every step. As a result, stochastic cycles inevitably break down in the long run, and fixation in one species becomes unavoidable, regardless of the stability of the associated deterministic cycle. Remember that the result of [Benaïm and Weibull, 2003] we quoted here is valid for a finite time span T .

Indeed, as observed in [Dua et al., 2021], the deviation from the deterministic cycle and absorption of the Markov process can occur over just a few cycles. We stress that the example cycle in [Dua et al., 2021] is *unstable*. The pay-off matrix used there is equivalent to the unstable matrix in Table 3.

On the contrary, in our simulations of Moran processes associated with asymptotically stable cycles with large values of N , we never saw any occurrence of absorption, indicating that the mean absorption time is probably much larger than the number of simulated time steps. We conjecture that the mean time to fixation in one of the absorbing states increases exponentially with N in the asymptotically stable case. This expectation is based on analogy with results in [Antal and Scheuring, 2006], rigorously proved in [Pires, 2022], on mean fixation times for the Moran process for two types of individuals, in the case that a stable coexistence equilibrium for the replicator equation exists.

Dua, Ma and Newton [Dua et al., 2021] state, pessimistically, that a highly accurate tumor model is necessary to design an effective and repeatable adaptive treatment routine, because the stochastic errors will quickly degrade its effectiveness. We may be more optimistic, because we saw here that the stochastic fluctuations will not degrade too much a routine if it is based on an *asymptotically stable* cycle.

In summary, we have shown that the deviation from the planned cycle in stochastic routines is influenced by the stability or instability of the deterministic counterpart, and that this effect is more pronounced for smaller population sizes. We saw that stochastic fluctuations do not rapidly degrade a stochastic routine if it is based on an asymptotically stable cycle.

6 DISCUSSION

The integration of ecological and evolutionary concepts into the study of cancer treatments introduces new and promising possibilities to the field. Adaptive therapy offers a novel approach to controlling tumors and delaying or even preventing the onset of drug resistance. This can be achieved over long periods using cyclic treatment routines. There exists already an extensive literature exploring these routines in different cancer contexts [West et al., 2023, Gedye and Navani, 2022, Zhang et al., 2023] and with different models [Zhang et al., 2017, West et al., 2018, Wang et al., 2021, West et al., 2019, 2020, Dua et al., 2021] but, even though they are already being translated into clinical trials successfully, the mathematical foundation behind this process lacks the proper attention.

We demonstrate in this thesis, both for Lotka-Volterra and adjusted replicator models for tumors, that the design of treatment cyclic routines is indeed feasible under very reasonable biological conditions. Furthermore, we show that there exists an abundance of such cycles, which can be leveraged to identify treatment strategies that are better tailored to individual patient conditions.

Along with the proof of existence of cycles, we also provided an algorithm to design cyclic routines given a starting point and the time to be spent in the first treatment phase. From a mathematical perspective, the generality of this algorithm and the geometrical arguments used in the demonstrations are not only applicable to the Lotka-Volterra and adjusted replicator models discussed here, but can also be extended to other models that exhibit planar dynamics.

The role of stability of deterministic routines was also taken into account in this study. Beyond simply identifying cycles that return the tumor to its initial state, it is essential to ensure that such cycles are resilient to perturbations that are inherent to their translation into clinic treatments, such as fluctuations in estimated parameters, or small delays in drug administration intervals and dosage variations. An asymptotically stable cycle, as defined in this study, would

absorb such deviations and maintain long-term tumor control despite real-world imperfections. Conversely, applying an unstable cycle could lead to outcomes that diverge significantly from theoretical predictions, potentially compromising patient safety and treatment efficacy. For a single treatment routine, stability can be checked relatively easily using computer simulations. However, when many different routines must be tested to find the best one, this process quickly becomes difficult, as verifying stability for each candidate cycle can be computationally expensive and time-consuming.

We stated a conjecture that, if proved, will predict the stability of a given cycle in terms of the dynamics of the time-averaged system. On the biological side, this conjecture brings out an interesting interpretation: in order to control long-term tumor progression as a whole, it is crucial to design treatments that, on average, foster the stable coexistence of tumor cell types. Moreover, given that the stability of the averaged system is an open property, this further implies that the treatment is likely to behave in clinical trials as predicted by the theoretical model, reinforcing the practical relevance of incorporating stability analysis into treatment design. Furthermore, since the presence of a stable coexistence equilibrium in both Lotka–Volterra and replicator equation can be determined directly from their parameters, this result suggests an important, but easily verifiable constraint for treatment optimization routines: the stability of the averaged treatment.

Finally, following [Dua et al., 2021, Park and Newton, 2023] we also examined the Moran process as a stochastic counterpart of the adjusted replicator model. We introduced a stochastic counterpart to the Lotka–Volterra model, too. We showcased here, with both mathematical reasoning and numeric experiments, that the stability or instability of a deterministic routine heavily impacts the expected outcome of the associated stochastic routine when the treatment is applied for long enough. While studies showing that stochastic fluctuations may cause failure of a treatment routine are present in the literature, our work establishes that routines based on stable cycles may be useful for longer. This discovery may enhance our ability to manage and optimize treatment outcome.

A first limitation of our results is that, although the methods to construct cycles may be applicable to other models, non-planar systems remain particularly challenging, since ODE orbits do not limit regions in population spaces with three or more dimensions.

A second limitation is that, while this work provides a rigorous theoretical framework, its direct application to clinical practice remains to be explored. To translate these results into

actionable treatment strategies, close collaboration with clinicians is essential, both to validate the models against real patient data and to ensure that the proposed routines are feasible and safe in practice.

Importantly, from a clinical perspective, our results imply that, in the presence of competitive interactions among cancer cell subtypes, it is in principle possible to design cyclic treatment protocols based on the parameter values describing these intra-tumor dynamics. However, functional quantification of the nature and strength of cell-cell interactions remains rare and challenging. Recent studies have begun to address this gap, quantifying paracrine interactions in mouse models of mammary carcinoma [Noble et al., 2021] and in vitro interactions among breast cancer cell lineages [Emond et al., 2023]. We see strong potential for integrating our modeling framework with such data to address clinically relevant questions: do empirically measured intratumoral interactions satisfy the constraints required for the existence of treatment cycles? Are these cycles stable? Furthermore, since different drugs affect specific cellular interactions and thereby modify the time-averaged system, our results raise another interesting question in the context of drug development: given a set of quantified cell-cell interactions, how might therapies be chosen or designed to promote the conditions necessary for the existence and stability of a therapeutic cycle?

In summary, we establish a mathematical framework for devising cyclic adaptive protocols, highlighting the importance of cycle stability and its connection to stochastic robustness, thereby offering theoretical support for the design of clinically resilient adaptive cancer therapies.

PART III

POPULATION MODELS FOR HSC

DYNAMICS

This part of the thesis is based on the pre-print “Long-term clonal analysis using stochastic models reveals heterogeneity and quiescence of hematopoietic stem cells”, submitted to *Computers in Biology and Medicine* in September, 2025.

7 INTRODUCTION

Hematopoietic stem cells (HSCs) are crucial for sustained blood cell production by balancing self-renewal and differentiation. These rare, multipotent cells reside predominantly in the bone marrow and generate all blood lineages while at the same time maintaining their own stem cell pool [Morrison et al., 1995]. In order to provide a life-long supply of blood cells for a healthy organism, the activity of these stem cells is tightly regulated by both intrinsic genetic and epigenetic mechanisms, as well as by external signals from the bone marrow microenvironment, also referred to as the *stem cell niche* [Ema et al., 2014, Cullen et al., 2014, Pinho and Frenette, 2019, Pereira et al., 2024]. Consequently, understanding the mechanisms by which stem cell numbers and divisional activity are controlled, including the influence of the bone marrow niche on the balance between self-renewal and differentiation, is a key step in understanding the dynamics of the hematopoietic system as a whole. However, few studies addressed the dynamical regulation of stem cell numbers and divisional activity with a particular focus on how the niche can regulate the balance between self-renewal and differentiation and thereby influence stem cell numbers [Wilson et al., 2008, Glauche et al., 2009, Greenbaum et al., 2013, Pedersen et al., 2021].

An important source of HSC data comes from *clonal studies*, where the direct offspring of individual stem cells (*clones*) are tracked over time. In this context, a clone is defined as the population of cells derived from a single hematopoietic stem cell through successive rounds of self-renewal and differentiation. Such clones can be followed experimentally through artificial marking events, either in vitro [Cornils et al., 2014, Lu et al., 2011, Perié et al., 2014, Gerrits et al., 2010] or in vivo [Pei et al., 2017, Sun et al., 2014], or by tracing naturally occurring mutations [Lee-Six et al., 2018]. By monitoring individual clones over time, such studies have elucidated key aspects of HSC self-renewal, long-term persistence and aging, shaping our understanding of hematopoiesis and its regulatory mechanisms [Sun et al., 2014, Lu et al., 2011, Verovskaya et al.,

2013, Mitchell et al., 2022]. Clonal analysis also elucidated how HSCs and other progenitor cells contribute to blood production and how these differentiation pathways are challenged under stress conditions. While gene therapy settings are a possible access to study post-transplantation clonality in humans [Biasco et al., 2016], many aspects of dynamical engraftment and long-term polyclonality are still incompletely understood.

Non-human primates (NHPs) serve as an important clinical model for studying the clonal dynamics of HSCs in a more clinically relevant context [Kim et al., 2014, Radtke et al., 2017]. A remarkable contribution was the recent work by Radtke et al. [2023] that followed individual HSC and progenitor clones over extended periods after autologous transplantation. The high-resolution tracking revealed that long-term persisting HSCs contribute already to early hematopoietic recovery, challenging the assumption that only committed progenitors drive short-term hematopoiesis. Additionally, while their data showed a rapid loss of unique clones and the formation and expansion of long-term persisting clones, it also revealed the continued detection of new clones over time.

Radtke and colleagues used a stochastic, one-compartment model of HSC competition to quantitatively explain the manifestation of long-term persisting clones. This feature is known as clonal conversion and represents a hallmark of neutral competition in which stochastic processes lead to the extinction of some clones and the growth of others. The underlying model assumes that all HSCs are equally likely to divide or differentiate. Although the model explains the loss of clonal diversity between six months and four years post transplantation, it does not account for the delayed emergence of previously undetected clones even after four years. Moreover, this model is insufficient to fully explain the distribution of clone sizes and their respective changes over time.

These observations led us to hypothesize that some key element missing from the applied approach, without which the observed clonal dynamics could not be explained. A candidate improvement to the model comes from the concept HSCs do not constitute a homogeneous pool of cells but rather a dynamic system in which cells can reversibly transition between an active state, characterized by division or differentiation, and a quiescent, potentially regenerative state (Figure 17). This interpretation follows a history of conceptual models that value this dynamic state transitions as an explicit feature of HSC organization, in which the hematopoietic niche plays a crucial role to support HSC quiescence [Wilson et al., 2008, Essers et al., 2009, Morcos et al., 2022, Ibneeva and Grinenko, 2024]. One key piece of evidence supporting this view is

that a homogeneous pool of continuously proliferating and differentiating cells fails to account for the outcomes of label dilution experiments [Glauche et al., 2009]. The concept has also been applied to disease settings, such as leukemia, in which this tight regulation is disturbed and confers an advantage for the mutated clones [Roeder et al., 2005, Fassoni et al., 2018, Hoffmann et al., 2020, Harnisch et al., 2024].

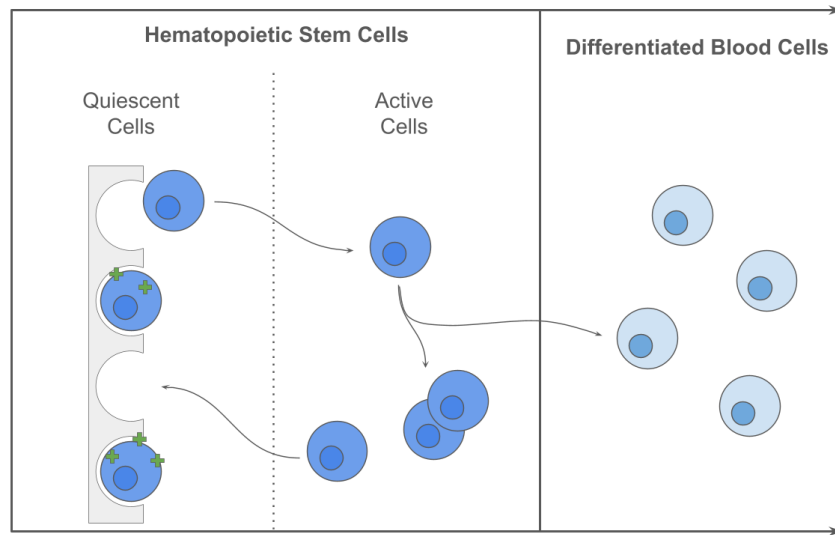


Figure 17 – Schematic representation of hematopoietic stem cell dynamics. HSCs constitute a dynamic system in which cells can reversibly transition between an active state, characterized by division or differentiation, and a quiescent, potentially regenerative state. Active and quiescent HSCs are shown as dark blue cells, while progenitor states with lower self-renewal potential (e.g. multipotent myeloid progenitors) are shown in light blue. Arrows indicate possible transitions and fates.

In this part of the thesis, we propose a stochastic, two-compartment model of clonal hematopoiesis based on the deterministic model proposed by Hoffmann et al. [2020], where HSCs can transition between an active cycling phase and a quiescent state within the niche. Testing this model against a simpler one-compartment model, we find that hematopoietic reconstitution involves not only stochastic fate decisions but also regulated transitions between dormancy and activation. We show that introducing a second compartment provides a mechanistic explanation for the persistence and delayed activation of certain clones, aligning better with long-term tracking data from the respective NHP models. The quantitative evidence for the existence of a quiescent compartment also suggests that interactions within the stem cell niche play a crucial role in long-term hematopoiesis.

This part of the thesis is organized as follows: Chapter 8 introduces models for HSC dynamics under two different assumptions: (i) all HSCs are continuously cycling, and (ii) HSCs transition between quiescent and active states. For each model, we present both a deterministic

formulation (Section 8.1) and a stochastic formulation (Section 8.2), highlighting their similarities and differences. We then discuss different ways to measure the mean time to different events (Section 8.3), followed by a discussion on multiclonal extensions of these models in Section 8.4. In Chapter 9, we describe the procedures used to fit the models to the data, including data selection (Section 9.1), the optimization method (Section 9.2), and performance evaluation criteria (Section 9.3). Finally, Chapter 10 presents the fitting results for the two main models and discusses the hypotheses that can be drawn from these experiments. In Section 10.3, we introduce a third model encompassing different proliferation mechanisms and examine its relevance to the context of long-term homeostatic dynamics. This part concludes with final remarks in Chapter 11.

8 MATHEMATICAL MODELS

In this chapter we present two model structures representing distinct assumptions on the HSC population dynamics, and analyze their dynamical properties through the lens of both deterministic and stochastic models. Precisely, we analyze a model where the HSC constitute an homogeneous pool of continuously dividing *active cells* (*one-compartment model*), and a model where these cells can also enter a *quiescent state* where they stay dormant (*two-compartment model*).

The models discussed here are built upon a well-known framework known as *compartmental models*. In essence, a compartmental model classifies the population into distinct, mutually exclusive categories (or *compartments*), with dynamics governed by both the internal behavior within each compartment and the transitions between compartments (as well as to or from the external environment). In our case, a natural modeling choice is to define two compartments (Figure 18): one for *active* cells that proliferate and differentiate, and another for *quiescent* cells that remain in a dormant state. Cells in the active compartment can proliferate or differentiate, while cells in the quiescent compartment remain in dormant. Furthermore, cells transition between the two compartments through *deactivation* (active becomes quiescent) and *activation* (quiescent becomes active). Naturally, the one-compartment model is composed only of the active compartment.

In this thesis, focus on modeling a healthy, homeostatic system, so we assume that all cells share the same functional capabilities, and that the production of peripheral blood cells through differentiation proceeds at a constant, unperturbed rate. With that in mind, we proceed to the presentation of the models.

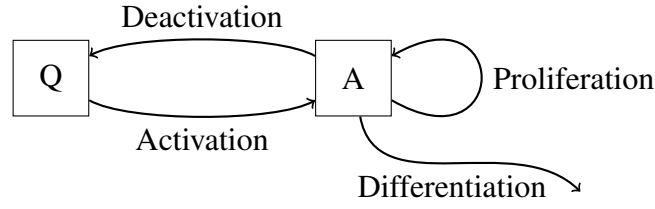


Figure 18 – Schematic representation of a two-compartment model for HSCs dynamics. The cells in the active compartment (represented by the box **A**) are able to either proliferate or differentiate, while cells in the quiescent compartment (represented by the box **Q**) stay dormant. Cells transition between the two compartments through *deactivation* (**A** to **Q**) and *activation* (**Q** to **A**).

8.1 Deterministic Formulation

One-compartment Model

We start our studies with the *one-compartment model*, where we consider only the “basic” dynamics of the HSCs: proliferation and differentiation. Let $x_A(t)$ denote the number of HSCs at time t , and let p_A and d_A be their proliferation and differentiation rates, respectively. Assuming that proliferation is regulated by a logistic term with carrying capacity K_A , the model is

$$x'_A = p_A x_A \left(1 - \frac{x_A}{K_A} \right) - d_A x_A. \quad (8.1)$$

We immediately notice the resemblance to the logistic growth model, with the differentiation acting as an additional loss term. Solving $x'_A = 0$, we see that this model admits two equilibria: an *extinction equilibrium* $x_A \equiv 0$, always present, and a *homeostatic equilibrium* $x_A^* = K_A \frac{p_A - d_A}{p_A}$, which is biologically meaningful only if $p_A > d_A$.

Similar to the logistic model, this one can also be solved exactly by separation of variables, and it has two possible dynamics:

- If $p_A < d_A$, only the extinction equilibrium is biologically relevant, and it is globally asymptotically stable¹.
- If $p_A > d_A$, the extinction state becomes unstable, and the homeostatic equilibrium x_A^* is globally asymptotically stable¹.

Biologically, this conveys a simple but important principle: the balance between proliferation and differentiation dictates whether the HSC population extinguish or persists. If self-renewal

¹ For positive initial conditions.

outpaces differentiation, the population stabilizes at a fraction of its carrying capacity; if not, it declines to extinction.

Two-compartment Model

Next, we present the *two-compartment model*, based on the work of Hoffmann et al. [2020] regarding HSC dynamics in the context of Acute Myeloid Leukemia. In a healthy context, it can be seen as an extension to the one-compartment model including a quiescent compartment and bidirectional transitions between the two states (recall Figure 18). Specifically, cells deactivate from the active to the quiescent state at rate t_{AQ} , and activate from the quiescent to the active state at rate t_{QA} . We also introduce a carrying capacity K_Q for the quiescent compartment and assume that the transitions are regulated by logistic-like terms.

Formally if $x_Q(t)$ and $x_A(t)$ denote the number of quiescent and active cells at time t , respectively, the model is given by:

$$\begin{cases} x'_Q = t_{AQ}x_A \left(1 - \frac{x_Q}{K_Q}\right) - t_{QA}x_Q \left(1 - \frac{x_A}{K_A}\right), \\ x'_A = -t_{AQ}x_A \left(1 - \frac{x_Q}{K_Q}\right) + t_{QA}x_Q \left(1 - \frac{x_A}{K_A}\right) + p_Ax_A \left(1 - \frac{x_A}{K_A}\right) - d_Ax_A. \end{cases} \quad (8.2)$$

A first remark on the dynamics of this model is that, unlike in the other deterministic models presented in this thesis, here the coordinate axes are not invariant. Indeed, notice that if $x_A > 0$ and $x_Q = 0$, then $x'_Q \neq 0$. Similarly, if $x_A = 0$, $x_Q > 0$, then $x'_A > 0$. It follows that the direction on both axes point to the interior or R_2^+ and, consequently, this quadrant is forward invariant under this dynamics. Biologically, this tell us that if we start with non-negative a cell population, then this population stays non-negative as the model evolves.

It is immediate to see that the *extinction state* $\mathbf{x} = (0, 0)$ is an equilibrium. The Jacobian matrix at this point is

$$DF(0,0) = \begin{pmatrix} -t_{QA} & t_{AQ} \\ t_{QA} & -t_{AQ} + p_A - d_A \end{pmatrix},$$

with characteristic polynomial $\lambda^2 + b\lambda + c = 0$, where $b = t_{QA} + t_{AQ} - (p_A - d_A)$ and $c = -(p_A - d_A)t_{QA}$.

Notice that if $p_A < d_A$ then the product of the two roots, c , is positive, while their sum, $-b$, is negative. Therefore both eigenvalues have negative real part, implying that the origin is locally asymptotically stable.

If $p_A > d_A$, then $c < 0$, so the eigenvalues have opposite signs and we conclude that the origin is a saddle. Indeed, we can show that restricted to $\mathbb{R}_+^2$, the origin is in fact a repeller. Notably, if $x_A = 0$ and $x_Q > 0$, then $x'_A > 0$, so active cells increase (analogous for $x_Q = 0$ and $x_A > 0$). At the same time, for the total population $N = x_A + x_Q$, we have

$$N' = p_A x_A \left(1 - \frac{x_A}{K_A}\right) - d_A x_A, \quad (8.3)$$

which, regardless of x_Q , is positive whenever $x_A \in \left(0, K_A \frac{p_A - d_A}{p_A}\right)$ if $p_A > d_A$. Thus, trajectories cannot approach the origin from $\mathbb{R}_+^2$ (see Figure 20).

No other equilibria lie on the coordinate axes, since as already observed, in this axes the direction field point to the interior of R_2^+ . Thus, any other biologically relevant equilibrium must have strictly positive populations x_A and x_Q .

In order to determine them, we notice that when $x'_Q = 0$, the equation for x'_A reduces to the one-compartment formulation, implying that any coexistence equilibrium $\mathbf{x}^*$ must satisfy

$$x_A^* = K_A \frac{p_A - d_A}{p_A}.$$

From the same identity, we also obtain $1 - \frac{x_A^*}{K_A} = \frac{d_A}{p_A}$, which substituting into $x'_Q = 0$ yields

$$x_Q^* = K_Q \frac{t_{AQ}(p_A - d_A)K_A}{t_{AQ}(p_A - d_A)K_A + t_{QA}d_A K_Q} = K_Q \frac{1}{1 + \frac{K_Q t_{QA}}{K_A t_{AQ}} \frac{1 - \alpha}{\alpha}},$$

where $\alpha = \frac{p_A - d_A}{p_A}$. Observing that $\alpha < 1$, it is immediate to see that this coexistence equilibrium is biologically relevant if and only if $\alpha > 0$, i.e., $p_A > d_A$. A Jacobian analysis at $\mathbf{x}^*$ shows that it is locally asymptotically stable whenever it exists. We refer the curious reader to the supplementary material of Harnisch et al. [2024], where this analysis is carried out in detail.

To extend the local stability analysis to global dynamics, we first observe that all trajectories in $\mathbb{R}_+^2$ are bounded. Indeed, fix $N^* > K_A + K_Q$ and consider the region $R \subset \mathbb{R}_+^2$ bounded by the curves

$$\begin{aligned} \gamma_Q &= \{(x_Q, x_A) \in \mathbb{R}_+^2 \mid x_Q = N^* - K_A\}, \\ \gamma_A &= \{(x_Q, x_A) \in \mathbb{R}_+^2 \mid x_A = N^* - K_Q\}, \\ \gamma_N &= \{(x_Q, x_A) \in \mathbb{R}_+^2 \mid x_A + x_Q = N^*\}, \end{aligned} \quad (8.4)$$

as illustrated in Figure 19. On the boundary of R , the vector field points strictly to the interior of the region:

- If $\mathbf{x} \in \partial R \cap \gamma_Q$, then $x_Q > K_Q$ and $x_A < K_A$, hence $x'_Q < 0$, so x_Q decreases.

- If $\mathbf{x} \in \partial R \cap \gamma_A$, then $x_A > K_A$ and $x_Q < K_Q$, hence $x'_A < 0$, so x_A decreases.
- If $\mathbf{x} \in \partial R \cap \gamma_N$, then $x_A > K_A$, so the total population satisfies

$$N' = x'_A + x'_Q = p_A x_A \left(1 - \frac{x_A}{K_A}\right) - d_A x_A < 0.$$

Notice that N' is precisely the inner product of the model's direction field (x'_Q, x'_A) with the vector $(1, 1)$, which is normal to the curve γ_N . Since $N' < 0$, the angle between these two vectors is greater than $\pi/2$. Moreover, because $(1, 1)$ points *outwards* from the region R , it follows that the direction field along γ_N indeed points inwards, toward the region R .

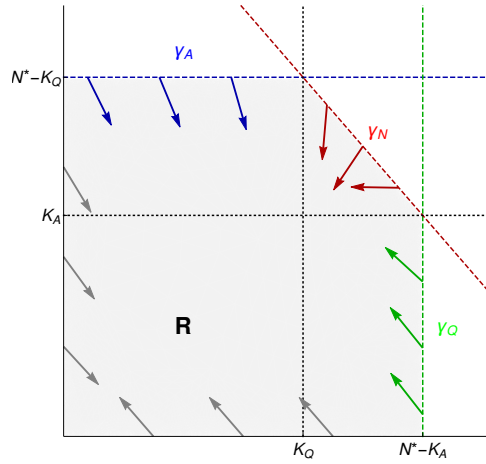


Figure 19 – Direction field of the Hoffmann model along the boundary of the compact region R . The curves γ_A , γ_N , and γ_Q , defined in equation (8.4), are shown in blue, red, and green, respectively. The black dashed horizontal and vertical lines correspond to $x_A = K_A$ and $x_Q = K_Q$. On every boundary segment lying in the interior of $\mathbb{R}_+^2$, as well as in the x_Q and x_A axis (gray arrows), the vector field points strictly inward, ensuring that all trajectories remain confined within R .

In addition, as already observed, in the coordinate axes the direction field points into the interior of $\mathbb{R}_+^2$. It follows that every trajectory starting inside R remains in R for all $t > 0$. A region R can be designed to contain any initial condition in $\mathbb{R}_+^2$, from Theorem 2 follows that all orbits are defined for all future time.

Since the orbits are limited and we are in a planar system, the Poincaré–Bendixson theorem (see, e.g., Section 3.7 of Perko [1991]) guarantees that all orbits must converge to a single equilibrium point, to a periodic orbit or to a finite set of equilibrium points along with orbits connecting them. Furthermore, the Dulac's criterion (see, e.g., Proposition 1.195 of Chicone [2006]) applied with the Dulac function $\Phi(x_Q, x_A) = 1/(x_Q x_A)$, excludes the possibility of periodic orbits, so the only possibility is the convergence to some equilibria².

² Along with the orbits connecting them, if more than one.

Now, when $p_A < d_A$ the only equilibrium is the extinction state, thus it is globally asymptotically stable. When $p_A > d_A$, the extinction state becomes a repeller in the first quadrant, so no orbit can approach it and, consequently, the homeostatic equilibrium $\mathbf{x}^*$ is a globally attracting equilibrium (see Figure 20).

Together with the analysis carried out in the previous subsection, we can summarize the behavior of the deterministic models presented here as follow:

Proposition 7. *Let $\mathbf{x}(t)$ be a positive solution of (8.1) or (8.2). Then*

- *if $p_A < d_A$, then $\mathbf{x}(t) \rightarrow \mathbf{0}$ when $t \rightarrow \infty$;*
- *if $p_A > d_A$, then $\mathbf{x}(t) \rightarrow \mathbf{x}^*$ when $t \rightarrow \infty$, where $\mathbf{x}^*$ is the homeostatic equilibrium of the respective model.*

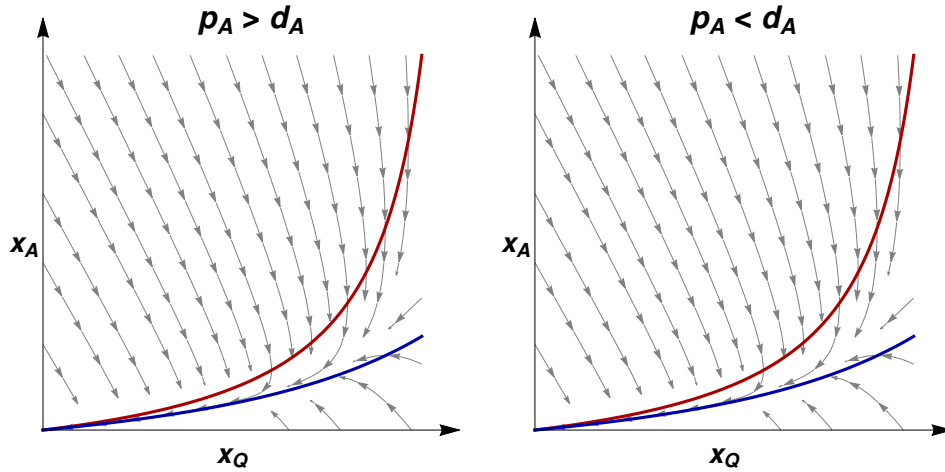


Figure 20 – Possible dynamics of the two-compartment model. If proliferation exceeds differentiation (left), the cell population reaches an homeostatic equilibrium. Otherwise (right), the HSCs eventually go extinct.

Biologically, Proposition 7 states that if proliferation exceeds differentiation, a *homeostatic equilibrium* $\mathbf{x}^*$ exists and it attracts all interior trajectories. Otherwise, the *extinction equilibrium* is globally asymptotically stable, meaning that, overtime, the HSCs are not able to maintain their population.

8.2 Stochastic Formulation

We now derive the stochastic version of the two-compartment model presented in the previous section. Naturally, a stochastic one-compartment model can be recovered from this description by removing the quiescent compartment and the associated transitions.

Once again, let $\mathbf{X} = (X_Q, X_A) \in \mathbb{N}^2$ denote the number of quiescent and active cells. Using the same parameters as in the deterministic models, we define a process in which each cell *independently* performs an action with probabilities equal to the corresponding deterministic rates normalized by a *base rate*

$$\Phi \geq \max\{p_A + d_A + t_{AQ}, t_{QA}\}.$$

The *action probabilities* are as follows:

$$\begin{aligned} \pi_{d_A} &= \frac{d_A}{\Phi}, & (\text{differentiation of an active cell}), \\ \pi_{p_A} &= \max\left\{0, \frac{p_A \left(1 - \frac{X_A}{K_A}\right)}{\Phi}\right\}, & (\text{proliferation of an active cell}), \\ \pi_{t_{AQ}} &= \max\left\{0, \frac{t_{AQ} \left(1 - \frac{X_Q}{K_Q}\right)}{\Phi}\right\}, & (\text{deactivation of an active cell}), \\ \pi_{\emptyset_A} &= 1 - (\pi_{d_A} + \pi_{p_A} + \pi_{t_{AQ}}), & (\text{no action for an active cell}), \\ \pi_{t_{QA}} &= \max\left\{0, \frac{t_{QA} \left(1 - \frac{X_A}{K_A}\right)}{\Phi}\right\}, & (\text{activation of a quiescent cell}), \\ \pi_{\emptyset_Q} &= 1 - \pi_{t_{QA}}, & (\text{no action for a quiescent cell}). \end{aligned} \tag{8.5}$$

We introduce the random variables

$$\begin{aligned} (N_{p_A}(\mathbf{X}), N_{d_A}(\mathbf{X}), N_{t_{AQ}}(\mathbf{X}), N_{\emptyset_A}(\mathbf{X})) &\sim \text{Multinom}(X_A; \pi_{p_A}, \pi_{d_A}, \pi_{t_{AQ}}, \pi_{\emptyset_A}), \\ (N_{t_{QA}}(\mathbf{X}), N_{\emptyset_Q}(\mathbf{X})) &\sim \text{Multinom}(X_Q; \pi_{t_{QA}}, \pi_{\emptyset_Q}). \end{aligned}$$

The value $N_\lambda(\mathbf{X})$ represents number of cells that performed action λ when the system was at state $\mathbf{X}$.

The stochastic process is then defined by

$$\begin{cases} X_A^{k+1} = N_{\emptyset_A}(\mathbf{X}^k) + 2N_{p_A}(\mathbf{X}^k) + N_{t_{QA}}(\mathbf{X}^k), \\ X_Q^{k+1} = N_{\emptyset_Q}(\mathbf{X}^k) + N_{t_{AQ}}(\mathbf{X}^k). \end{cases} \tag{8.6}$$

To simplify the notation, we omit, from now on, the dependence of the random variables on $\mathbf{X}^k$.

Observe that, once again, the stochastic model is as a discrete-time Markov chain. The discrete-time nature is immediate, since the process evolves discrete steps. Moreover, the Markov

property holds since each action probability depend only on the current state $\mathbf{X}^k$ and not on the past trajectory.

However, unlike the other processes analyzed in this thesis, the process considered here *is not* of finite-state. Indeed, if $K_A \geq 2$, for any non-null initial condition, there exists a trajectory with positive probability that reaches arbitrarily large numbers of cells.

To see this, consider an initial state with precisely one active cell, $\mathbf{X}^0 = (X_Q^0, 1)$. Any other non-null state can reach such a configuration with positive probability in one step: if there are initially more than one active cell, $\mathbf{X}^0$ is reached if all but one active cell differentiates; if there are no active cells, $\mathbf{X}^0$ is reached if exactly one quiescent cell activates. For simplicity, assume that $K_A \in \mathbb{N}$ (otherwise take $\lfloor K_A \rfloor$).

Now consider the following trajectory:

Step 1 to $K_A - 2$: Quiescent cells remain quiescent, while exactly one active cell proliferates per step. By the end of these steps, the state is

$$\mathbf{X}^{K_A-2} = (X_Q^0, K_A - 1).$$

Step $K_A - 1$: All quiescent cells activate simultaneously, while active cells remain unchanged, so

$$\mathbf{X}^{K_A-1} = (0, X_Q^0 + K_A - 1).$$

Step K_A : All but one active cell deactivate, giving

$$\mathbf{X}^{K_A} = (X_Q^0 + K_A - 2, 1).$$

Notice that all of the events above has *positive* probability. For steps $i = 1, \dots, K_A - 2$, since $X_A^i < K_A$, proliferation has positive probability. The same holds for the activation probability on step $K_A - 1$. In step K_A , since the quiescent compartment is empty, the deactivation probability is also positive. In all steps, since the probability of “*doing something*” is not 1, the probability of “*doing nothing*” ($\pi_{\emptyset_A}$ and $\pi_{\emptyset_Q}$) is positive.

Thus the entire path occurs with positive probability. By iterating this construction, for any $m \in \mathbb{N}$ we obtain the trajectory

$$\mathbf{X}^{mK_A+i} = \begin{cases} (X_Q^0 + m(K_A - 2), 1 + i), & \text{for } i = 0, \dots, K_A - 2, \\ (0, X_Q^0 + (m + 1)(K_A - 2) + 1), & \text{for } i = K_A - 1, \end{cases}$$

which again has strictly positive probability. Therefore, with positive probability the trajectory starting at $\mathbf{X}^0$ reaches arbitrarily large numbers of quiescent cells, showing that the state space is, in fact, unbounded.

On the other hand, it can be shown that *all states but the extinction* are transient. By the definition of our process, it is clear that the extinction state $\mathbf{0} = (0, 0)$ is absorbing. Now, from any other state there is a positive probability of reaching extinction: starting from $\mathbf{X}^0 = (X_Q^0, X_A^0)$, one possible path is:

Step 1: All active cells differentiate while quiescent cells remain inactive.

Step 2: All quiescent cells activate.

Step 3: All newly active cells differentiate.

The probability of this sequence is

$$P(\mathbf{X}^3 = \mathbf{0} \mid \mathbf{X}^0) \geq \underbrace{(\pi_{d_A})^{X_A^0} (\pi_{\emptyset_Q})^{X_Q^0}}_{\substack{\text{Active} \rightarrow \text{diff.} \\ \text{Quiescent} \rightarrow \text{still}}} \underbrace{(\pi_{t_{QA}})^{X_Q^0}}_{\text{Quiescent} \rightarrow \text{active}} \underbrace{(\pi_{d_A})^{X_Q^0}}_{\text{New active} \rightarrow \text{diff.}},$$

which is strictly positive. Since the probability that the process leaves the origin is null, it follows from Theorem 3 that, indeed, state $\mathbf{X}_0$ is transient.

In the case $p_A > d_A$, both unbounded proliferation and cell extinction contrast sharply with the deterministic dynamics. Indeed, when proliferation exceeds differentiation, the deterministic system converges to a positive homeostatic equilibrium $\mathbf{x}^*$, hence, deterministic trajectories neither tend to extinction nor diverge to infinity, but instead remain bounded and approach $\mathbf{x}^*$ as $t \rightarrow \infty$.

Despite these divergent outcomes, what we observe *on average* is, in fact, that the stochastic process follows dynamics that are *close* to those of the deterministic models. Namely, let $\mathbf{X}^k$ denote the stochastic process defined above with $\mathbf{X}^0 \in [0, K_Q] \times [0, K_A]$, and let $\tilde{\mathbf{x}}^k$ be the approximation of the solution of (8.2) obtained via the first-order Euler method [Atkinson, 1989, Filho, 2007] with initial condition $\tilde{\mathbf{x}}^0 = \mathbf{X}^0$ and step size $h = 1/\Phi$, i.e.

$$\tilde{\mathbf{x}}^{k+1} = \tilde{\mathbf{x}}^k + h\Psi'(kh, \tilde{\mathbf{x}}^k).$$

The expected number of cells in the quiescent compartment after one step is

$$\begin{aligned}
\mathbb{E}[X_Q^1] &= \mathbb{E}[N_{\emptyset_Q}] + \mathbb{E}[N_{t_{AQ}}] \\
&= X_Q^0 (1 - \pi_{t_{QA}}) + X_A^0 \pi_{t_{AQ}} \\
&= X_Q^0 - X_Q^0 \frac{1}{\Phi} t_{QA} \left(1 - \frac{X_A^0}{K_A}\right) + X_A^0 \frac{1}{\Phi} t_{AQ} \left(1 - \frac{X_Q^0}{K_Q}\right) \\
&= X_Q^0 + \frac{1}{\Phi} \underbrace{\left[-t_{QA} X_Q^0 \left(1 - \frac{X_A^0}{K_A}\right) + t_{AQ} X_A^0 \left(1 - \frac{X_Q^0}{K_Q}\right) \right]}_{\text{Quiescent field in } \mathbf{X}^0} \\
&= \tilde{x}_Q^0 + h \Psi'_Q(0, \tilde{x}_Q^0) \\
&= \tilde{x}_Q^1.
\end{aligned} \tag{8.7}$$

An analogous computation for the active compartment gives a matching identity, so in vector form $\mathbb{E}[\mathbf{X}^1] = \tilde{\mathbf{x}}^1$.

In the particular case where $\mathbf{X}^0 = \mathbf{x}^*$ is a deterministic equilibrium, we have $\tilde{\mathbf{x}}^k = \mathbf{x}^*$ for all k , and consequently $\mathbb{E}[\mathbf{X}^1] = \mathbf{x}^*$, that is, the equilibrium is preserved in expectation. In fact, in our experiments, when starting from the homeostatic equilibrium, the average trajectory remained at equilibrium for all simulated steps. Nevertheless, the *variance* of the stochastic process increased with time.

From the one-step derivation above, we notice two factors that influence the drift and the spread of the stochastic trajectory around the deterministic solution $\mathbf{x}(t)$:

- As the *cell counts* increase, the random variables N_λ concentrate around their expectations, so relative fluctuations diminish. Consequently, the spread of stochastic trajectories around their own expectation, the Euler approximation, also diminishes.
- As the *base rate* Φ increases, the time step $h = 1/\Phi$ decreases, reducing the Euler approximation error with respect to the true deterministic trajectory. Consequently, the average stochastic trajectory also approaches the deterministic one.

Both effects are illustrated in Figure 21.

8.3 Mean Time to Events

The frequency of cellular events is a central aspect of biological dynamics, as it defines how often key processes such as division or differentiation occur. Fitting mechanistic models

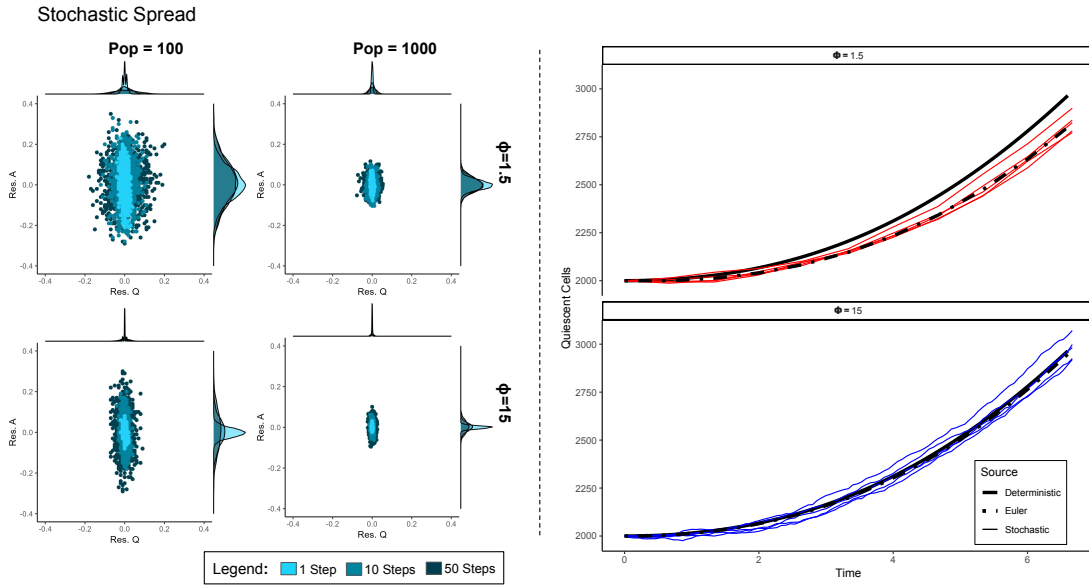


Figure 21 – Stochastic deviations from the deterministic model. *Left:* Spread of stochastic trajectories around the deterministic homeostatic equilibrium, based on 1,000 simulations *starting on equilibrium*. Each point represents the relative difference between stochastic and deterministic populations of quiescent (x-axis) and active (y-axis) cells, normalized by the deterministic equilibrium values. Marginal density plots illustrate the distributions of deviations at selected time steps. *Right:* Example of five trajectories of the number of quiescent cells. Stochastic simulations are shown as solid colored lines, Euler method approximations as dashed black lines, and exact deterministic solutions as solid black lines, for two values of the base rate Φ . All simulations start from one-tenth of the equilibrium population size. As Φ increases, the Euler approximation converges toward the deterministic solution and so do the stochastic trajectories.

to empirical data enables the estimation of these frequencies or, equivalently, the inverse of it, which is the average time until an event occurs.

In a stochastic framework, these quantities can be empirically measured on simulations by monitoring the time required for individual cells to undergo a specific event. As illustrated in Figure 22, the estimated average time for an event depends on the simulation period and stabilizes as this period increases well beyond the measured event time scale. This convergence behavior, typical of average-based estimations, highlights the importance of longer experiments to accurately capture the true temporal dynamics, especially for infrequent events.

Apart from this computational approach, one can also estimate these average times analytically from a deterministic model. In mathematical models based on ODEs, this quantity is typically calculated as the inverse of the rate governing that process and can often be derived analytically (see, e.g., Murray [2002] or Edelstein-Keshet [2005] for examples in epidemiology). For the Hoffmann model, all rates, except for the differentiation, depend on the population sizes. Nevertheless, we have shown that all trajectories converge either to extinction or to the

homeostatic equilibrium $\mathbf{x}^*$. At the extinction state no events occur, so this case is not of particular interest. At the homeostatic equilibrium, however, cells continue to perform actions even though the average number of cells in each compartment remains constant. Consequently, at equilibrium the effective rates become constant.

For example, the effective proliferation rate at equilibrium is

$$p_A \left(1 - \frac{x_A^*}{K_A} \right) = p_A \left(1 - \frac{K_A \cdot \frac{p_A - d_A}{p_A}}{K_A} \right) = d_A,$$

which matches the differentiation rate. In the same way, one can derive the effective activation and deactivation rates at equilibrium. These rates are summarized in Table 4, together with the corresponding average times between actions for single cells.

As illustrated in Figure 22, the empirically derived time converges toward its analytically predicted counterpart as the simulation time increases. This asymptotic agreement between the deterministic and stochastic formulations highlights the potential of coupled studies using both approaches, as it suggests that insights derived from one framework can potentially inform interpretations in the other.

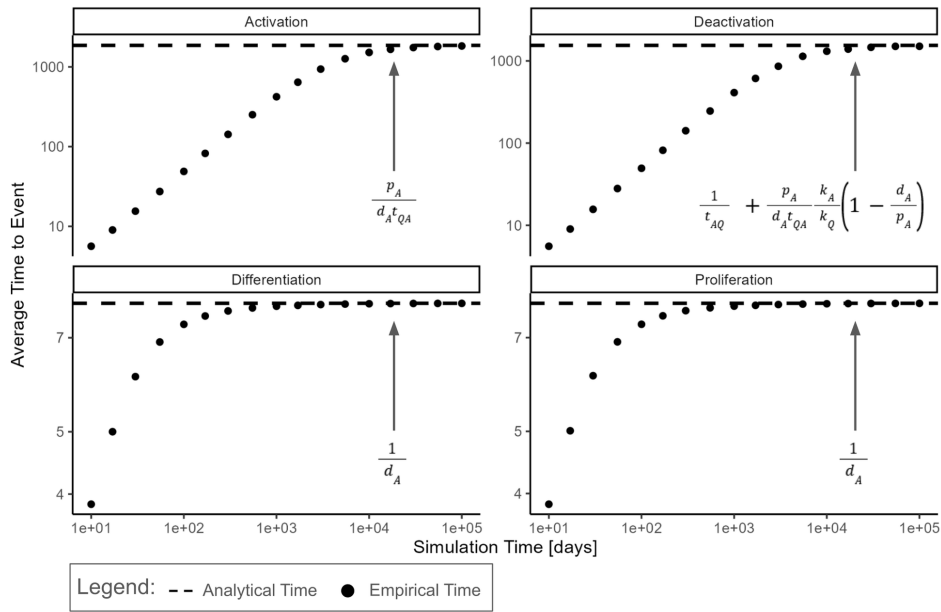


Figure 22 – Empirical and analytical estimates for the average time to event at homeostatic equilibrium. Dots indicate the estimated average time between events recorded in the simulations at different time points. As the simulation period increases, the empirical estimates (dots) converge toward its analytically predicted counterpart (dashed lines).

State	Action	Rates	Average Time
Active	Proliferation	d_A	$\frac{1}{d_A}$
	Differentiation	d_A	$\frac{1}{d_A}$
	Deactivation	$\frac{t_{AQ}t_{QA}d_AK_Q}{t_{AQ}(p_A - d_A)K_A + t_{QA}d_AK_Q}$	$\frac{1}{t_{AQ}} + \frac{K_A}{K_Q} \frac{p_A - d_A}{d_A t_{QA}}$
Quiescent	Activation	$\frac{d_A t_{QA}}{p_A}$	$\frac{p_A}{d_A t_{QA}}$

Table 4 – Analytical rates and average time to cellular events at homeostatic equilibrium of the Hoffmann model (8.2). The average time value represents the expected time between two occurrences of a given action for a single cell, assuming the system has reached steady state.

8.4 Multiclonal Dynamics

Clonality data provides an important perspective on HSC population dynamics, as it enables the study of how different hematopoietic stem cell clones coexist, compete, and evolve over time. All models presented in the previous sections can be extended to a multiclonal setting by subdividing the compartments to accommodate the different clonal subpopulations. Precisely, assuming that all clones are functionally identical, i.e., they all have the same rates for proliferation, differentiation, etc., the extension of the Hoffmann model to n clones takes the form

$$\begin{cases} x'_{Q,i} = t_{AQ}x_{A,i} \left(1 - \frac{\sum_{j=1}^n x_{Q,j}}{K_Q}\right) - t_{QA}x_{Q,i} \left(1 - \frac{\sum_{j=1}^n x_{A,j}}{K_A}\right), \\ x'_{A,i} = -t_{AQ}x_{A,i} \left(1 - \frac{\sum_{j=1}^n x_{Q,j}}{K_Q}\right) + t_{QA}x_{Q,i} \left(1 - \frac{\sum_{j=1}^n x_{A,j}}{K_A}\right) + \\ \quad + p_A x_{A,i} \left(1 - \frac{\sum_{j=1}^n x_{A,j}}{K_A}\right) - d_A x_{A,i}, \end{cases} \quad (8.8)$$

where $x_{Q,i}$ and $x_{A,i}$ represents the abundance of clone i in the quiescent and active compartments, respectively, for $i = 1, \dots, n$. A schematic representation of this model is presented on Figure 23.

Since all clones share the same parameters $(p_A, d_A, t_{AQ}, t_{QA}, K_A, K_Q)$, the qualitative dynamics of the multiclonal system are identical to those of the single-clone Hoffmann model.

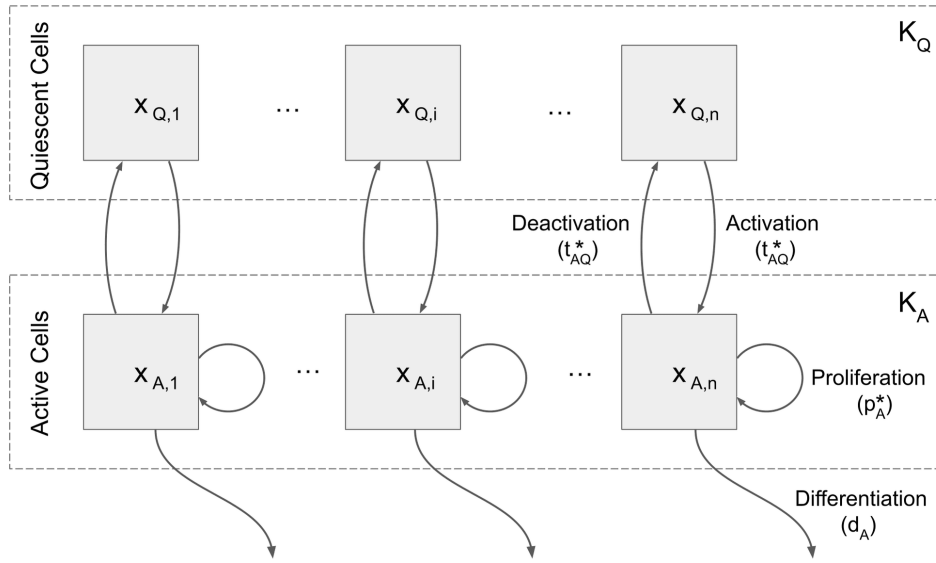


Figure 23 – Two-compartment model of multiclonal HSCs dynamics. Dashed boxes indicate the “active” and “quiescent” compartments with their respective carrying capacities, K_A and K_Q , and one sub-compartment for each clone represented as gray boxes. $x_{A,i}$ and $x_{Q,i}$ indicate the abundance of clone i in the active and quiescent compartments, respectively. Arrows indicate the flux of cells when the corresponding action is taken with rates as in equation (8.8). A one-compartment model can be obtained by removing the quiescent compartment along with the actions that interact with it.

Indeed, if we consider the total number of active cells $x_A(t) = \sum x_{A,i}(t)$ and the total number of quiescent cells $x_Q(t) = \sum x_{Q,i}(t)$, it is immediate that $(x_Q(t), x_A(t))$ is a solution of the single-clone Hoffmann model (8.2). Therefore, if $p_A > d_A$, the HSC population approaches homeostatic equilibrium, and if $p_A < d_A$, the total number of cells tends to zero and, consequently, all clones go extinct.

Moreover, we see that the deterministic dynamics preserve clonal proportions. Precisely, if

$$\mathbf{x}(t) = (x_{Q,1}(t), \dots, x_{Q,n}(t), x_{A,1}(t), \dots, x_{A,n}(t))$$

is a solution of (8.8) with

$$\alpha_i = \frac{x_{Q,i}(0)}{x_Q(0)} = \frac{x_{A,i}(0)}{x_A(0)}, \quad (8.9)$$

for $i = 1, \dots, n$, then a straight derivation shows that

$$\tilde{\mathbf{x}}(t) = (\alpha_1 x_Q(t), \dots, \alpha_n x_Q(t), \alpha_1 x_A(t), \dots, \alpha_n x_A(t))$$

is also a solution with the same initial conditions. Indeed, by definition, it is clear that $\tilde{x}_{Q,i}(0) =$

$x_{Q,i}(0)$ and $\tilde{x}_{A,i}(0) = x_{A,i}(0)$. Furthermore, we have

$$\begin{aligned}
 \tilde{x}'_{Q,i} &= \alpha_i x'_{Q,i} \\
 &= \alpha_i \left[-t_{QA} x_Q \left(1 - \frac{x_A}{K_A} \right) + t_{AQ} x_A \left(1 - \frac{x_Q}{K_Q} \right) \right] \\
 &= -t_{QA} (\alpha_i x_Q) \left(1 - \frac{x_A \sum_{j=1}^n \alpha_j}{K_A} \right) + t_{AQ} (\alpha_i x_A) \left(1 - \frac{x_Q \sum_{j=1}^n \alpha_j}{K_Q} \right) \\
 &= -t_{QA} \tilde{x}_{Q,i} \left(1 - \frac{\sum_{j=1}^n \tilde{x}_{A,j}}{K_A} \right) + t_{AQ} \tilde{x}_{A,i} \left(1 - \frac{\sum_{j=1}^n \tilde{x}_{Q,j}}{K_Q} \right).
 \end{aligned}$$

Analogously, we have that $\tilde{x}_{A,i}$ satisfies the differential equation for the active cells in (23).

By the uniqueness of solutions, we have that, indeed, $\mathbf{x}(t) = \tilde{\mathbf{x}}(t)$ not only for $t = 0$, but for all $t > 0$. Therefore, in the deterministic model, the initial proportions of clone sizes remain unchanged over time. In particular, if all clones start with the same number of cells *in both compartments*, they remain equal in size for all time.

Not surprisingly, the situation changes when we move to the stochastic setting: via the analogous stochastic dynamics with identical parameters for all clones, random fluctuations inevitably break the symmetry and lead to the progressive loss of clones until eventually a single clone dominates.

As showed earlier, the *total* number of active and quiescent cells in the multi-clone context still follows the single-clone dynamics. Hence, as shown in Section 8.2, this aggregated measure in the stochastic model also closely follows its deterministic counterpart. In particular, when starting from the homeostatic equilibrium, the total population size remains approximately constant.

At the clonal level, however, stochasticity has a pronounced impact: even though they all share identical rates, individual cells differentiate, proliferate, and transition at different *moments*, causing the composition of the clonal pool gradually change over time. This leads to an eventual *decrease of the number of distinct clones*, while the *average size of the surviving clones increases*, keeping the total HSC population close to its deterministic homeostatic equilibrium (Figure 24).

This *clonal convergence effect*, i.e., the gradual reduction of clonal diversity compensated by the expansion of the surviving clones, is a well-documented phenomenon in diverse clonal systems [Klein and Simons, 2011], and, as we will shortly see, it is also observed in the data analyzed in this study.

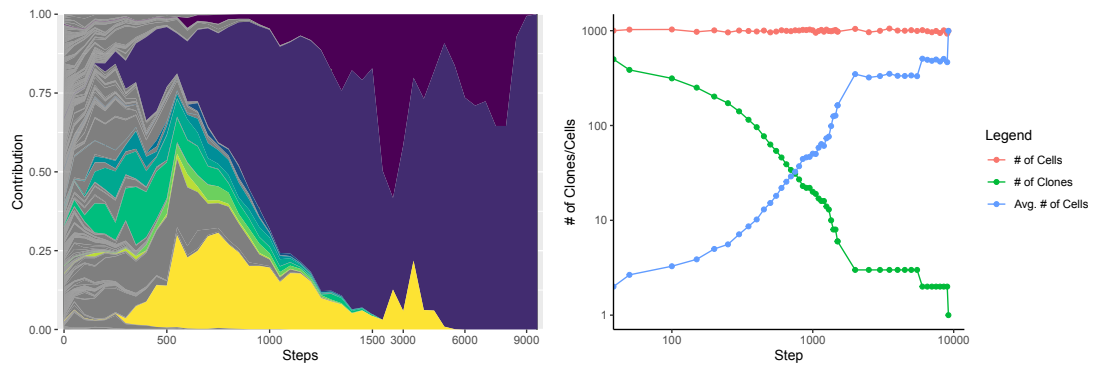


Figure 24 – Loss of clonal diversity in the stochastic model. Plots relative to a simulation *starting at deterministic equilibrium* with 500 clones with one active and one quiescent cell. (Left) Relative contribution of each surviving clone over time. Only the final 10 clones are shown in color. (Right) Number of clones (green), total number of cells (red) and average size of surviving clones (blue) over time. While the number of clones decreases sharply in the beginning and slowly in the end of the process, the total cell count remains almost unchanged.

9 MODEL FITTING

9.1 Clonality Data

Tracking hematopoietic clonal contributions across species, including mice [Gerrits et al., 2010, Lu et al., 2011, Cornils et al., 2014], non-human primates [Kim et al., 2014, Radtke et al., 2017], and humans [Biasco et al., 2016], has revealed rich dynamic behaviors. Following clonal marking events, some clones expand while others contract, resulting in a gradual loss of clonal diversity over time. This phenomenon occurs even in the absence of selective advantages and reflects the inherent stochastic competition among individual cells. The term *clonal conversion* refers to this dynamic of neutral reshaping of the clonal landscape.

In a 2023 study, Radtke et al. investigated clonal dynamics following autologous (“*of the same individual*”) HSC transplantation after full-body irradiation in pigtail macaques, tracking the process at the clonal level from early recovery to long-term engraftment. Their high-density sampling approach uses genetic barcodes to produce a rich dataset containing temporal clonal information for two monkeys (Z13264 and Z14004) spanning over four years of follow-up. For our study, we restricted our attention to unsorted white blood cells (WBC) gathered from the peripheral blood, since this is the most complete and consistent data source over the whole experimental period (Figure 25). Assuming that no clone has a differentiation bias, clonal composition of WBCs is a sufficient approximation of clonality at the HSC level.

This comprehensive dataset allows deriving diversity measures beyond conventional metrics, enabling detailed quantification of different aspects of the clonal dynamics over extended time periods. Precisely, from the extracted data we derived the following metrics (Figure 26):

- The *clonal diversity* is an approximation of the number of different clones which is still present in the recipient animal after transplantation. Since there are no further transplantations inducing additional clones during the follow up, we reason that the number of clones

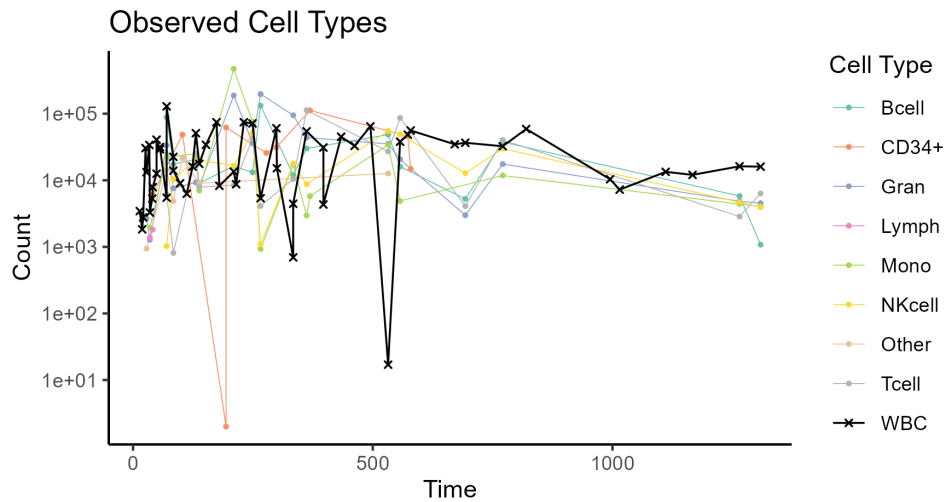


Figure 25 – Number of cells recovered for animals Z13264 and Z14004 from the data of Radtke et al. sorted by cell type. Points represent the observed number of cells other than white blood cells (WBC), represented by the x marks; lines connect observations for clarity. Colors denote the different cell types as indicated in the legend.

progressively declines over time. Therefore clonal diversity at a given time point refers to the number of remaining clones that are measured at this time point or any time point thereafter.

- The *single-time occurring (SO) clones* describe the fraction of clones that were observed exclusively at a single time point. Specifically, at a given time point, the SO fraction is defined as the number of clones detected only at that time point, divided by the total number of clones observed at the same time point. This metric approximates the abundance of small clones that are likely lost over time.
- The *relative clone size distribution* describes the frequency at which normalized clone sizes occur at any time point for which clonal measurements are available. Specifically, this metric is computed by dividing the size of each clone by the total number of reads at that time point, thereby obtaining a measure of relative clonal abundance. Clones are grouped according to their relative abundance on a logarithmic scale. The resulting histogram provides an empirical estimate of the underlying clone size distribution.

It is interesting to see that clonal diversity drops rapidly in the early post-transplant phase, most likely as a result of the transient contributions from short-lived, lineage-committed progenitors, i.e., cells that are able to proliferate, but not to the point of maintaining their own population, and differentiate into specific cell lineages. However, on longer time scales, clonal

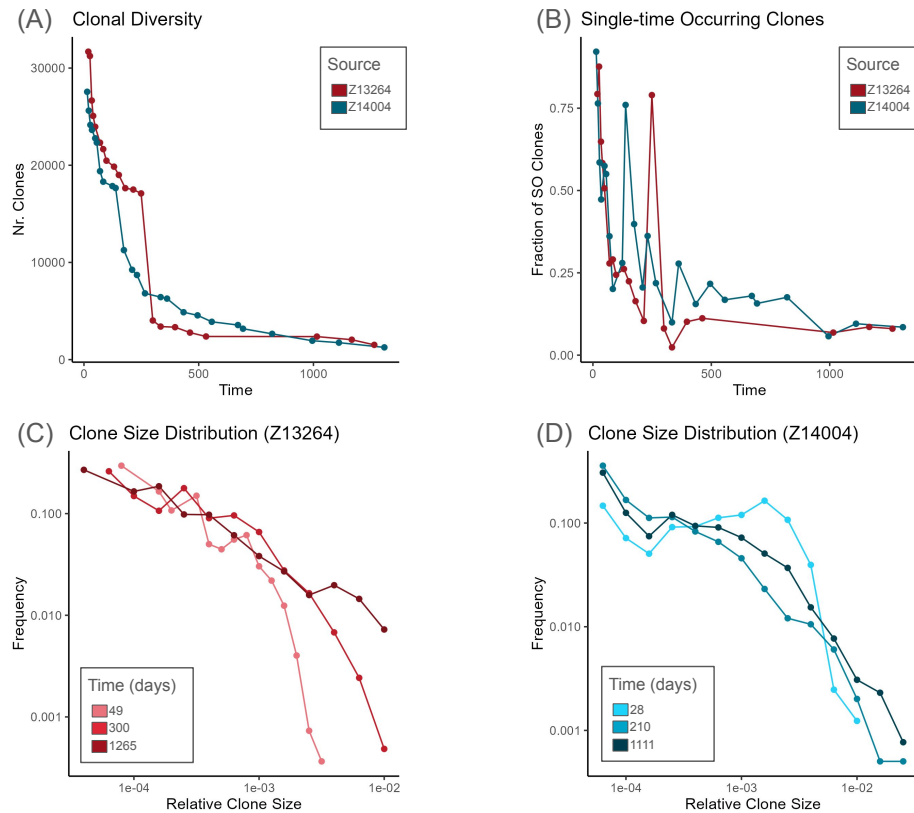


Figure 26 – Metrics of HSC clonality, as derived from Radtke et al. (A) Clone diversity, measured as the number of distinct clones detected over time. (B) Fraction of single-time occurring (SO) clones shown at each time point, i.e., the number of clones observed only at that time point divided by the total number of clones observed on that time point. (C-D) Clone size distribution for animals Z13264 and Z14004, respectively, shown as histograms of clone relative abundance (clone size divided by total of reads) on a log-log scale. Dots indicate observed data; lines connect observations for clarity. Colors denote animal id and sampling time as indicated in the legend.

diversity only decreases slowly, consistent with *clonal convergence effect* discussed in last chapter (Figure 26A). Similarly remarkable is the observation that even after several years post transplantation there is an occurrence of previously undetected clones that only occur once (Figure 26B). Changes in clone size distribution indicate that larger clones become more dominant while there is still a substantial fraction of smaller clones remaining (Figure 26C,D).

Samples taken on day 249 for Z13264 and on day 138 for Z14004 are remarkable as they have an increased sequencing depth and detect a higher number of especially small clones. This manifests by the peak of SO clones in Figure 26B and the corresponding sharp drop in clone diversity in Figure 26A.

Data for model fitting

Although the dataset from [Radtke et al., 2023] offers a valuable asset to study HSC dynamics, an important step of model fitting is selecting the data that is most relevant to that specific context. In that sense, not all of the derived data points are aligned with the scope of our model, which explains the homeostatic dynamics of HSCs.

We acknowledge that the early post-transplant phase is characterized by dynamic fluctuations and transient contributions from short-lived, lineage-committed progenitors. In humans, it has been shown that these cells dominate hematopoiesis for several months, while long-term HSCs only gradually take over full hematopoietic functionality, typically after 6 to 12 months [Biasco et al., 2016]. Likewise, studies in pigtailed macaques have shown that the hematopoietic system stabilizes over a period of 3 to 12 months before reaching homeostatic equilibrium [Radtke et al., 2017]. To avoid confounding short-term effects and capture more stable, long-term HSC-driven hematopoiesis, we focus our model analysis on time points beyond 360 days.

Moreover, as we are working with stochastic models, we aim to limit our analysis to more reliable measurements with lower intrinsic variability. For the SO clone fraction metric, we exclude the sample of animal Z13264 on time 532, as it only contains information on 17 cells. At such a small sample size even minor stochastic fluctuations across simulations tend to heavily impact this metric. Similarly, for the relative clone size distribution, we focus our fitting routine on small and medium-sized clones, with relative size typically around 10^{-4} and 10^{-3} . In the data, these clones typically account for more than 85% of the total number of clones, providing a robust basis for model comparison. Evaluated on a logarithmic scale, these abundant clones are less influenced by stochastic fluctuations, whereas the rarer large clones (relative size $10^{-3} \sim 10^{-2}$) display higher variability across simulations with the same parameters.

Furthermore, fitting of the clone size distribution is limited to two representative time points per animal (days 397 and 462 for Z13264, and 362 and 495 for Z14004). These have been selected based on high sequencing depth and temporal proximity to the beginning of the homeostatic phase. The limitation to two time points only is warranted to prevent an over-representation of this metric compared to the clonal diversity and the SO clones.

In Table 5 we present a summary of the available data points and those used for model fitting.

	Z13264	Z14004
Diversity	Total Points: 21 Fitted Points: 6	Total Points: 25 Fitted Points: 10
	Criteria: Sampling day > 360.	
SO Clone Contribution	Total Points: 21 Fitted Points: 5	Total Points: 25 Fitted Points: 10
	Criteria: Sampling day > 360 and exclude day 532 of Z13264.	
Clone Size Distribution	Total Points: 235 Fitted Points: 17	Total Points: 313 Fitted Points: 20
	Criteria: Selected days and relative clone size $\leq 10^{-3}$.	
	Selected Days: 397 and 462.	Selected Days: 362 and 495.
Summary	Total Points: 277 Fitted Points: 28	Total Points: 373 Fitted Points: 40

Table 5 – Summary of data points used for model fitting across the three clonal aspects for animals Z13264 and Z14004. Diversity and SO clones fits were restricted to time points after homeostatic equilibrium, with SO clones further requiring over 1,000 reads. Clone size distribution was fitted with data on small and medium-sized clones from two early homeostatic-phase time points with high read coverage per animal.

9.2 Simulation Routines and Optimization

In order to derive parameter configurations for each model class that optimally describe the experimental data, we minimize the squared sum of the distance between model and data on logarithmic scale (residual sum of squares, RSS) across all metrics.

To generate model predictions, we employ a customized R routine that simulates clonal dynamics and blood sampling over time. Technically, given the number of clones and the parameters described in the previous section, at each time-step, the simulation determines the number of cells from each clone that will perform a possible action based on the probabilities (Equations 8.5) with fixed carrying capacities of 10^5 cells per compartment when working with two compartments, and $2 \cdot 10^5$ (active) cells when working with a single compartment (compare Cosgrove et al. [2021] and Mitchell et al. [2022] for references in mice and men). The corresponding compartments are then updated accordingly. At time points when blood was sampled in the in vivo experiment, the same number of cells was randomly chosen from the active compartment and used for clonal analysis. From the progression of these cell pools over

time, we derive temporal representations of the clonal diversity, relative clone size distribution and the fraction of single-time occurring clones.

For the optimization, we compare the model's predictions with the actual data and quantify the difference in terms of the residuals for each data point (formally, the squared distance between model and data on logarithmic scale). Since the comparison is based on all three available metrics (see Chapter 9.1), we first normalize the residuals by the typical variation at each particular data point (obtained by repeatedly simulating with a previously defined parameter set), and then average the normalized residuals within each metric. Finally, the difference is calculated as the sum of the average normalized residuals across all three metrics. This approach balances the contribution of each metric, ensuring that no single one disproportionately influences the fitting process.

For a given set of model rates, this approach returns a numerical value indicating the quality of the model's fit to the data. We use this measure within a genetic algorithm [Holland, 1975] to explore the parameter space and identify optimal parameter sets for the initial number of clones and the event rates (p_A , d_A , t_{AQ} , t_{QA} and p_Q . See Supplementary Table 1). Technically, we use the GA package in R [Scrucca, 2013], which at each generation (iteration), generates a population of instances (i.e. set of parameters) which are evaluated with respect to the above quality measure. The best-performing instances (parameter sets within the population) are then used to generate the next generation, thereby gradually improving the fit over successive iterations. The process continues until convergence is reached, defined as the point where no further improvement in the fit is observed over several generations.

9.3 Model Selection Criteria

The approach presented in the last section enables the fit of a given model to the selected data, aiming to determine the best parameter sets to represent it. However, when comparing different models with varying numbers of parameters, it is important to consider not only the fit of each model but also their complexity. Indeed, a model may appear to fit the data better not because it captures an essential underlying mechanism of the process, but simply because the different parameters allow a finer adjust to the data.

In order to compare the different models, we employ two model quality metrics: Akaike Information Criterion (AIC) [Akaike, 1998] and the Bayesian Information Criterion (BIC) [Schwarz, 1978]. Both criteria are grounded in information theory and aim to balance two

competing aspects: the likelihood that a given model generated the observed data (encoded in the residuals of the best fit) and the penalty for model complexity (encoded in the number of parameters). In this sense, these criteria help to identify models that achieve a balance between goodness-of-fit and parsimony.

In the case of fits obtained by the Least Squares method, i.e., minimizing the Residual Sum of Squares, assuming normally distributed errors with constant variance (hence the normalization of residuals), the criteria reduce to the following forms (see, for instance, section 2.2 of Burnham and Anderson [2002]):

$$AIC = 2K + N \log \left(\frac{RSS}{N} \right), \quad BIC = K \log(N) + N \log \left(\frac{RSS}{N} \right),$$

where K is the number of estimated parameters, N the number of data points, and RSS the residual sum of squares of the best fit of a model.

10 ANALYSIS OF THE MODEL FIT

10.1 Model based description of the data using a one-compartment model

To assess how well different mechanistic assumptions capture the outlined clonal dynamics, we systematically evaluated the performance of the one- and two-compartment models introduced on Section 8.2.

The one-compartment model represents a simplified view of hematopoiesis, where all HSCs remain in an homogeneous state, in which they continuously proliferate and differentiate. This model is equivalent to the one presented by Radtke et al. [2017] and, provides a reasonable fit for the overall measure of clonal diversity. However, when attempting to fit all metrics combined, the overall fit deteriorates significantly. Notably, in order to reproduce even a minimal fraction of SO clones beyond the one-year horizon (Figure 27B), the model requires an unrealistically large number of initial clones. This, in turn, leads to an overestimation of total clone numbers (Figure 27A) and a failure to reproduce the observed clone size distribution (Figure 27C–D). Even under these conditions, the predicted SO clone fraction remains well below experimental values, which consistently shows the appearance of new clones even at later time points.

Optimizing the model with lower weight on the SO fraction yields visually improved fits, yet substantial discrepancies remain. Even the best fit achieved with this modification fails to capture the abundance of small clones, which appear on the left tail of the clone size distributions (Supplementary Figure 1). This limitation also affects the model-based approximation of the clonal diversity, leading to an underestimation of the overall number of clones.

Conceptually, these aspects highlight a key limitation of the one-compartment model: when only a homogeneously proliferating and differentiating HSC population is considered, small clones are vulnerable to the competitive pressure from larger clones, making them unable to persist over long periods. This limitation suggests that additional dynamics are necessary to

fully explain the observed clonal behavior.

10.2 Model based description of the data using a two-compartment model

We propose that an HSC intrinsic heterogeneity, namely a transiently quiescent state, improves the model predictions, as this additional state can account for the long-term persistence and occasional occurrence of smaller clones. To test this hypothesis, we introduce the transition dynamics, leading to an extended two-compartment model.

In the two-compartment model, long-term HSCs are able to transition between an active state, where they proliferate and differentiate, and a quiescent state, where they remain proliferatively inactive. Analyzing the relative clonal profile distribution generated by this model, we confirm that the inclusion of the quiescent state renders the expected result. Importantly, using this model we were able to capture the continued emergence of SO clones throughout the entire sampling period (Figure 27F). Moreover, both the overall number of clones (Figure 27E) and their contributions to the HSC pool (Figure 27G and 27H) were in close agreement with the experimental data.

To quantify and compare the goodness of fit for each model, we evaluated the residual sum of squares (RSS), the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). As shown in Figure 28, the RSS for the two-compartment model is consistently lower than that of the one-compartment model across all evaluated metrics, confirming the visual intuition that the two-compartment model provides a significantly better fit to the data. Furthermore, both AIC and BIC scores are lower for the two-compartment model, indicating that even after penalizing the additional parameters, the two-compartment model remains better suited to describe the combined metrics (Figure 31).

We still notice an underestimation in the SO clone fraction during an intermediate period for animal Z14004, suggesting that additional mechanisms may be required to fully account for the increased presence of these small clones in this window. Despite this limitation, the two-compartment model successfully reproduces key features of the clonal data, reinforcing the idea that transitions between active and quiescent states are essential for shaping long-term persistence of clonal diversity, appearance and contribution patterns.

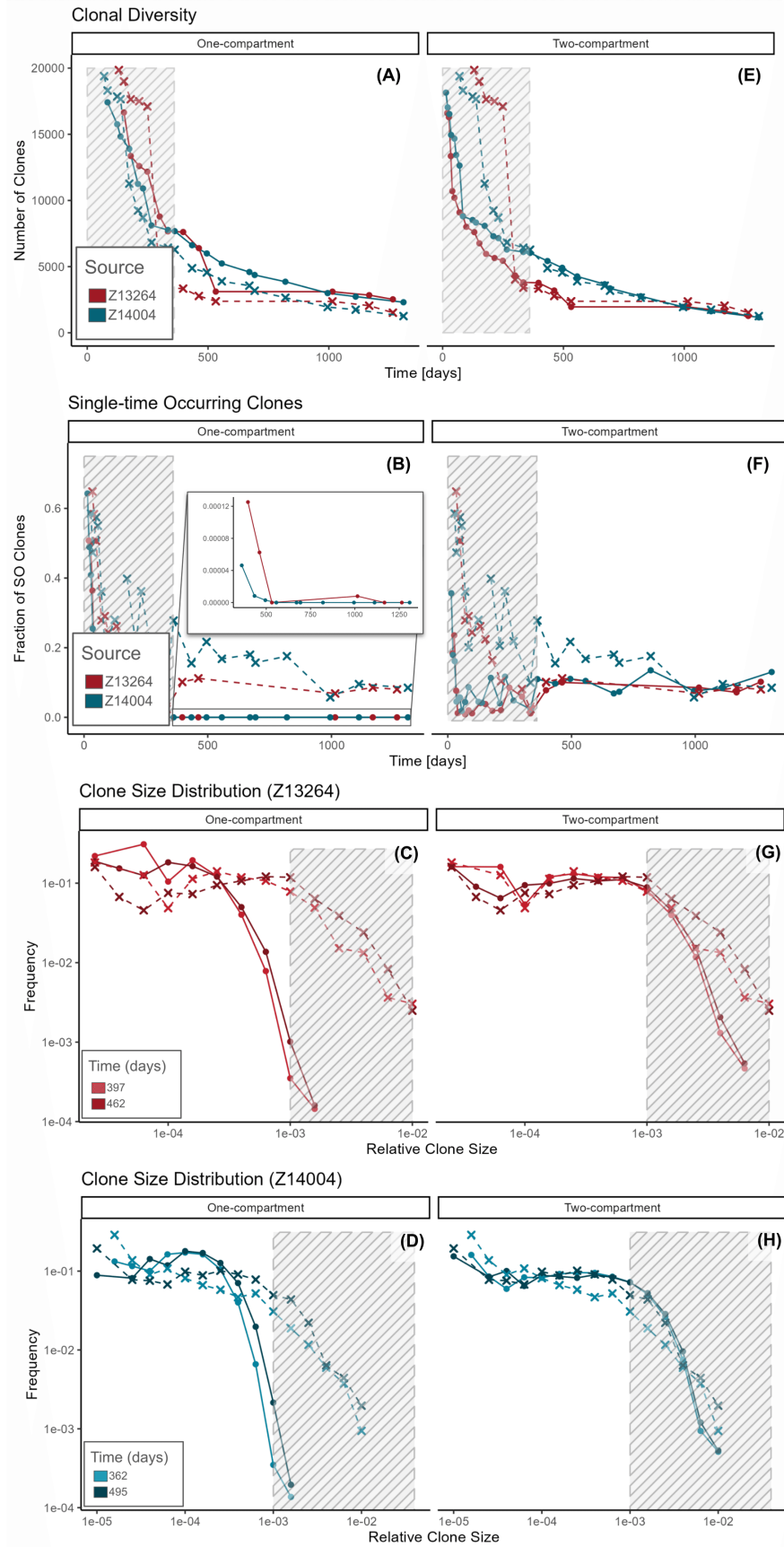


Figure 27 – Best fit for the one (A-D) and two-compartment models (E-H) optimized with respect to the metrics derived from data by Radtke et al. [2023] (See Supplementary Table 2). Dots indicate model predictions with lines connecting observations for clarity; experimental data is indicated by x marks. Striped areas indicate data which is not used for fitting.

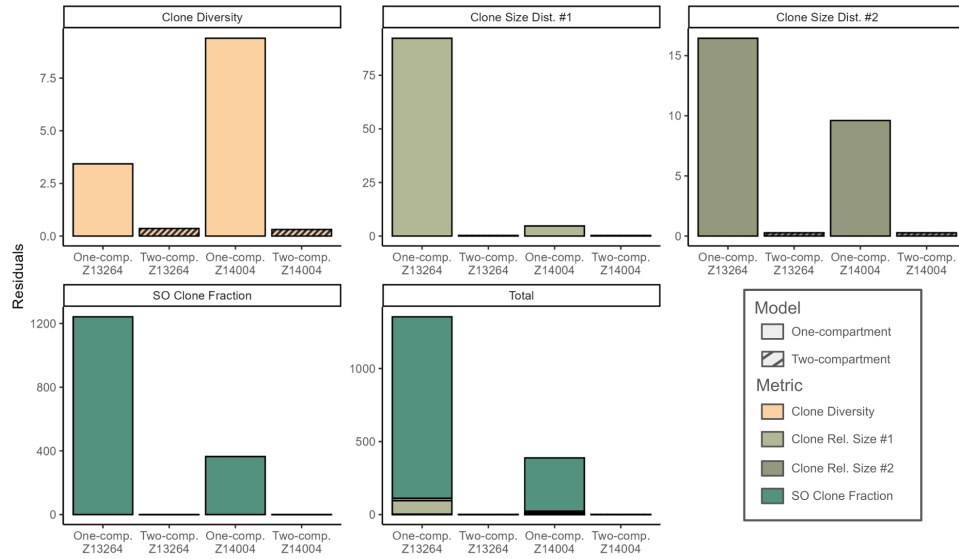


Figure 28 – Residual Sum of Squares (RSS) of best fits of one and two-compartment models. Stacked bar plots show the RSS of each metric - diversity, SO clone fraction and clone size distribution in two different time points (see Table 5) - for two animals (Z13264 and Z14004). The two-compartment model provides consistently lower RSS across both animals and all clonal metrics, suggesting improved overall performance.

Mean Time to Events

Evaluating the average time between events for the optimized two-compartment model, we observe that division-related processes (proliferation and differentiation) occur on a much faster timescale compared to state transitions (activation and deactivation). Namely, it takes on average one week for a cell to divide, while it takes on average several years for a cell to activate (Table 6). At first glance, this disparity might suggest that transitions can be neglected. However, omitting them effectively reduces the model to the one-compartment version, which cannot explain long-term persistence of small clones.

It is interesting to note that transition times follow an exponential distribution (Figure 29), implying that early transition events are still frequent despite long averages. While 50% of the cells transition within the first three and a half years, the fastest 25% of transitions occur within the first one and a half years. This characteristic of exponential distributions explains why such transitions, though rare in homeostatic situations, substantially influence the overall system behavior. In summary, although state transitions occur much less frequently than division events, their presence is crucial for the long-term dynamics of the hematopoietic system.

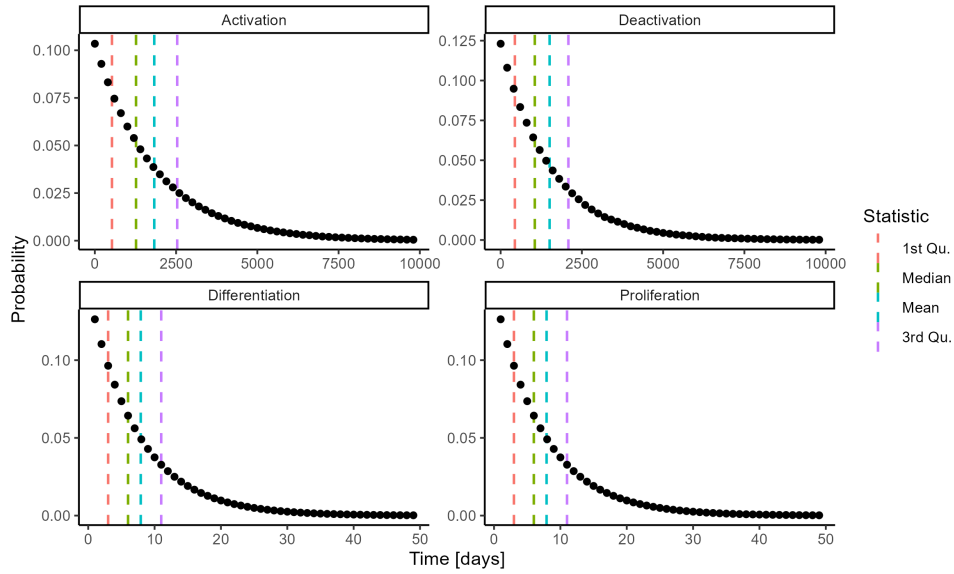


Figure 29 – Probability distributions of waiting times for individual cell events (activation, deactivation, differentiation, and proliferation) in the two-compartment model, estimated using best-fit parameters. Black dots indicate the probability that a single cell performs the indicated event after a given time. Vertical dashed lines show the 1st quartile (red), median (green), mean (blue), and 3rd quartile (purple) of each distribution, as summarized in Table 6. Due to the exponential nature of these distributions, most events take place earlier than the mean waiting time, as shown by the 1st quartile and median statistics.

Action	1st Quartile	Median	Mean	3rd Quartile
Activation	527	1268	1828	2535
Deactivation	438	1051	1507	2085
Differentiation	3	6	8	11
Proliferation	3	6	8	11

Table 6 – Summary statistics of simulated waiting times for cell events in the two-compartment model. Estimates were derived from simulations run for 100,000 time steps using best-fit parameters.

10.3 Different Mechanisms of HSC Proliferation

In the scope of the two-compartment model, the quiescent state is associated with niche regulation that maintains cells in a non-proliferative condition, while active cells divide independently of any previous actions (niche-independent proliferation). Recent work [Pedersen et al., 2023, Harnisch et al., 2024], including our own, has suggested an alternative mechanism in which proliferation is directly linked to the transition from quiescence to activation (*niche-dependent proliferation*). This mechanism can be particularly important on scenarios of therapy that target proliferative cells (e.g., Acute Myeloid Leukemia treatments with anthracycline [Venkatesh and Kasi, 2025]), where a longer permanence of some clone in the quiescent compartment translates to a fitness advantage.

One formulation of this idea, derived from the model proposed by Pedersen et al. [2023], is attributing cell division not to the active compartment itself, but rather to a fraction p_Q of cells that leave quiescence. To combine these different proliferation mechanisms, a unified model that incorporates both processes has been proposed (Figure 30).

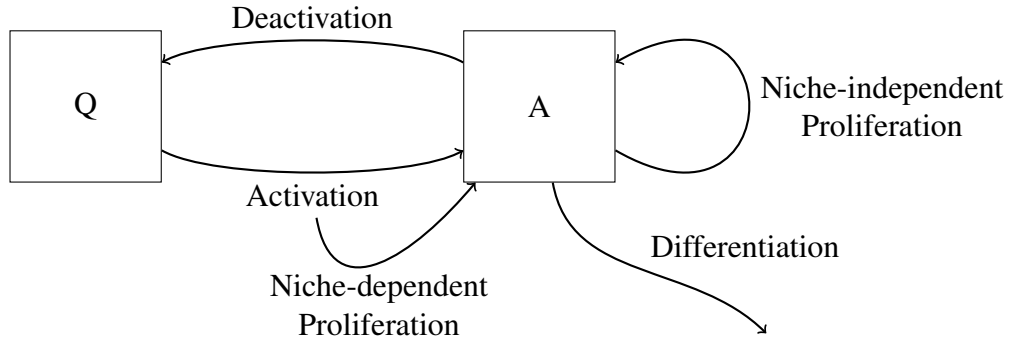


Figure 30 – Schematic representation of the unified model for HSCs dynamics. In addition to random divisions of active cells (*niche-independent proliferation*), a fraction of cells exiting quiescence immediately undergo proliferation (*niche-dependent proliferation*). The two-compartment model can be recovered by removing the niche-dependent proliferation. A model equivalent to the one proposed by Pedersen et al. [2023] is recovered by removing the niche-independent proliferation.

Precisely, this model is given by the ODE

$$\begin{cases} x'_Q = t_{AQ}x_A \left(1 - \frac{x_Q}{K_Q}\right) - t_{QA}x_Q \left(1 - \frac{x_A}{K_A}\right), \\ x'_A = -t_{AQ}x_A \left(1 - \frac{x_Q}{K_Q}\right) + t_{QA}x_Q \left(1 - \frac{x_A}{K_A}\right) + \\ \quad + (p_A x_A + p_Q t_{AQ} x_Q) \left(1 - \frac{x_A}{K_A}\right) - d_A x_A. \end{cases} \quad (10.1)$$

Notably, the two-compartment model discussed earlier can be seen as a particular case of the unified model with $p_Q = 0$, while a model composed only of niche-dependent proliferation is achieved with $p_A = 0$. We raise the question whether this unified model is better suited to explain the long latency of hematopoietic clones after transplantation.

Incorporating this additional parameter p_Q in our fitting routine for the unified model, we observe that it achieves a fit similar to that of the earlier two-compartment model. Both reproduce key trends across all metrics with comparable accuracy (Supplementary Figure 2). This visual similarity is also reflected in overlapping ranges of the RSS (Figure 31A). Model selection criteria such as AIC and BIC (Figure 31B and C) further reinforce this interpretation, showing only small differences between the two models. This suggests that niche-dependent proliferation

may not be robustly identified from the shape of clonal dynamics under steady-state conditions, while a more prominent role in perturbed or stressed hematopoietic environments should not be excluded. However, comparing these metrics with the corresponding one-compartment model further supports the hypothesis that niche influence, whether through a dormant compartment or via its role in proliferation, is pivotal to explain the observed dynamics.

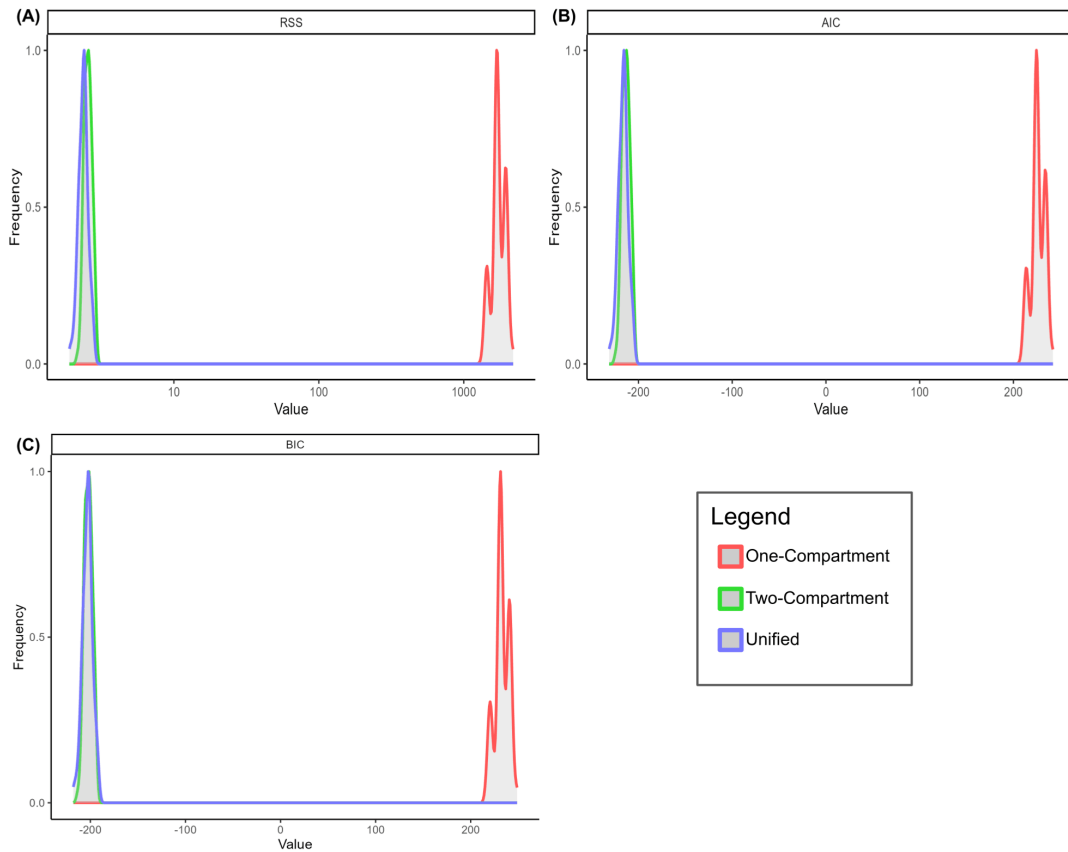


Figure 31 – Distribution of Residual Sum of Squares (RSS) and Akaike Information Criterion (AIC) values from 100 simulations using the best-fit parameters of the one-compartment, two-compartment and unified models. Colors represent the different models, as in legend.

11 DISCUSSION

In this study, we employed high-resolution clonal tracking of hematopoietic cells in macaques to demonstrate that an additional layer of heterogeneity among HSCs is necessary to particularly explain the late emergence of smaller clones. Precisely, referring to a previously established concept, in which HSCs reversibly transit between a proliferatively inactive/quiescent state and an active, proliferative state, clearly outperforms simpler models that consider HSCs as a single, homogeneous population. Using a set of particular metrics to describe key features of the temporal clonal data set, we show that while a model incorporating only proliferation and differentiation can reproduce the gradual loss of clonal diversity, it fails to capture the observed distribution of clone sizes and entirely misses the long-term persistence of small clones. Introducing heterogeneity into the HSC population by distinguishing between actively cycling cells and niche-bound quiescent cells, yielded a remarkably improved fit to all metrics. Our results confirm that the addition of a quiescent compartment allows for a sustained coexistence of small and large clones and, in particular, the long-term emergence of single-time occurring clones throughout the time course, which is in good agreement with experimental observations.

These findings highlight the importance of acknowledging a duality in the appearance of HSCs, to be either in an active or quiescent state [Viale and Pelicci, 2009]. This concept has been long discussed on a theoretical level [Loeffler and Roeder, 2002, Glauche et al., 2009], and is increasingly gaining recognition to explain biological data sets [Wilson et al., 2008, Glauche et al., 2009, Essers et al., 2009, Morcos et al., 2022, Ibneeva and Grinenko, 2024]. A prominent example arises from label dilution experiments where the biphasic decline of label-retaining HSCs, characterized by an initial rapid loss followed by a slower secondary phase, can only be reconciled by assuming that a substantial subset of HSCs exhibits markedly slower turnover dynamics.

It is an open conceptual question whether quiescence and activation in HSCs are truly

two distinct and discrete states, or whether they are better understood as two extremes along a continuous spectrum of functional activity. In the latter view, individual HSCs may gradually shift their proliferative activity, potentially in response to external cues, but also to spatial and metabolic conditions. This would be in contrast to a sharp transition between binary states, as we considered it in our analysis. However, the two-state model offers conceptual clarity, is easier to represent in terms of mathematical or conceptual models, and also aligns well with experimentally accessible features such as flow cytometry based cell classification. In turn, it may only represent a coarse-grained abstraction of a more gradual biological reality.

While the total number of cells and the division rates estimated for the two-compartment model is compatible with other data sources [Wilson et al., 2008, Glauche et al., 2009, Catlin et al., 2011, Busch et al., 2015, Biasco et al., 2016, Cosgrove et al., 2021], we notice that this model describes the observed data with fewer clones compared to the one-compartment model. It appears that the latter model structure tries to compensate for the rapid loss of clones by a higher initial clone number. Furthermore, while transitions between active and quiescent states emerge as a crucial mechanism to explain the data, their rate is much lower than the cell division rate. Still, the model fits clearly indicate that the slow but continuous flux of cells between compartments is a requirement for the steady renewal of the active pool, thereby accounting for the persistent emergence of new clones even at late sampling times.

Although our model provides a good overall fit to the data, some aspects remain unexplained. For instance, it can not simultaneously reproduce the rapid loss of diversity observed shortly after transplantation and the slower dynamics that dominate during later phases. Moreover, while the model predicts the existence of single-occurring clones, it fails to reproduce their observed frequency at intermediate time points. These limitations suggest the need for more refined models that incorporate, for instance, heterogeneity in clonal capacities or context-dependent regulatory mechanisms. Importantly, such extensions can be naturally accommodated within the same two-compartment framework, offering a clear path forward for future work.

In summary, our analysis demonstrates that describing HSC as a heterogeneous population comprising two complementary functional activity states provides a suitable explanation for the resulting clonal dynamics. It supports the notion of reversible HSC quiescence as a key element to understand long-term maintenance of clonal diversity. Our approach highlights the potential of using clonality data to explore the underlying mechanisms of HSC interaction and regulation, enabling the simultaneous evaluation of several different metrics in a unified framework and

leading to more robust models.

A SUPPLEMENTAL FIGURES AND TABLES

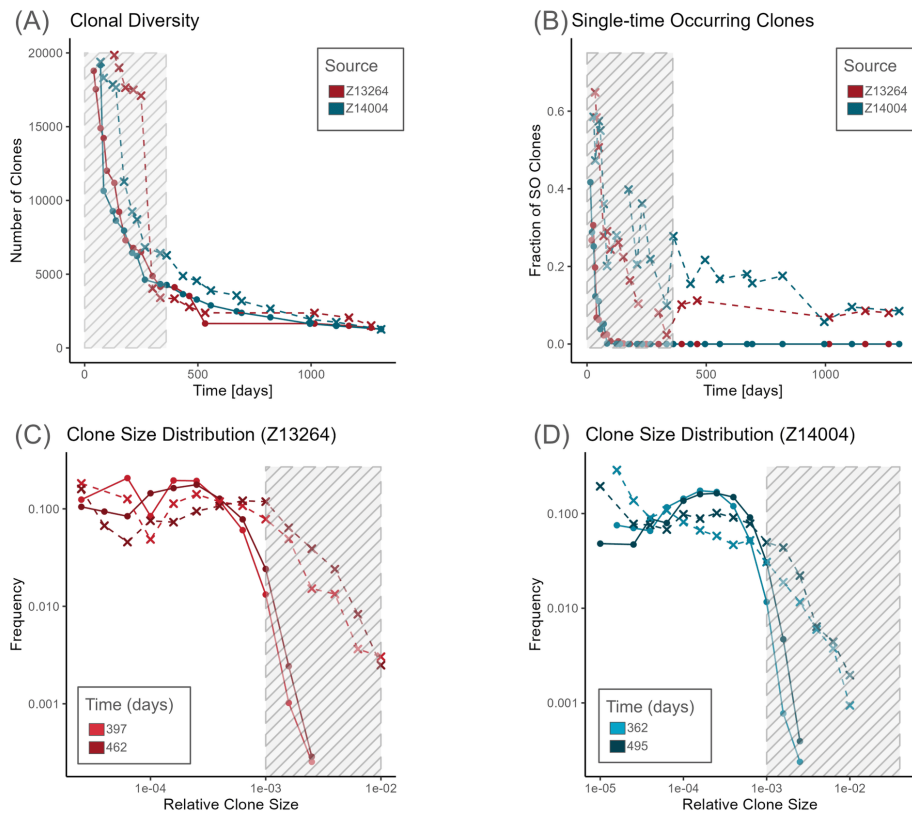


Figure 1 – Best fit of one-compartment model with SO clones residuals downweighted by a factor of 0.05 optimized with respect to the metrics derived from data by Radtke et al. [2023] (See Supplementary Table 2). Dots indicate model predictions with lines connecting observations for clarity; experimental data is indicated by x marks. Striped areas indicate data which is not used for fitting.

Parameter Name	Definition	Default Value
d_A	Base differentiation rate for active cells.	–
p_A	Base proliferation rate for active cells.	–
t_{AQ}	Base deactivation rate for active cells.	0
t_{QA}	Base activation rate for quiescent cells.	0
$clone_mult$	Multiplicative factor for the number of clones. $Nr. Clones = clone_mult \cdot 10^4$	–
p_Q	Probability of a quiescent cell to proliferate when activating.	0

Table 1 – Summary of parameters used in the optimization routine. The model rates (p_A , d_A , t_{QA} , t_{AQ} , and p_Q) are directly optimized. Carrying capacities (K_A and K_Q) are fixed to 10^5 cells per compartment when working with two compartments, and $2 \cdot 10^5$ (active) cells when working with a single compartment. The total number of clones is scaled by $clone_mult$. Default values are shown where applicable.

Parameter	One-cpt.	One-cpt. (Dw. SO)	Two-cpt.	Unified
Nr. Clones	$17.67 \cdot 10^4$	$9.07 \cdot 10^4$	$9.72 \cdot 10^4$	$10.35 \cdot 10^4$
d_A	0.045	0.088	0.126	0.108
p_A	0.435	0.620	0.708	0.494
k_A	$2 \cdot 10^5$	$2 \cdot 10^5$	10^5	10^5
t_{AQ}	0	0	0.061	0.055
t_{QA}	0	0	0.003	0.002
k_Q	0	0	10^5	10^5
p_Q	0	0	0	0.054

Table 2 – Best fit parameters for three models of HSC clonal dynamics. Values were determined by a customized R routine based on genetic algorithms.

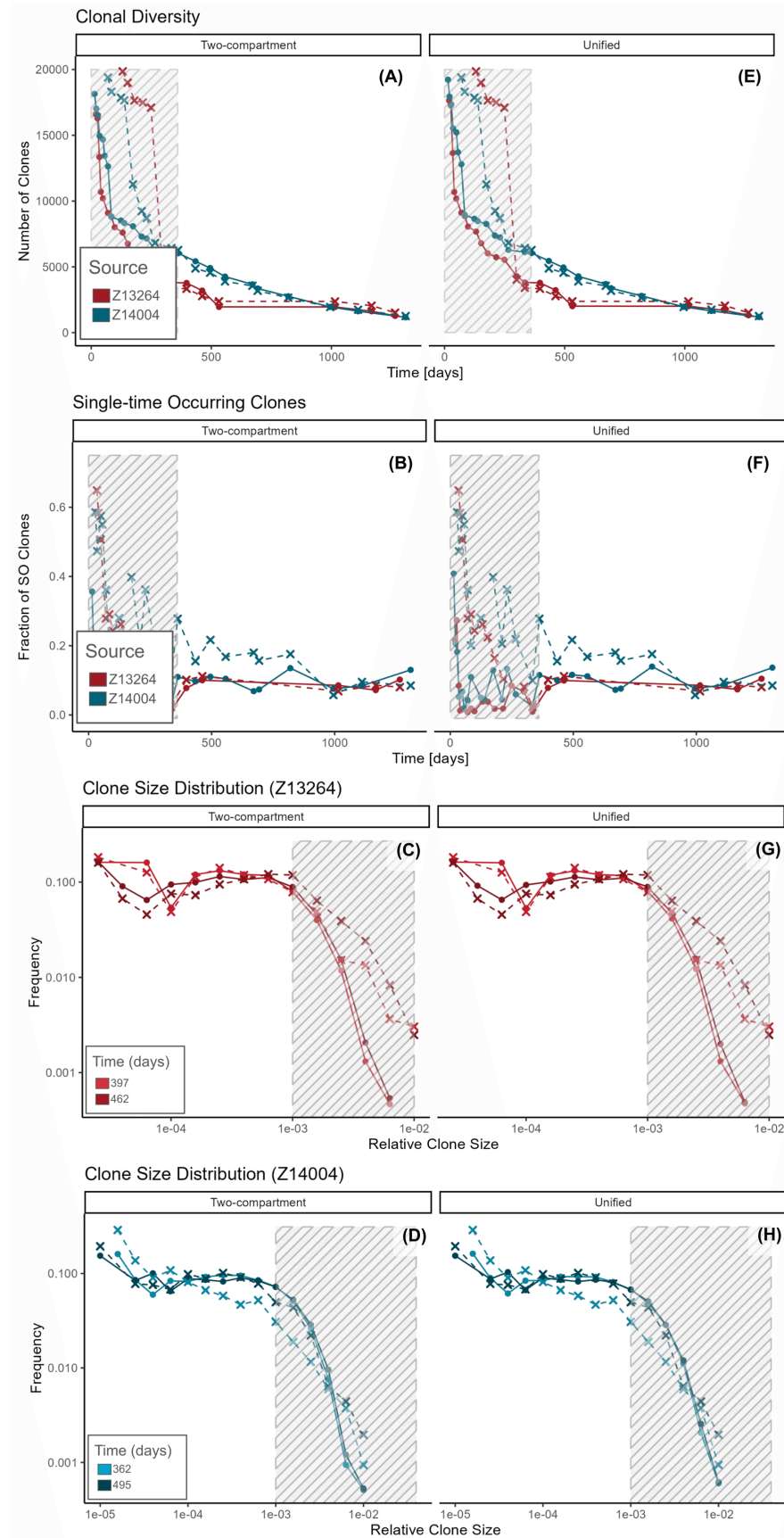


Figure 2 – Best fit for the two-compartment (A-D) and unified models (E-H) optimized with respect to the metrics derived from data by Radtke et al. [2023] (See Supplementary Table 2). Dots indicate model predictions with lines connecting observations for clarity; experimental data is indicated by x marks. Striped areas indicate data which is not used for fitting.

PART IV

CONCLUSION AND FUTURE WORKS

CONCLUSION AND FUTURE WORKS

The work presented in this thesis illustrates how mathematical modeling, grounded in biological principles and complemented by computational methods, can effectively address central questions in cell biology. For that end, two distinct problems were investigated: the design and stability of adaptive cancer-therapy routines, and the mechanisms underlying long-term HSC clonal dynamics. Across these two biological applications, it was shown how population models can be tailored to address complex problems both deepening our understanding of the mechanisms that govern them, and highlighting key factors to be considered in their potential clinical translation.

In the context of adaptive therapy, a formal analysis of treatment models based on Lotka–Volterra and Replicator dynamics was presented, with the derivations of biologically relevant conditions under which cyclic routines can be designed and the presentation an algorithm to do so. Furthermore, a conjecture linking the stability of the cycle with the stability of the average treatment was stated, which, if proved true, may lead to important constraints when searching for optimal cyclic therapies. Finally, the results were extended to stochastic settings, showing that while all stochastic cycles break, those corresponding to unstable deterministic routines deteriorate significantly faster. Together, these contributions advance the theoretical understanding of adaptive therapy and raise hypotheses under which cyclic treatment strategies may be robust enough to inform clinical applications.

In the study of hematopoietic stem cell dynamics, it was argued that modeling HSCs dynamics with reversible transitions between active and quiescent states provides a suitable explanation for their long-term clonal behavior, reinforcing the established view of a quiescent compartment as a key element in hematopoiesis. The results presented here indicate that this mechanism is particularly important in maintaining long-term clonal diversity as, under homeostatic conditions, transitions between compartments occur on much longer timescales than cell

divisions. Moreover, this work highlights the value of clonal data as a tool for studying cellular mechanisms: by enabling the simultaneous evaluation of multiple metrics, these data support the development of models that are more faithful to biological reality.

Naturally, in both research fronts, important challenges remain. For adaptive therapy, a central open problem is to rigorously determine the conditions that ensure cycle stability. Progress on this question would greatly strengthen the ability of clinicians to assess when adaptive therapy is a viable option. Moreover, while our analysis provides a careful theoretical foundation, its empirical validation in preclinical and clinical settings is still lacking and represents a clear direction for future work. On the HSC side, although our models satisfactorily captured key features of the experimental data, a mismatch on short-term dynamics and in the magnitude of single-occurring clones contributions suggest that additional mechanisms may be necessary to fully account for the observed dynamics. Identifying and incorporating such mechanisms will be essential to advance the modeling of a global hematopoietic dynamics. On the theoretical front, an important direction is to refine our formal understanding of the stochastic dynamics, including the asymptotic relationship between stochastic and deterministic models presented here.

An important next step that is already planned, is to extend the knowledge gained in this thesis to the study of critical biological mechanisms in the evolution and treatment of hematological tumors. Specifically, future work aims to evaluate the impact of heterogeneity in the proliferation of HSCs during the progression of acute myeloid leukemia, and, on another front, to identify indicators of clinical success or relapse in CAR-T cell therapy for multiple myeloma. The experience acquired through the development of this thesis provides a solid foundation for tackling these challenges and, more broadly, to guide future research efforts on tightening the gap between theoretical models and clinical applications.

BIBLIOGRAPHY

- Hiroto Akaike. Information Theory and an Extension of the Maximum Likelihood Principle, pages 199–213. Springer New York, New York, NY, 1998. ISBN 978-1-4612-1694-0. doi: 10.1007/978-1-4612-1694-0_15. URL https://doi.org/10.1007/978-1-4612-1694-0_15. Cited on page 107.
- Athena Aktipis, Virginia S.Y. Kwan, Kathryn A. Johnson, Steven L. Neuberg, and Carlo C. Maley. Overlooking evolution: A systematic analysis of cancer relapse and therapeutic resistance research. PLoS ONE, 6, 2011. ISSN 19326203. doi: 10.1371/journal.pone.0026100. Cited on page 50.
- Tibor Antal and István Scheuring. Fixation of strategies for an evolutionary game in finite populations. Bulletin of Mathematical Biology, 68, 2006. ISSN 00928240. doi: 10.1007/s11538-006-9061-4. Cited 2 times on pages 48 and 77.
- Kendall E.. Atkinson. An introduction to numerical analysis. Wiley, 1989. ISBN 9780471500230. Cited on page 94.
- Michel Benaïm and Jörgen W. Weibull. Deterministic approximation of stochastic evolution in games. Econometrica, 71, 2003. ISSN 00129682. doi: 10.1111/1468-0262.00429. Cited 3 times on pages 43, 44, and 77.
- Luca Biasco, Danilo Pellin, Serena Scala, Francesca Dionisio, Luca Basso-Ricci, Lorena Leonardelli, Samantha Scaramuzza, Cristina Baricordi, Francesca Ferrua, Maria Pia Cicalese, Stefania Giannelli, Victor Neduva, David J Dow, Manfred Schmidt, Christof Von Kalle, Maria Grazia Roncarolo, Fabio Ciceri, Paola Vicard, Ernst Wit, Clelia Di Serio, Luigi Naldini, and Alessandro Aiuti. In vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases. Cell stem cell, 19:107–19, 7

2016. ISSN 1875-9777. doi: 10.1016/j.stem.2016.04.016. Cited 4 times on pages 83, 102, 105, and 117.
- Immanuel M. Bomze. Lotka–volterra equation and replicator dynamics: A two-dimensional classification. *Biological Cybernetics*, 48:201–211, 1983. Cited on page 40.
- Karol Bukowski, Mateusz Kciuk, and Renata Kontek. Mechanisms of multidrug resistance in cancer chemotherapy. *International Journal of Molecular Sciences*, 21, 2020. ISSN 14220067. doi: 10.3390/ijms21093233. Cited on page 50.
- Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, 2nd edition, 2002. doi: 10.1007/b97636. Cited on page 108.
- Katrin Busch, Kay Klapproth, Melania Barile, Michael Flossdorf, Tim Holland-Letz, Susan M Schlenner, Michael Reth, Thomas Höfer, and Hans-Reimer Rodewald. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, 518:542–6, 2 2015. ISSN 1476-4687. doi: 10.1038/nature14242. Cited on page 117.
- Sandra N Catlin, Lambert Busque, Rosemary E Gale, Peter Gutter, and Janis L Abkowitz. The replication rate of human hematopoietic stem cells in vivo. *Blood*, 117:4460–6, 4 2011. ISSN 1528-0020. doi: 10.1182/blood-2010-08-303537. Cited on page 117.
- Carmen Chicone. *Ordinary Differential Equations with Applications*, volume 34. Springer New York, 2006. ISBN 978-0-387-30769-5. doi: 10.1007/0-387-35794-7. Cited on page 90.
- Kerstin Cornils, Lars Thielecke, Svenja Hüser, Michael Forgber, Michael Thomaschewski, Nadja Kleist, Kais Hussein, Kristoffer Riecken, Tassilo Volz, Sebastian Gerdes, Ingmar Glauche, Andreas Dahl, Maura Dandri, Ingo Roeder, and Boris Fehse. Multiplexing clonality: combining rgb marking and genetic barcoding. *Nucleic acids research*, 42:e56, 4 2014. ISSN 1362-4962. doi: 10.1093/nar/gku081. Cited 2 times on pages 82 and 102.
- Jason Cosgrove, Lucie S P Hustin, Rob J de Boer, and Leïla Perié. Hematopoiesis in numbers. *Trends in immunology*, 42:1100–1112, 12 2021. ISSN 1471-4981. doi: 10.1016/j.it.2021.10.006. Cited 2 times on pages 106 and 117.
- Sean M Cullen, Allison Mayle, Lara Rossi, and Margaret A Goodell. Hematopoietic stem cell development: an epigenetic journey. *Current topics in developmental biology*, 107:39–75,

2014. ISSN 1557-8933. doi: 10.1016/B978-0-12-416022-4.00002-0. Cited 2 times on pages 15 and 82.
- Rajvir Dua, Yongqian Ma, and Paul K. Newton. Are adaptive chemotherapy schedules robust? a three-strategy stochastic evolutionary game theory model. *Cancers*, 13, 2021. ISSN 20726694. doi: 10.3390/cancers13122880. Cited 9 times on pages 14, 51, 52, 53, 61, 68, 77, 78, and 79.
- Rick Durrett. *Essentials of Stochastic Processes*. Springer, New York, 3rd edition, 2019. ISBN 978-3-030-18532-5. Cited 2 times on pages 23 and 24.
- Leah Edelstein-Keshet. *Mathematical Models in Biology*. Society for Industrial and Applied Mathematics, 1 2005. ISBN 978-0-89871-554-5. doi: 10.1137/1.9780898719147. Cited on page 96.
- Hideo Ema, Yohei Morita, and Toshio Suda. Heterogeneity and hierarchy of hematopoietic stem cells. *Experimental hematology*, 42:74–82.e2, 2 2014. ISSN 1873-2399. doi: 10.1016/j.exphem.2013.11.004. Cited 2 times on pages 15 and 82.
- Rena Emond, Jason I. Griffiths, Vince Kornél Grolmusz, Aritro Nath, Jinfeng Chen, Eric F. Medina, Rachel S. Sousa, Timothy Synold, Frederick R. Adler, and Andrea H. Bild. Cell facilitation promotes growth and survival under drug pressure in breast cancer. *Nature Communications*, 14(1), 2023. doi: 10.1038/s41467-023-39242-6. Cited on page 80.
- Marieke A G Essers, Sandra Offner, William E Blanco-Bose, Zoe Waibler, Ulrich Kalinke, Michel A Duchosal, and Andreas Trumpp. Ifn α activates dormant haematopoietic stem cells in vivo. *Nature*, 458:904–8, 4 2009. ISSN 1476-4687. doi: 10.1038/nature07815. Cited 2 times on pages 83 and 116.
- Artur C Fassoni, Christoph Baldow, Ingo Roeder, and Ingmar Glauche. Reduced tyrosine kinase inhibitor dose is predicted to be as effective as standard dose in chronic myeloid leukemia: a simulation study based on phase iii trial data. *Haematologica*, 103:1825–1834, 11 2018. ISSN 1592-8721. doi: 10.3324/haematol.2018.194522. Cited on page 84.
- Eliza M. Ferreira and Armando G.M. Neves. Fixation probabilities for the Moran process with three or more strategies: general and coupling results. *Journal of Mathematical Biology*, 81, 2020. ISSN 14321416. doi: 10.1007/s00285-020-01510-0. Cited 3 times on pages 36, 41, and 43.

Frederico Ferreira. Campos Filho. Algoritmos numericos. LTC, 2007. ISBN 9788521615378. Cited on page 94.

Robert A. Gatenby, Ariosto S. Silva, Robert J. Gillies, and B. Roy Frieden. Adaptive therapy. Cancer Research, 69, 2009. ISSN 00085472. doi: 10.1158/0008-5472.CAN-08-3658. Cited 2 times on pages 14 and 50.

Craig Gedye and Vishal Navani. Find the path of least resistance: Adaptive therapy to delay treatment failure and improve outcomes. Biochimica et Biophysica Acta - Reviews on Cancer, 1877, 2022. ISSN 18792561. doi: 10.1016/j.bbcan.2022.188681. Cited 3 times on pages 14, 50, and 78.

Alice Gerrits, Brad Dykstra, Olga J Kalmykova, Karin Klauke, Evgenia Verovskaya, Mathilde J C Broekhuis, Gerald de Haan, and Leonid V Bystrykh. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood, 115:2610–8, 4 2010. ISSN 1528-0020. doi: 10.1182/blood-2009-06-229757. Cited 2 times on pages 82 and 102.

Ingmar Glauche, Kateri Moore, Lars Thielecke, Katrin Horn, Markus Loeffler, and Ingo Roeder. Stem cell proliferation and quiescence—two sides of the same coin. PLoS computational biology, 5:e1000447, 7 2009. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000447. Cited 5 times on pages 15, 82, 84, 116, and 117.

Adam Greenbaum, Yen-Michael S Hsu, Ryan B Day, Laura G Schuettelpelz, Matthew J Christopher, Joshua N Borgerding, Takashi Nagasawa, and Daniel C Link. Cxcl12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. Nature, 495:227–30, 3 2013. ISSN 1476-4687. doi: 10.1038/nature11926. Cited on page 82.

Charles M. Grinstead and J. Laurie Snell. Introduction to Probability. American Mathematical Society, 2nd edition, 2003. Cited on page 24.

Christoph Harnisch, Ingmar Glauche, Friedemann Uchner, Ingo Roeder, and Artur César Fassoni. Mechanisms of leukemic stem cell interaction in aml: insights from complementary mathematical modeling approaches., 2 2024. Cited 3 times on pages 84, 89, and 113.

Morris W. Hirsch, Stephen Smale, and Robert L. Devaney. Differential Equations, Dynamical Systems, and an Introduction to Chaos. Academic Press, Boston, 3rd edition, 2013. ISBN 978-0123820105. Cited 3 times on pages 18, 19, and 32.

- J. Hofbauer and K. Sigmund. Evolutionary Games and Population Dynamics. Cambridge University Press, 1998. doi: 10.1017/CBO9781139173179. Cited 6 times on pages 25, 27, 32, 33, 39, and 52.
- H. Hoffmann, C. Thiede, I. Glauche, M. Bornhaeuser, and I. Roeder. Differential response to cytotoxic therapy explains treatment dynamics of acute myeloid leukaemia patients: insights from a mathematical modelling approach. Journal of The Royal Society Interface, 17(170): 20200091, 2020. doi: 10.1098/rsif.2020.0091. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2020.0091>. Cited 2 times on pages 84 and 88.
- J.H Holland. Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, 1975. Cited on page 107.
- Liliia Ibneeva and Tatyana Grinenko. Dissecting dormancy and quiescence in hematopoietic stem cells. Frontiers in Hematology, 3, 9 2024. ISSN 2813-3935. doi: 10.3389/frhem.2024.1401713. Cited 2 times on pages 83 and 116.
- Sangu Kim, Namshin Kim, Angela P Presson, Mark E Metzger, Aylin C Bonifacino, Mary Sehl, Samson A Chow, Gay M Crooks, Cynthia E Dunbar, Dong Sung An, Robert E Donahue, and Irvin S Y Chen. Dynamics of hspc repopulation in nonhuman primates revealed by a decade-long clonal-tracking study. Cell stem cell, 14:473–85, 4 2014. ISSN 1875-9777. doi: 10.1016/j.stem.2013.12.012. Cited 2 times on pages 83 and 102.
- Allon M. Klein and Benjamin D. Simons. Universal patterns of stem cell fate in cycling adult tissues. Development, 138:3103–3111, 8 2011. ISSN 1477-9129. doi: 10.1242/dev.060103. Cited on page 100.
- Henry Lee-Six, Nina Friesgaard Øbro, Mairi S Shepherd, Sebastian Grossmann, Kevin Dawson, Miriam Belmonte, Robert J Osborne, Brian J P Huntly, Inigo Martincorena, Elizabeth Anderson, Laura O'Neill, Michael R Stratton, Elisa Laurenti, Anthony R Green, David G Kent, and Peter J Campbell. Population dynamics of normal human blood inferred from somatic mutations. Nature, 561:473–478, 9 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0497-0. Cited on page 82.
- Markus Loeffler and Ingo Roeder. Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models—a conceptual approach. Cells, tissues, organs, 171:8–26, 2002. ISSN 1422-6405. doi: 10.1159/000057688. Cited on page 116.

A. J. Lotka. Elements of Physical Biology. Williams and Wilkins, 1925. Cited on page 13.

Rong Lu, Norma F Neff, Stephen R Quake, and Irving L Weissman. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nature biotechnology, 29:928–33, 10 2011. ISSN 1546-1696. doi: 10.1038/nbt.1977. Cited 2 times on pages 82 and 102.

Emily Mitchell, Michael Spencer Chapman, Nicholas Williams, Kevin J Dawson, Nicole Mende, Emily F Calderbank, Hyunchul Jung, Thomas Mitchell, Tim H H Coorens, David H Spencer, Heather Machado, Henry Lee-Six, Megan Davies, Daniel Hayler, Margarete A Fabre, Krishna Mahbubani, Federico Abascal, Alex Cagan, George S Vassiliou, Joanna Baxter, Inigo Martincorena, Michael R Stratton, David G Kent, Krishna Chatterjee, Kourosh Saeb Parsy, Anthony R Green, Jyoti Nangalia, Elisa Laurenti, and Peter J Campbell. Clonal dynamics of haematopoiesis across the human lifespan. Nature, 606:343–350, 6 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04786-y. Cited 2 times on pages 83 and 106.

P. A.P. Moran. Random processes in genetics. Mathematical Proceedings of the Cambridge Philosophical Society, 54, 1958. ISSN 14698064. doi: 10.1017/S0305004100033193. Cited 3 times on pages 13, 41, and 52.

Mina N F Morcos, Congxin Li, Clara M Munz, Alessandro Greco, Nicole Dressel, Susanne Reinhardt, Katrin Sameith, Andreas Dahl, Nils B Becker, Axel Roers, Thomas Höfer, and Alexander Gerbaulet. Fate mapping of hematopoietic stem cells reveals two pathways of native thrombopoiesis. Nature communications, 13:4504, 8 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-31914-z. Cited 2 times on pages 83 and 116.

S J Morrison, N Uchida, and I L Weissman. The biology of hematopoietic stem cells. Annual review of cell and developmental biology, 11:35–71, 1995. ISSN 1081-0706. doi: 10.1146/annurev.cb.11.110195.000343. Cited 2 times on pages 15 and 82.

J D Murray. Mathematical Biology : I . An Introduction , Third Edition, volume 1. Springer New York, NY, 2002. doi: 10.1086/421587. Cited 4 times on pages 25, 27, 52, and 96.

Armando G. M. Neves. Aplicações biológicas de cadeias de markov. Notas de minicurso ministrado no III Colóquio de Matemática da Região Norte, 2014. Cited 2 times on pages 23 and 24.

- Robert J. Noble, Viola Walther, Christian Roumestand, Michael E. Hochberg, Urszula Hibner, and Patrice Lassus. Paracrine behaviors arbitrate parasite-like interactions between tumor subclones. Frontiers in Ecology and Evolution, 9, 2021. doi: 10.3389/fevo.2021.675638. Cited on page 80.
- Martin A. Nowak. Evolutionary dynamics: exploring the equations of life. Harvard University Press, 2006. Cited 3 times on pages 33, 41, and 52.
- J. Park and P. K. Newton. Stochastic competitive release and adaptive chemotherapy. Physical Review E, 108, 2023. ISSN 24700053. doi: 10.1103/PhysRevE.108.034407. Cited 4 times on pages 52, 53, 61, and 79.
- Rasmus Kristoffer Pedersen, Morten Andersen, Thomas Stiehl, and Johnny T. Ottesen. Mathematical modelling of the hematopoietic stem cell-niche system: Clonal dominance based on stem cell fitness. Journal of Theoretical Biology, 518:110620, 6 2021. ISSN 00225193. doi: 10.1016/j.jtbi.2021.110620. Cited on page 82.
- Rasmus Kristoffer Pedersen, Morten Andersen, Vibe Skov, Lasse Kjær, Hans C Hasselbalch, Johnny T Ottesen, and Thomas Stiehl. Hsc niche dynamics in regeneration, pre-malignancy, and cancer: Insights from mathematical modeling. Stem cells (Dayton, Ohio), 41:260–270, 3 2023. ISSN 1549-4918. doi: 10.1093/stmcls/sxac079. Cited 2 times on pages 113 and 114.
- Weike Pei, Thorsten B Feyerabend, Jens Rössler, Xi Wang, Daniel Postrach, Katrin Busch, Immanuel Rode, Kay Klapproth, Nikolaus Dietlein, Claudia Quedenau, Wei Chen, Sascha Sauer, Stephan Wolf, Thomas Höfer, and Hans-Reimer Rodewald. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. Nature, 548:456–460, 8 2017. ISSN 1476-4687. doi: 10.1038/nature23653. Cited on page 82.
- Ana Luisa Pereira, Serena Galli, and Cesar Nombela-Arrieta. Bone marrow niches for hematopoietic stem cells. HemaSphere, 8, 8 2024. ISSN 2572-9241. doi: 10.1002/hem3.133. Cited 2 times on pages 15 and 82.
- Leïla Perié, Philip D Hodgkin, Shalin H Naik, Ton N Schumacher, Rob J de Boer, and Ken R Duffy. Determining lineage pathways from cellular barcoding experiments. Cell reports, 6: 617–24, 2 2014. ISSN 2211-1247. doi: 10.1016/j.celrep.2014.01.016. Cited on page 82.

- Lawrence Perko. Differential equations and dynamical systems. Springer-Verlag, Berlin, Heidelberg, 1991. ISBN 0387974431. Cited 2 times on pages 21 and 90.
- Sandra Pinho and Paul S Frenette. Haematopoietic stem cell activity and interactions with the niche. Nature reviews. Molecular cell biology, 20:303–320, 5 2019. ISSN 1471-0080. doi: 10.1038/s41580-019-0103-9. Cited 2 times on pages 15 and 82.
- Rosângela A. Pires. Análise assintótica dos tempos médios de fixação no Processo de Moran. PhD thesis, Universidade Federal de Minas Gerais, 2022. URL <http://www.mat.ufmg.br/posgrad/wp-content/uploads/TesesDissertacoes/Tese180.pdf>. Cited 2 times on pages 48 and 77.
- Stefan Radtke, Jennifer E Adair, Morgan A Giese, Yan-Yi Chan, Zachary K Norgaard, Mark Enstrom, Kevin G Haworth, Lauren E Schefter, and Hans-Peter Kiem. A distinct hematopoietic stem cell population for rapid multilineage engraftment in nonhuman primates. Science translational medicine, 9, 11 2017. ISSN 1946-6242. doi: 10.1126/scitranslmed.aan1145. Cited 4 times on pages 83, 102, 105, and 109.
- Stefan Radtke, Mark Enstrom, Dnyanada Pande, Elizabeth R Duke, E Fabian Cardozo-Ojeda, Ravishankar Madhu, Staci Owen, Greta Kanestrom, Margaret Cui, Anai M Perez, Joshua T Schiffer, and Hans-Peter Kiem. Stochastic fate decisions of hscs after transplantation: early contribution, symmetric expansion, and pool formation. Blood, 142:33–43, 7 2023. ISSN 1528-0020. doi: 10.1182/blood.2022018564. Cited 6 times on pages 83, 102, 105, 111, 119, and 121.
- Ingo Roeder, Leonie M Kamminga, Katrin Braesel, Bert Dontje, Gerald de Haan, and Markus Loeffler. Competitive clonal hematopoiesis in mouse chimeras explained by a stochastic model of stem cell organization. Blood, 105:609–16, 1 2005. ISSN 0006-4971. doi: 10.1182/blood-2004-01-0282. Cited on page 84.
- Gideon Schwarz. Estimating the dimension of a model. The Annals of Statistics, 6(2):461–464, 1978. doi: 10.1214/aos/1176344136. Cited on page 107.
- Luca Scrucca. Ga : A package for genetic algorithms in r. Journal of Statistical Software, 53, 2013. ISSN 1548-7660. doi: 10.18637/jss.v053.i04. Cited on page 107.

- Jianlong Sun, Azucena Ramos, Brad Chapman, Jonathan B Johnnidis, Linda Le, Yu-Jui Ho, Allon Klein, Oliver Hofmann, and Fernando D Camargo. Clonal dynamics of native haematopoiesis. Nature, 514:322–7, 10 2014. ISSN 1476-4687. doi: 10.1038/nature13824. Cited on page 82.
- Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71, 2021. ISSN 0007-9235. doi: 10.3322/caac.21660. Cited on page 50.
- Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. Mathematical Biosciences, 40, 1978. ISSN 00255564. doi: 10.1016/0025-5564(78)90077-9. Cited 3 times on pages 13, 33, and 52.
- Arne Traulsen, Jens Christian Claussen, and Christoph Hauert. Coevolutionary dynamics in large, but finite populations. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 74, 2006. ISSN 15393755. doi: 10.1103/PhysRevE.74.011901. Cited 2 times on pages 36 and 43.
- P. Venkatesh and A. Kasi. Anthracyclines. In StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL), 2025. URL <https://www.ncbi.nlm.nih.gov/books/NBK538187/>. [Updated 2023 Jan 30; cited 2025 Jan-]. Cited on page 113.
- Evgenia Verovskaya, Mathilde J C Broekhuis, Erik Zwart, Martha Ritsema, Ronald van Os, Gerald de Haan, and Leonid V Bystriykh. Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood, 122: 523–32, 7 2013. ISSN 1528-0020. doi: 10.1182/blood-2013-01-481135. Cited on page 82.
- Andrea Viale and Pier G. Pelicci. Awakening stem cells from dormancy: growing old and fighting cancer. EMBO Molecular Medicine, 1:88–91, 5 2009. ISSN 1757-4676. doi: 10.1002/emmm.200900019. Cited on page 116.
- Vito Volterra. Variazioni e fluttuazioni del numero d’individui in specie animali conviventi. Memorie della Reale Accademia Nazionale dei Lincei, 2:31–113, 1926. Cited on page 13.
- Dujuan Wang and Jinzhi Lei. Optimal adaptive therapeutic schedules for metastatic castrate-resistant prostate cancer based on bilevel optimization problem. J. Math. Biol., 90, 2025. doi: 10.1007/s00285-025-02220-1. Cited on page 52.

- Jiali Wang, Yixuan Zhang, Xiaoquan Liu, and Haochen Liu. Optimizing adaptive therapy based on the reachability to tumor resistant subpopulation. *Cancers*, 13, 2021. ISSN 20726694. doi: 10.3390/cancers13215262. Cited 3 times on pages 14, 50, and 78.
- Jeffrey West, Yongqian Ma, and Paul K. Newton. Capitalizing on competition: An evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment. *Journal of Theoretical Biology*, 455, 2018. ISSN 10958541. doi: 10.1016/j.jtbi.2018.07.028. Cited 3 times on pages 14, 50, and 78.
- Jeffrey West, Li You, Jingsong Zhang, Robert A. Gatenby, Joel S. Brown, Paul K. Newton, and Alexander R.A. Anderson. Towards multidrug adaptive therapy. *Cancer Research*, 80, 2020. ISSN 15387445. doi: 10.1158/0008-5472.CAN-19-2669. Cited 3 times on pages 14, 51, and 78.
- Jeffrey West, Fred Adler, Jill Gallaher, Maximilian Strobl, Renee Brady-Nicholls, Joel Brown, Mark Roberson-Tessi, Eunjung Kim, Robert Noble, Yannick Viossat, David Basanta, and Alexander R.A. Anderson. A survey of open questions in adaptive therapy: Bridging mathematics and clinical translation. *eLife*, 12, 2023. ISSN 2050084X. doi: 10.7554/elife.84263. Cited 3 times on pages 14, 50, and 78.
- Jeffrey B. West, Mina N. Dinh, Joel S. Brown, Jingsong Zhang, Alexander R. Anderson, and Robert A. Gatenby. Multidrug cancer therapy in metastatic castrate-resistant prostate cancer: An evolution-based strategy. *Clinical Cancer Research*, 25, 2019. ISSN 15573265. doi: 10.1158/1078-0432.CCR-19-0006. Cited 3 times on pages 14, 51, and 78.
- Anne Wilson, Elisa Laurenti, Gabriela Oser, Richard C van der Wath, William Blanco-Bose, Maike Jaworski, Sandra Offner, Cyrille F Dunant, Leonid Eshkind, Ernesto Bockamp, Pietro Lió, H Robson Macdonald, and Andreas Trumpp. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, 135:1118–29, 12 2008. ISSN 1097-4172. doi: 10.1016/j.cell.2008.10.048. Cited 4 times on pages 82, 83, 116, and 117.
- Yonghui Xia and Maoan Han. New conditions on the existence and stability of periodic solution in lotka-volterra's population system. *SIAM Journal on Applied Mathematics*, 69, 2009. ISSN 00361399. doi: 10.1137/070702485. Cited on page 67.

Jingsong Zhang, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nature Communications, 8, 2017. ISSN 20411723. doi: 10.1038/s41467-017-01968-5. Cited 4 times on pages 14, 50, 67, and 78.

Lei Zhang, Jianli Ma, Lei Liu, Guozheng Li, Hui Li, Yi Hao, Xin Zhang, Xin Ma, Yihai Chen, Jiale Wu, Xinheng Wang, Shuai Yang, and Shouping Xu. Adaptive therapy: a tumor therapy strategy based on darwinian evolution theory. Critical Reviews in Oncology/Hematology, 192, 2023. ISSN 18790461. doi: 10.1016/j.critrevonc.2023.104192. Cited 3 times on pages 14, 50, and 78.