

Article

Reinforcement Learning-Driven Evolutionary Stackelberg Game Model for Adaptive Breast Cancer Therapy

Fatemeh Tavakoli ¹, Davud Mohammadpur ¹, Javad Salimi Sartakhti ² and Mohammad Hossein Manshaei ^{3,*}

¹ Department of Electrical and Computer Engineering, University of Zanjan, Zanjan 45371-38791, Iran; dmp@znu.ac.ir (D.M.)

² Department of Electrical and Computer Engineering, University of Kashan, Kashan 87317-53153, Iran; salimi@kashanu.ac.ir

³ Department of Computer Science, Hunter College, City University of New York, New York, NY 10065, USA

* Correspondence: mohammad.manshaei@hunter.cuny.edu

Abstract

In this paper, we present an integrative framework based on Evolutionary Stackelberg Game Theory to model the strategic interaction between a physician, acting as a rational leader, and a heterogeneous population of treatment-sensitive and treatment-resistant breast cancer cells. The model incorporates ecological competition, evolutionary adaptation, and spatial heterogeneity, enabling prediction of tumor progression under clinically relevant treatment protocols. Using tumor volume data obtained from breast cancer-bearing mice treated with Capecitabine and Gemcitabine, we estimated treatment and subject-specific parameters via the GEKKO optimization package in Python. Benchmarking against classical tumor growth models (Exponential, Logistic, and Gompertz) showed that while classical models capture monotonic growth, they fail to reproduce complex, non-monotonic behaviors such as treatment-induced regression, rebound, and phenotypic switching. The game-theoretic approach achieved superior alignment with experimental data across Maximum Tolerated Dose, Dose-Modulation Adaptive Therapy, and Intermittent Adaptive Therapy protocols. To enhance adaptability, we integrated reinforcement learning (RL) for both single-agent and combination chemotherapy. The RL agent learned dosing policies that maximized tumor regression while minimizing cumulative drug exposure and resistance, with combination therapy exploiting dose diversification to improve control without exceeding total dose budgets. Incorporating reaction diffusion equations allowed the model to capture spatial dispersal of sensitive (cooperative) and resistant (defector) phenotypes, revealing that spatially aware adaptive strategies more effectively suppress resistant clones than non-spatial approaches. These results demonstrate that evolutionarily informed, spatially explicit, and computationally optimized strategies can outperform conventional fixed-dose regimens in reducing resistance, lowering toxicity, and improving efficacy. This framework offers a biologically interpretable tool for guiding evolution-aware, patient-tailored cancer therapies toward improved long-term outcomes.

Keywords: breast cancer modeling; Stackelberg game theory; adaptive therapy; phenotypic dynamics; diffusion; treatment-sensitive cells; treatment-resistant cells



Received: 12 October 2025
Revised: 19 November 2025
Accepted: 25 November 2025
Published: 5 December 2025

Citation: Tavakoli, F.; Mohammadpur, D.; Sartakhti, J.S.; Manshaei, M.H. Reinforcement Learning-Driven Evolutionary Stackelberg Game Model for Adaptive Breast Cancer Therapy. *Math. Comput. Appl.* **2025**, *30*, 134. <https://doi.org/10.3390/mca30060134>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer is a biologically complex and heterogeneous disease, often leading to drug resistance and treatment failure. Traditional approaches such as the maximum tolerated dose (MTD) aim for aggressive tumor eradication but frequently accelerate the

emergence of resistant clones [1,2]. Adaptive therapy offers a promising alternative by dynamically adjusting drug dosage to maintain a manageable tumor burden and suppress resistance through intratumoral competition [3,4].

To design such evolution-informed strategies, computational modeling plays a crucial role. Frameworks including cellular automata, differential equations, agent-based simulations, and especially game-theoretic models [5–8] enable researchers to simulate tumor dynamics and optimize treatment protocols. Among these, evolutionary game theory provides biologically realistic insights into strategic interactions among cancer cells [9,10]. Recent models based on Stackelberg evolutionary games conceptualize the physician as a rational leader who anticipates the adaptive responses of cancer cells [11,12].

Building on this foundation, our study introduces the Breast Cancer Stackelberg game model to analyze the strategic interactions between a physician and a heterogeneous tumor cell population. The schematic representation of this hierarchical leader–follower interaction framework is presented in Figure 1, highlighting the physician’s role as a rational leader who anticipates and influences the evolutionary responses of cancer cells.

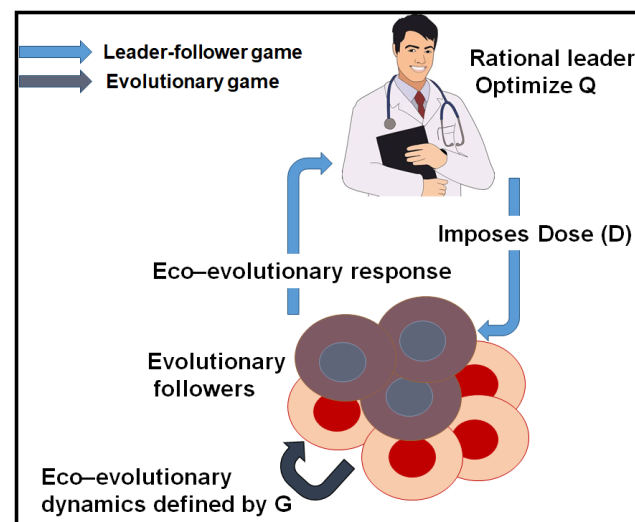


Figure 1. The structure of evolutionary Stackelberg game in cancer.

To complement our theoretical framework, we utilized experimental data from xenograft mice treated with Capecitabine and Gemcitabine, as reported in [13]. Moving beyond purely theoretical approaches, we incorporate patient-specific tumor measurements to estimate key biological parameters through the Python-based GEKKO optimization package (version 1.3.0), enhancing model realism and predictive accuracy. The resulting model closely reproduces observed tumor behavior under three therapeutic protocols: dose-modulated adaptive therapy, intermittent adaptive therapy, and MTD regimens.

To capture the strategic interplay between treatment-sensitive and -resistant cancer cell populations, we formulate their interactions as a Stackelberg evolutionary game. In this hierarchical framework, the physician acting as a rational leader selects treatment doses to optimize a predefined objective function linked to patient quality of life. We evaluate three therapeutic strategies: MTD, ecologically enlightened therapy (EET), and evolutionarily enlightened therapy (EVT), following the conceptual framework introduced by Stein et al. [14]. Our analysis, applied to our own dataset, demonstrates that EVT consistently achieves the best outcomes: minimizing drug usage, reducing resistance emergence, and maximizing the objective function. In contrast, the MTD strategy accelerates resistance and compromises treatment durability.

In EET, the physician dynamically adjusts dosing based on observed resistance levels, leading to a static Nash equilibrium, where both the cancer cells’ resistance and the

physician's dosage mutually stabilize in response to one another. However, in EVT, the physician anticipates the future adaptive responses of cancer cells, achieving a Stackelberg equilibrium in which resistant populations settle into a stable eco-evolutionary state. This foresight allows the physician to proactively steer tumor evolution toward more manageable trajectories.

Further, to address spatial heterogeneity in the tumor microenvironment, we integrate reaction-diffusion equations into the model. This extension captures phenotypic diffusion and spatial expansion of resistant clones, underscoring the critical role of spatial dynamics in shaping treatment outcomes and resistance patterns.

Building on this calibrated framework, we integrate reinforcement learning (RL) to design adaptive treatment strategies for both single-agent and dual-drug settings. In the single-agent case, the RL agent learns dosing policies that balance tumor regression against cumulative drug exposure and resistance suppression. For combination therapy, the agent dynamically allocates doses between Capecitabine and Gemcitabine within pharmacodynamic and toxicity constraints, exploiting dose diversification to improve tumor control without exceeding total dose budgets. In both scenarios, the RL-driven policies adapt to evolving tumor states, offering a flexible and data-informed approach to therapy optimization.

By leveraging limited early-stage tumor data, our approach enables reliable forecasts of tumor evolution and facilitates comparative evaluation of alternative treatment strategies, including their long-term efficacy and resistance outcomes. This predictive capability suggests that the framework may serve as a useful decision-support tool for designing personalized, evolution-aware cancer therapies that combine biologically grounded modeling with adaptive, AI-driven treatment planning.

The remainder of this paper is organized as follows. Section 2 presents a Stackelberg-based evolutionary modeling framework that strategically captures tumor progression by explicitly representing the interactions among heterogeneous breast cancer cell populations. Section 3 presents the main findings, including parameter estimation from experimental data, analysis of sensitive and resistant cell dynamics, spatial modeling of phenotype dispersal within the tumor microenvironment, and reinforcement learning-based optimization of adaptive therapy (single-drug and dual-drug). Finally, Section 4 discusses the broader implications of our findings, highlighting how the integration of game theory, reinforcement learning, and spatial ecology can inform the design of evolutionarily informed, resistance-aware treatment protocols in precision oncology.

2. Stackelberg Game Framework for Breast Cancer Therapy

The Stackelberg-Based Evolutionary Model provides a strategic framework for analyzing tumor progression by incorporating hierarchical interactions between breast cancer cells. By leveraging game theory principles, this model optimizes treatment strategies, balancing the competition between sensitive and resistant cell populations to enhance therapeutic efficacy.

This model enables physicians to make informed treatment decisions by dynamically adjusting drug dosages based on tumor response. Through its predictive capabilities, it offers personalized approaches that mitigate resistance development, ultimately improving patient outcomes in breast cancer therapy.

In this study, we propose an evolutionary Stackelberg game theory model to describe interactions among breast cancer cells. To evaluate the model's applicability to physician-tumor dynamics, we analyzed experimental data from mice treated with two distinct therapeutic agents: Capecitabine and Gemcitabine. Model parameters were estimated

based on observed data for each protocol, enabling a comparative evaluation of drug dosage effects and tumor burden variations.

The model was subsequently extended to incorporate two cancer cell populations, treatment-sensitive and treatment-resistant cells, allowing for a more comprehensive analysis of their spatial diffusion dynamics within a two-dimensional framework. This extension provides deeper insights into tumor evolution and treatment optimization, enhancing the predictive capabilities of the proposed model in guiding therapeutic decision-making.

Evolutionary Stackelberg game theory offers a promising framework for optimizing cancer treatment strategies. Within this approach, breast cancer therapy is modeled as an interaction between the physician and a heterogeneous population of cancer cells. To capture the evolutionary dynamics of tumor growth and phenotypic adaptations leading to resistance over time, a fitness generating function (G-function) is employed. This function accounts for logistic tumor growth, which is effectively suppressed by multiple anticancer drugs. The generating function is defined as

$$G(u, x, D) = r \left[1 - \frac{x}{K(u)} \right] - \frac{D}{k + bu}. \quad (1)$$

r indicates intrinsic growth rate and x is normalized tumor burden. The expression $\frac{D}{k+bu}$ represents the drug-induced death rate, where D denotes the administered drug dose, k represents the intrinsic resistance of cancer cells, and b quantifies the impact of resistance traits on diminishing drug efficacy. In this framework, D takes values between 0 and 1, with $D = 1$ corresponding to MTD. The carrying capacity, denoted as $K(u)$, is modeled as a function of the resistance level u , reflecting the trade-off between resistance and ecological fitness. This formulation follows previous work by Pressley et al. [15], where similar relationships were used to capture treatment-induced evolutionary dynamics:

$$K(u) = K_{\max} \cdot \exp(-g \cdot u). \quad (2)$$

Here, $K_{\max}$ represents the maximum possible carrying capacity, and g determines the magnitude of the resistance-associated cost. This formulation accounts for the impact of resistance on population growth, ensuring that as resistance levels increase, the effective carrying capacity is reduced. The ecological dynamics governing the system are defined as follows:

$$\frac{dx}{dt} = xG(u, x, D). \quad (3)$$

The evolution of cancer cell resistance in response to the physician's treatment choices is governed by specific biological and ecological dynamics. These dynamics define how resistance levels change over time. The formal mathematical representation of these resistance dynamics is defined as follows:

$$\frac{du}{dt} = \sigma \frac{\partial G}{\partial u}, \quad (4)$$

The parameter σ determines the evolutionary speed of resistance adaptation. The following functions describe the calculation of the D value, adjusted for different treatment protocols.

Standard Treatment Protocol: The standard treatment protocol involves the continuous administration of the drug at MTD, ensuring a consistent therapeutic effect. This approach maintains a high level of drug exposure to maximize tumor suppression, although it may lead to increased side effects and potential resistance over time.

Dose Modulation Adaptive Therapy: In this adaptive therapy protocol, drug dosage is dynamically adjusted based on tumor burden fluctuations. If the tumor burden decreases by more than 10% compared to the previous measurement, the dose is reduced by 50%, preventing unnecessary drug exposure. Conversely, if the tumor burden increases by more

than 10%, the dose is increased by 50%, ensuring sufficient drug pressure to counteract tumor progression.

Additionally, treatment is suspended when the tumor burden falls below a predefined threshold, allowing for drug holidays to minimize toxicity and prevent unnecessary intervention. Therapy is resumed once tumor growth surpasses the threshold again, ensuring effective control of disease progression. The parameters governing this adaptive strategy include $\alpha = 0.5$ and $\beta = 0.1$, which regulate the degree of dose modulation. The dose adjustment function for this protocol is defined as

$$D(x_i) = \begin{cases} (1 + \alpha)D_{i-1} & \text{if } x_i > (1 + \beta)x_{i-1}, \\ (1 - \alpha)D_{i-1} & \text{if } x_i \leq (1 - \beta)x_{i-1}, \\ D_0 & \text{otherwise.} \end{cases} \quad (5)$$

Here, D_0 represents the initial dose at the start of treatment, D_{i-1} is the drug dose in phase $(i - 1)$, and i indicates the current stage. The variable x_i denotes the tumor burden at time step i , while x_{i-1} corresponds to the burden measured in the previous interval. The time steps i and $i - 1$ represent consecutive measurement intervals, each corresponding to one unit of time based on the experimental sampling schedule.

Intermittent Adaptive Therapy Protocol: This protocol initiates treatment by administering the drug at MTD. If the tumor burden decreases to less than 50% of its initial value, treatment is temporarily discontinued to minimize unnecessary drug exposure and reduce toxicity. However, if the tumor burden exceeds its original load, therapy is resumed at the MTD for each drug. The function D for this protocol is defined as

$$D(x_i) = \begin{cases} 0 & \text{if } x_i < 0.5x_0, \\ 1 & \text{if } x_i > x_0, \\ D_0 & \text{otherwise.} \end{cases} \quad (6)$$

Here, $D_0 = \text{MTD}$ denotes the maximum tolerated dose. This innovative approach provides a comprehensive depiction of tumor population dynamics and the evolutionary trajectories of cellular resistance mechanisms in response to therapeutic interventions. By employing this modeling technique, evolutionary game theory offers critical insights into the complex interactions between therapeutic strategies and tumor evolution. These findings contribute to the development of personalized treatment protocols, enabling the optimization of therapeutic outcomes and improving long-term cancer management.

3. Results

This section is organized into six interconnected parts, each addressing a key component of the study. First, in Estimation of Model Parameters (Section 3.1), we present the methodology for deriving both treatment-specific and patient-specific parameters from experimental tumor volume data, ensuring accurate calibration of the model to individual and regimen-level characteristics. Second, in Sensitive and Resistant Cell Interactions (Section 3.2), we investigate the ecological and evolutionary dynamics between treatment-sensitive and treatment-resistant phenotypes, highlighting their competitive interplay under different therapeutic protocols. Third, in Integration of Experimental Data into the Simulation Framework (Section 3.4), we explain how real-world data is incorporated into the simulation environment and how it informs model behavior and predictions. Fourth, in Diffusion in Sensitive and Resistant Cells (Section 3.3), we incorporate spatial heterogeneity through reaction-diffusion modeling to capture the dispersal patterns of both phenotypes

and assess their impact on tumor progression and treatment response. Fifth, in Application of Reinforcement Learning for Adaptive Cancer Therapy (Section 3.5), we evaluate the performance of an RL-based framework in learning dosing strategies that balance tumor regression, cumulative drug exposure, and resistance suppression. Finally, in Extension from Single-Agent to Combination Therapy Modeling (Section 3.6), we expand the RL approach to dual-drug regimens, analyzing dose allocation under pharmacodynamic and toxicity constraints. This structured presentation enables a comprehensive evaluation of the model's predictive capabilities and its potential to inform the design of evolutionarily informed, resistance-aware cancer therapies.

3.1. Estimation of Model Parameters

To evaluate the proposed model for cancer cell interactions with the physician, experimental data from treated mice were analyzed. Initially, the dataset underwent MinMax scaling normalization to standardize values, ensuring uniformity and consistency in the analytical process. This preprocessing step allowed for a more accurate comparison of treatment effects and model predictions. Tumors were classified into two distinct categories: increasing (Up) and decreasing (Down) based on their initial response to treatment. This classification criterion is based on the tumor's early trend upon treatment initiation, distinguishing between cases where the tumor shows immediate growth versus regression. Clinically, patients are categorized as non-responders if their tumor burden increases initially, while those showing early tumor shrinkage are labeled as responders. Notably, this initial reaction to treatment may provide valuable predictive insights into future tumor progression, making it a meaningful clustering approach.

Parameter estimation was conducted using the Python package GEKKO, a specialized tool for machine learning and optimization tasks. The model incorporates both treatment-specific and patient-specific parameters to enhance predictive accuracy. The treatment-specific parameters $g, k, b, K_{\max}$ define therapeutic response mechanisms. These parameters are estimated independently for each therapy, while the patient-specific parameters σ, r, u_0 account for individual variations in disease progression and treatment sensitivity.

The distinction between treatment-specific and patient-specific parameters is grounded in both biological reasoning and modeling precedent. In our framework, parameters such as $K_{\max}$ which denotes the effective carrying capacity of the tumor under therapy are modeled as treatment-specific because they reflect the pharmacodynamic impact of the administered drug on tumor growth constraints. Prior studies have shown that therapeutic agents can modulate the tumor microenvironment, vascularization, and immune infiltration, all of which influence the maximum sustainable tumor burden under treatment [3,16]. In contrast, parameters like intrinsic growth rate (r) or resistance kinetics (σ, u_0) are considered patient-specific, as they are more tightly linked to host biology, genetic background, and tumor phenotype [17,18].

The variables and parameters of the model are summarized in Table 1. The optimal parameters for each group were selected based on their ability to achieve the best fit with observed data. A summary of their values is provided in Table 2, highlighting key parameter estimates.

Table 1. Variables and parameters of the model. Parameter ranges were informed by prior literature, particularly the study by Benzekry et al. on resistance modeling in non-small cell lung cancer [19], which provides biologically plausible bounds for resistance-related dynamics.

Parameter	Description	Range
r	Intrinsic growth rate	$[0, 1]$
σ	Evolution speed	$(0, 0.1]$
u_0	Initial resistance to treatment	Non-negative
k	Intrinsic resistance	$[0, 20]$
$K_{\max}$	Maximum carrying capacity	Non-negative
b	Benefit of resistance traits reducing drug efficacy	$[0, 20]$
g	Cost of resistance	$[0, 1]$
x	Normalized tumor burden (scaled)	$[0, 1]$
D	Drug dosage amount	$[0, 1]$
u	Resistance intensity to treatment	Non-negative

Table 2. The model parameters have been estimated.

Therapy	Increasing				Decreasing				Mouse
	b	k	$K_{\max}$	g	b	k	$K_{\max}$	g	
Capecitabine Therapy									
MTD	1.0	10.33	1.0	0.04	–	–	–	–	19,20,21,22,23,24
Dose modulation	–	–	–	–	19.92	1.02	1.0	0.99	31,34
Intermittent	1.0	9.98	1.0	0.18	1.0	19.99	1.0	0.83	25,26,27,29,28,30
Gemcitabine Therapy									
MTD	1.0	5.53	1.0	0.1	19.99	19.99	1.0	1.0	1,2,4,3,5,6
Dose modulation	1.0	7	1.0	0.05	19.93	18	1.0	0.89	13,15,16,17,18
Intermittent	1.0	6.55	1.0	1.0	1.0	3.38	1.0	0.13	7,8,9,10,11,12

Note: Parameter values were estimated using the GEKKO optimization package. The variation in parameter b reflects differences in tumor response dynamics across treatment protocols, as determined by nonlinear least-squares fitting to experimental data.

The parameters $b, k, K_{\max}$ and g were estimated under the assumption that they are treatment-specific but remain constant across all patients receiving the same therapeutic regimen. In contrast, the parameters σ, r , and the initial value of u were considered patient-specific, as summarized in Table 1. These subject-specific parameters were estimated independently for each tumor, with the primary objective of minimizing the discrepancy between observed experimental data and the predictions generated by the model.

Figure 2 illustrate actual tumor growth data and corresponding model-generated graphs based on the estimated parameters for a subset of mice, providing a visual representation of the model's predictive accuracy and its ability to capture tumor dynamics. As illustrated in the figure, the graph representing the actual data for Mice 2, 5, 12, 22, 25, and 31 closely aligns with the graph generated using the estimated parameters. This similarity indicates the accuracy and reliability of the parameter estimation process. The specific parameters associated with each mouse (u_0, σ, r) are detailed in the figure caption, providing insights into individual variations in tumor progression and response to treatment. Such consistency validates the effectiveness of the modeling approach in capturing key biological dynamics.

Since we have assumed that a specific set of parameters namely $K_{\max}, b, g$ and k are treatment-specific, we can utilize them to simulate potential outcomes for patients who received different therapeutic protocols. This simulation enables comparative analysis of treatment efficacy, providing insights into how variations in drug administration strategies may have influenced tumor progression and patient responses.

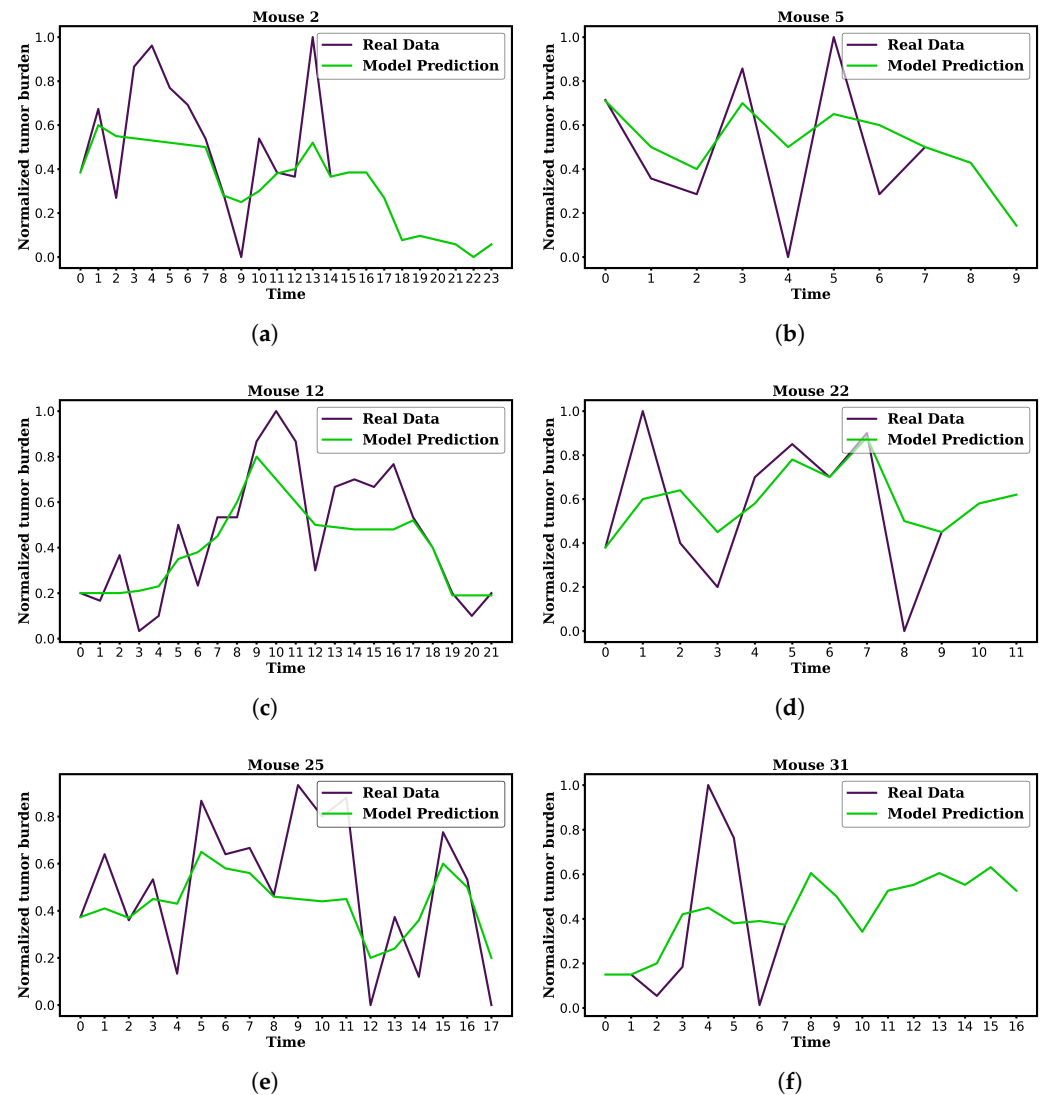


Figure 2. Comparison between experimental data and simulated population dynamics for Mice 2, 5, 12, 22, 25, and 31 using subject-specific estimated parameters. (a) Mouse 2: $r = 0.20$, $u_0 = 1.5478$, $\sigma = 0.0499$; (b) Mouse 5: $r = 0.2$, $u_0 = 10.0$, $\sigma = 0.05$; (c) Mouse 12: $r = 0.1999$, $u_0 = 1.1651$, $\sigma = 0.05$; (d) Mouse 22: $r = 0.2001$, $u_0 = 9.5948$, $\sigma = 0.0499$; (e) Mouse 25: $r = 0.1907$, $u_0 = 0.5102$, $\sigma = 0.0499$; (f) Mouse 31: $r = 0.2039$, $u_0 = 3.1046$, $\sigma = 0.05$.

While the model generally captures the overall trends in tumor growth and treatment response, its predictions do not fully align with experimental data in certain cases—particularly for subjects exhibiting oscillatory or non-monotonic behavior. These discrepancies likely stem from the simplified structure of the model, which does not explicitly incorporate key biological mechanisms such as immune system interactions, microenvironmental influences, and stochastic variability. Moreover, the parameter estimation process was intentionally constrained to biologically plausible ranges to prevent overfitting and ensure interpretability, which may have further limited the model’s flexibility in reproducing complex dynamics. Despite these limitations, the framework remains effective in capturing key features of tumor ecology and provides a foundation for evaluating adaptive treatment strategies.

In order to demonstrate the advantages of our game-theoretic modeling framework, we first compared its performance against well-established classical tumor growth models that have long served as the foundation of mathematical oncology. Classical models such as the Exponential, Logistic, and Gompertz formulations provide simplified yet analytically

tractable descriptions of tumor expansion, each grounded in distinct biological assumptions. The Exponential model presumes unrestricted proliferation, making it suitable for early-stage growth but unrealistic for long-term dynamics due to the absence of resource limitations. The Logistic model introduces a fixed carrying capacity, thereby capturing saturation effects as tumor burden approaches physiological constraints. The Gompertz model, in turn, characterizes a gradual deceleration in growth rate, consistent with empirical observations of reduced proliferation in larger tumors. These paradigms have been extensively studied and benchmarked in both preclinical and clinical contexts, with recent systematic analyses confirming their utility and limitations [20,21].

Following the methodology outlined by Ghaffari Laleh et al. [20], we performed a comparative analysis using a dataset derived from murine tumor volume measurements across various tumor growth models. Tumor burden trajectories were fitted to each of the three classical models, and predictive accuracy was quantified using the Mean Squared Error (MSE) between model-predicted and experimentally measured tumor data. This benchmarking step provided a rigorous reference point for evaluating our game-theoretic framework, which explicitly incorporates ecological and evolutionary interactions between phenotypically distinct subpopulations within the tumor microenvironment. Unlike classical models that treat tumor growth as a homogeneous process, the game-theoretic approach accounts for competition, cooperation, and treatment feedback, thereby enabling the modeling of adaptive responses and evolutionary dynamics. Model parameters were estimated using the GEKKO optimization suite [22], which supports dynamic systems and differential-algebraic equations.

The comparative results revealed that while classical models can adequately reproduce monotonic growth patterns, they consistently fail to capture non-monotonic behaviors such as treatment-induced regressions, rebounds, and phenotypic switching phenomena increasingly recognized in both experimental and clinical datasets. In contrast, the game-theoretic model achieved superior alignment with these complex dynamics, underscoring its potential for more biologically grounded and clinically informative predictions. A summary of MSE values across different therapy strategies is presented in Table 3, highlighting the enhanced predictive accuracy of the game-theoretic framework, particularly in scenarios involving adaptive treatment.

An important advantage of the proposed modeling framework is its ability to incorporate patient- or therapy-specific parameter estimates, enabling the simulation and comparison of multiple therapeutic protocols for the same individual. By calibrating the model to experimental data, it becomes possible to explore how alternative treatment strategies would influence both tumor dynamics and cumulative drug exposure, thereby providing a quantitative basis for personalized therapy design.

Table 3. Mean Squared Error (MSE) comparison across classical tumor growth models and game theory.

Therapy Strategy	Exponential	Logistic	Gompertz	Game Theory
Capecitabine with MTD-Up	0.0788	0.0574	0.0694	0.0300
Capecitabine dose modulation-Down	0.0675	0.0434	0.058	0.0400
Capecitabine intermittent-Up	0.0809	0.0789	0.0803	0.0600
Capecitabine intermittent-Down	0.0533	0.0476	0.0141	0.0100
Gemcitabine with MTD-Up	0.0969	0.0805	0.0873	0.0300
Gemcitabine with MTD-Down	0.0848	0.0786	0.0785	0.0700
Gemcitabine dose modulation-Up	0.0966	0.0642	0.0641	0.0356
Gemcitabine intermittent-Up	0.0640	0.0525	0.1220	0.0400
Gemcitabine intermittent-Down	0.0809	0.0687	0.0700	0.0300

3.2. Sensitive and Resistant Cell Interactions

The analytical framework presented in this section builds upon the Stackelberg evolutionary game theory approach, which has been developed and extended over time in prior work [11,23–25]. In particular, Stein et al. [14] applied and further extended this framework in the context of cancer treatment, and we build on their formulation to explore treatment-induced evolutionary responses under varying therapeutic strategies, including EET and EVT. While the structure follows their established model, our simulations use modified parameters tailored to the current study, resulting in distinct graphical outcomes.

Tumors consist of heterogeneous populations of cancer cells with varying degrees of therapeutic resistance, all competing for limited resources. Adaptive therapy is a dose modulation strategy guided by treatment response, which leverages the intrinsic competition between sensitive and resistant cancer cells to optimize therapeutic efficacy. Treatment response depends on the adaptive capacity of cancer cells; therefore, adaptive therapy can incorporate evolutionary principles to better account for these adaptive dynamics.

By maintaining a substantial population of treatment-sensitive cells through controlled drug dosing or periodic treatment interruptions, adaptive therapy creates an environment in which sensitive cells suppress the proliferation of treatment-resistant subpopulations via competitive interactions. This strategic approach not only delays the emergence of drug resistance but also enhances long-term tumor control, offering a promising alternative to traditional high-dose regimens.

Evolutionary Stackelberg game theory provides a framework for refining cancer treatment strategies. In this context, breast cancer therapy is conceptualized as a dynamic interaction between a physician and a polymorphic population of cancer cells, encompassing both treatment-sensitive and treatment-resistant phenotypes. To model tumor growth dynamics and the evolutionary adaptations of resistance-related traits over time, a fitness-generating function (G-function) is employed. This function simulates the competitive interplay between subpopulations, offering insights into optimal therapeutic interventions that leverage evolutionary principles to suppress resistance and enhance treatment effectiveness.

The dynamics of two tumor subpopulations are described by the following system of equations. The variable x_S represents cells sensitive to treatment, while x_R denotes cells resistant to treatment. Both equations incorporate ecological competition and pharmacodynamic effects. The growth of the sensitive subpopulation x_S is governed by its intrinsic proliferation rate r_S . The term $\alpha_{SS}x_S + \alpha_{SR}x_R$ represents the total competitive pressure experienced by sensitive cells, where α_{SS} denotes intra-type competition among sensitive cells, and α_{SR} captures inter-type competition exerted by resistant cells. The cytotoxic effect of therapy is modeled by the term $\frac{D}{k+bu_S}$. Here, D denotes the administered drug dose, and u_S represents the treatment resistance in the sensitive cell population. In this scenario $u_S = 0$:

$$\frac{dx_S}{dt} = x_S \left(r_S \left(1 - \frac{\alpha_{SS}x_S + \alpha_{SR}x_R}{K(u)} \right) - \frac{D}{k + bu_S} \right), \quad u_S = 0. \quad (7)$$

The dynamics of the resistant subpopulation x_R are similarly governed by its intrinsic growth rate r_R . α_{RR} represents intra-type competition among resistant cells, and α_{RS} accounts for inter-type competition from sensitive cells. The treatment effect on resistant cells is captured by the term $\frac{D}{k+bu_R}$, where u_R denotes the treatment resistance in the resistant subpopulation:

$$\frac{dx_R}{dt} = x_R \left(r_R \left(1 - \frac{\alpha_{RS}x_S + \alpha_{RR}x_R}{K(u)} \right) - \frac{D}{k + bu_R} \right). \quad (8)$$

In our study, the experimental data correspond to tumor burden, which were normalized to the unit interval $[0, 1]$ using min–max scaling. Accordingly, $x = x_S + x_R$ denotes the normalized total tumor burden. To optimize treatment decisions, we define the physician's objective function Q as

$$Q(D, u_R, x^*(D, u_R)) = \begin{cases} Q^{\max} - C_1 \left(\frac{x^*(D, u_R)}{K} \right)^2 - C_2 u_R^2 - C_3 D^2, \\ (D, u_R) \in \text{stabilization region}, \\ \text{undefined, elsewhere.} \end{cases} \quad (9)$$

In Equation (9), $Q^{\max}$ represents the maximum attainable value of the objective function. The term $x^*(D, u_R)$ corresponds to the ecological equilibrium, while u_R denotes the resistance strategy adopted by treatment-resistant cells and D is treatment dose. The parameters C_1 , C_2 , and C_3 reflect the relative importance of tumor burden, resistance, and toxicity.

Progression time is defined as the moment when tumor burden exceeds a fraction δ of the carrying capacity. In this study, we set $\delta = 0.7$, and therefore the threshold is given by $\delta K_{\max} = 0.7K_{\max}$. The goal is to identify treatment doses and resistance levels that keep the tumor below this threshold, enabling sustained control. Tumor behavior is classified into elimination, stabilization, and progression regions.

We adopt the classification of therapeutic strategies proposed by Stein et al. [14], which distinguishes between ecologically and evolutionarily enlightened approaches. These are formalized through Nash and Stackelberg optimization frameworks. The Nash strategy assumes a fixed resistance level and optimizes dose accordingly:

$$D^*(u_R) = \arg \max_D Q(D, u_R, x^*(D, u_R)). \quad (10)$$

This strategy aligns with the intersection of the cancer cells' evolutionary response curve (ESS strategy $u_R^*(D)$) and the physician's best response curve $D^*(u_R)$. The Stackelberg strategy anticipates the cancer cells' evolutionary response:

$$D_S = \arg \max_D Q(D, u_R^*(D), x^*(D, u_R^*(D))). \quad (11)$$

In order to illustrate the comparative outcomes of different therapeutic strategies, we analyzed Mouse 7 under three distinct approaches: MTD, the ecologically enlightened strategy (Nash equilibrium), and the evolutionarily enlightened strategy (Stackelberg equilibrium). The parametrization was based on the framework introduced by Stein et al. [14], adapted to our experimental data (Figure 3). The model was parameterized with $Q^{\max} = 1$, $C_1 = 0.51$, $C_2 = 0.08$, and $C_3 = 0.41$, reflecting the relative weighting of tumor burden, treatment-induced resistance, and therapy toxicity in the physician's objective function. Transition rates between phenotypes were set as $\alpha_{RS} = 0.60$, $\alpha_{RR} = \alpha_{SS} = 1.0$, and $\alpha_{SR} = 0.25$.

Under this parametrization, all three strategies fell within the stabilization region, where tumor burden remains nonzero but manageable. Among them, the Stackelberg equilibrium achieved the highest objective function value, followed by the Nash equilibrium, while MTD resulted in the lowest value. It should be noted that the near-zero dosing outcome observed under the Stackelberg strategy is influenced by the toxicity weight C_3 , which strongly penalizes drug administration. This reflects the sensitivity of the framework to parameter selection and highlights the importance of careful calibration and biological validation.

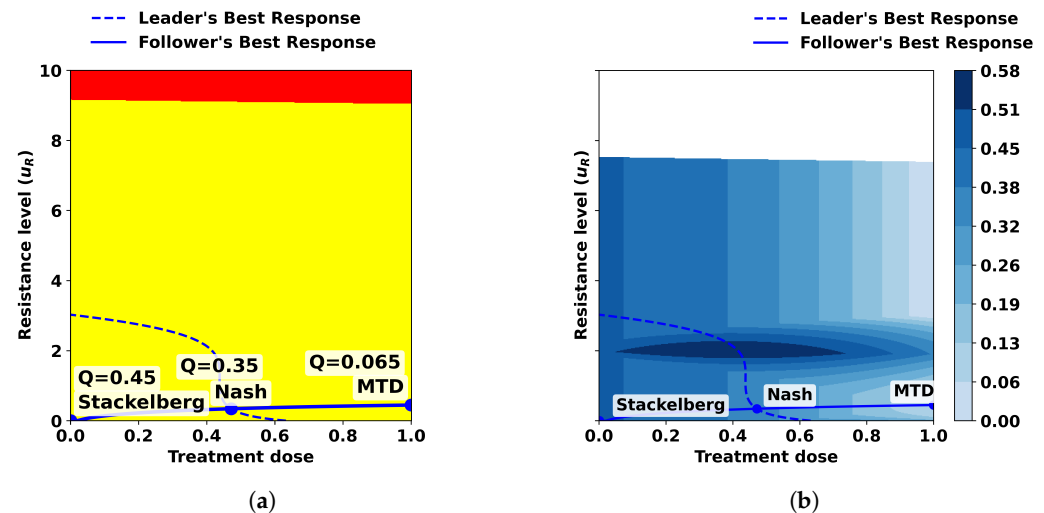


Figure 3. Outcomes of three therapeutic strategies for Mouse 7: MTD, Nash equilibrium, and Stackelberg equilibrium. (a) Regions of tumor stabilization (yellow) versus progression (red). (b) Physician's objective function under the chosen parametrization.

Furthermore, the Stackelberg strategy not only leads to a lower treatment dose and toxicity compared to both the Nash strategy and MTD but also reduces treatment-induced resistance, demonstrating its potential for optimizing therapeutic outcomes while minimizing adverse effects.

3.3. Diffusion in Sensitive and Resistant Cells

In this study, we adopt a conceptual framework based on evolutionary game theory, in which drug-sensitive cancer cells are represented as cooperative players and drug-resistant cells as opportunistic defectors. This abstraction has been widely used in the literature to model the ecological and evolutionary dynamics of tumor cell populations competing for limited resources within the tumor microenvironment [26–28].

Understanding spatial processes is essential, as tumor growth and treatment response are fundamentally shaped by the interplay between cellular motility, resource availability, and local interactions [27,29]. Although this classification offers a useful lens for analyzing treatment-induced selection pressures and spatial dynamics, we recognize that it simplifies the biological complexity of tumor behavior. For instance, under certain conditions, resistant cells may contribute to tumor-wide benefits—such as VEGF-mediated angiogenesis in hypoxic environments—which improve resource availability for both sensitive and resistant populations. This perspective aligns with recent work that integrates evolutionary game theory and control approaches to cancer therapy optimization [16,30]. These examples challenge the strict cooperator-defector dichotomy and underscore the context-dependent nature of cellular behavior.

To account for this nuance, our model focuses on the ecological consequences of differential diffusion and competition, rather than assigning fixed behavioral roles. By examining the conditions under which cooperative-like traits can suppress or coexist with defector-like strategies.

In spatially structured tumor environments, cells disperse to colonize new territories through migration and diffusion. The rate and pattern of this diffusion strongly influence population distribution and competitive dynamics. Slow diffusion of cooperative cells tends to channel their migration along high-benefit pathways, optimizing resource utilization and reinforcing mutualistic interactions. Conversely, rapid diffusion of defector cells

enables them to quickly locate and exploit favorable niches, often undermining cooperative expansion and destabilizing spatial organization [31,32].

These opposing forces cooperative resource optimization versus opportunistic exploitation can drive the coexistence of both phenotypes, producing a spectrum of spatial patterns ranging from stable segregation to dynamic spatial chaos. The local microenvironment is further shaped by the production and consumption of shared resources, which modulate population stability and competitive balance over time. Self-organization arising from these diffusion processes not only strengthens cooperative interactions but also generates habitat diversity, a fundamental driver of ecological complexity in biological systems [27].

Building on the evolutionary dynamics presented in previous sections, we now introduce a two-dimensional migration diffusion model to explicitly capture the role of spatial heterogeneity in shaping interactions between cancer subpopulations. This spatially explicit framework provides deeper insight into the mechanisms governing tumor progression and therapeutic response, offering a quantitative basis for the development of spatially informed, evolution-aware treatment protocols. The model is defined as

$$\frac{dx_S}{dt} = DI_S \nabla^2 x_S + x_S G \quad (12)$$

$$\frac{dx_R}{dt} = DI_R \nabla^2 x_R + x_R G. \quad (13)$$

The functions x_S and x_R represent the densities of cooperative and defector cells at time t , respectively. The diffusion constants DI_S and DI_R determine the migration rates for cooperative and defector cells, while ∇^2 denotes the diffusion operator. This continuous spatial expansion reveals a complex landscape of dynamic pattern formation within tumor microenvironments.

When defector cells exhibit a higher dispersal rate than cooperative cells ($DI_R \geq DI_S$), the resulting spatial dynamics do not support the dominance of cooperators. Instead, defectors engage in persistent exploration of resource-rich pathways. Slow migration fosters cooperator aggregation, stabilizing their presence in favorable regions, whereas rapid migration allows defectors to quickly identify cooperative pathways. However, excessive mobility also undermines their ability to fully exploit any specific pathway efficiently, leading to transient rather than sustained advantage.

Existing models for various cancers frequently evaluate cell fitness based solely on the relative abundance of phenotypes within the population. However, these models often overlook the spatial structure of cells beyond one dimension or fail to incorporate cell diffusion dynamics, which are critical for accurately capturing tumor behavior. In complex biological systems, such as the human body, cells, hormones, and growth factors exhibit mobility, influencing intercellular interactions and overall population dynamics [33–35]. The dispersal of these components plays a fundamental role in tumor progression and response to treatment. The rate at which these elements disperse is governed by the diffusion coefficient in the reaction-diffusion equation, which may vary across different spatial coordinates, further influencing population dynamics.

Since the equations presented in this section are second-order partial differential equations, their solutions require numerical methods for accurate computation. To approximate function values at each point in the spatial domain and across generations, numerical techniques such as the five-point stencil method are employed to solve all replicator equations, ensuring precision and practical applicability in modeling tumor evolution.

As proposed, the breast cancer game model is implemented within a two-dimensional spatial environment, represented as an $n \times n$ computational grid, where n denotes the number of cells along each coordinate direction. Each grid cell contains a type i cancer cell with a probability of p_i , while remaining vacant with a probability of $1 - p_i$, allowing for a probabilistic representation of tumor heterogeneity and spatial distribution.

Figure 4, depicting the diffusion dynamics of Mouse 22 across multiple generations, illustrates tumor progression trends. Specifically, at generation 90, the density of treatment-resistant cells surpasses that of treatment-sensitive cells, highlighting the selective pressures that drive resistance evolution over time.

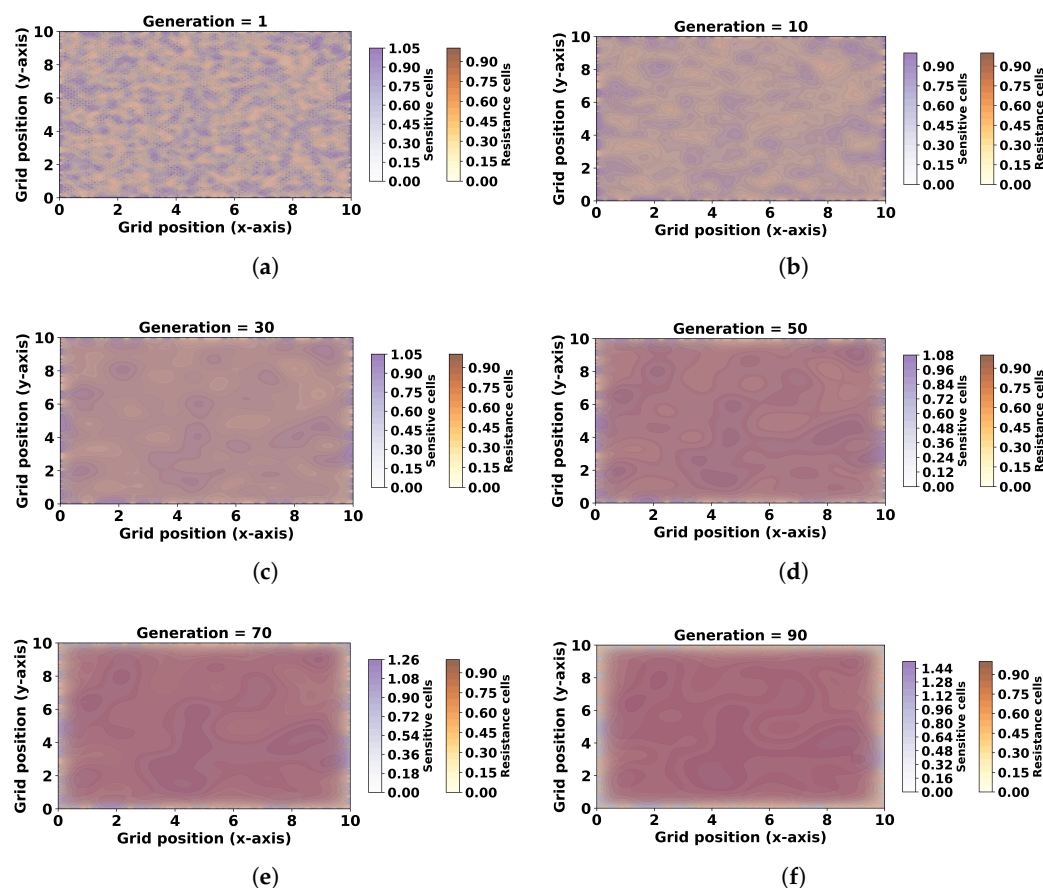


Figure 4. Two-dimensional representation of population dynamics and evolutionary progression under the condition $DI = 0.24$ on a 100×100 grid for Mouse 22, with initial values set to $x_S = 65\%$ and $x_R = 35\%$ of the initial population. (a) Population dynamics in generation 1; (b) Population dynamics in generation 10; (c) Population dynamics in generation 30; (d) Population dynamics in generation 50; (e) Population dynamics in generation 70; (f) Population dynamics in generation 90.

As illustrated in Figure 5, the population dynamics for Mouse 22 exhibit distinct shifts depending on the applied treatment protocol. When the Dose Modulation Adaptive Therapy protocol is implemented, the population predominantly transitions toward treatment-resistant phenotypes. Conversely, under the Intermittent Adaptive Therapy protocol, the population dynamics favor an increase in treatment-sensitive phenotypes, highlighting the differential impact of these therapeutic strategies on cellular evolution.

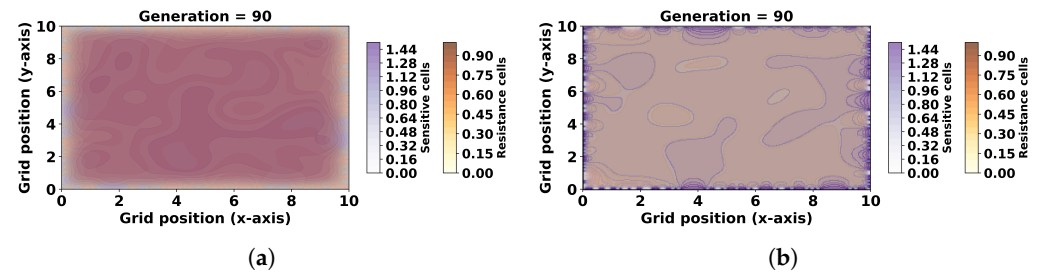


Figure 5. Two-dimensional representation of population dynamics and evolutionary progression under the condition $DI = 0.24$ on a 100×100 grid for Mouse 22: (a) Population state and evolutionary trajectory under the Dose Modulation Adaptive Therapy protocol. (b) Population state and evolutionary trajectory under the Intermittent Adaptive Therapy protocol.

3.4. Integration of Experimental Data into the Simulation Framework

To ensure biological relevance and model fidelity, we integrated real-world experimental data into the simulation environment. Specifically, tumor volume measurements obtained from murine xenograft models treated with Gemcitabine and Capecitabine under distinct therapeutic protocols were used as reference datasets. These data served as the foundation for estimating key pharmacodynamic parameters through nonlinear least-squares optimization using the Python-based GEKKO package, a specialized tool for dynamic modeling and parameter fitting.

The estimated parameters informed the construction of a biologically grounded simulation framework capable of capturing tumor dynamics under both single-agent and combination therapies. This calibration process enabled the model to reproduce observed treatment responses and provided a reliable basis for evaluating alternative therapeutic strategies, including reinforcement learning (RL)-based schedules. In our framework, we distinguish between two categories of parameters:

- Therapy-specific parameters ($g, k, b, K_{\max}$): These define the pharmacodynamic characteristics of each drug and are estimated independently for each treatment protocol. They remain constant across patients receiving the same therapy.
- Subject-specific parameters (σ, r, u_0): These capture individual tumor growth behavior, initial tumor burden, and biological variability. They are uniquely estimated for each subject based on their experimental data.

By combining these parameter sets, we simulate multiple treatment protocols for a given patient and assess their relative effectiveness. This enables personalized therapy planning by comparing outcomes such as tumor volume reduction and cumulative drug exposure across different regimens. The simulation framework thus supports the identification of optimal strategies tailored to individual patient profiles, balancing therapeutic efficacy with toxicity constraints.

As illustrated in Figure 6, Mouse 6 was treated experimentally under the Gemcitabine MTD protocol. Model based comparative analysis, however, indicates that if an intermittent adaptive therapy protocol had been applied, a more pronounced reduction in tumor burden could have been achieved, accompanied by a lower cumulative drug dose. These results suggest that intermittent scheduling may improve treatment effectiveness by better controlling the tumor while reducing the overall drug burden for Mouse 6.

Similarly, Figure 7 presents results for Mouse 25, comparing the real-data intermittent adaptive therapy protocol with a model simulated Dose Modulation adaptive therapy. In this case, the Dose Modulation approach produced a greater reduction in tumor burden, but at the expense of increased drug administration. Specifically, while the total dose delivered in the real data protocol was 16 units, the model required 22 units to achieve the observed

tumor suppression. Such analyses underscore the capacity of the modeling framework to quantify trade-offs between tumor control and drug exposure, thereby informing the selection of treatment strategies that balance efficacy with toxicity considerations.

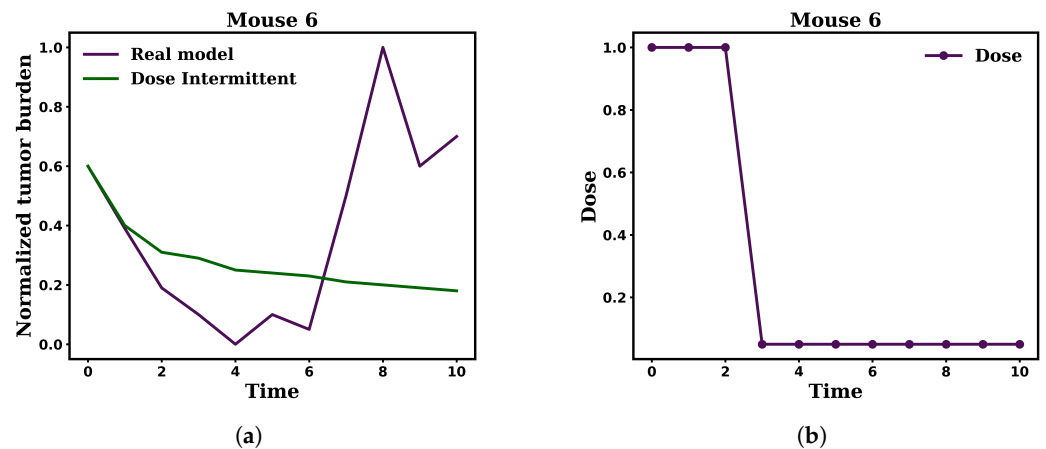


Figure 6. Model-based comparison of (a) population dynamics and (b) treatment dose rates for mouse 6. While the experimental protocol involved MTD administration of Gemcitabine, simulation results suggest that an intermittent adaptive therapy schedule could have achieved greater tumor suppression with reduced cumulative drug exposure.

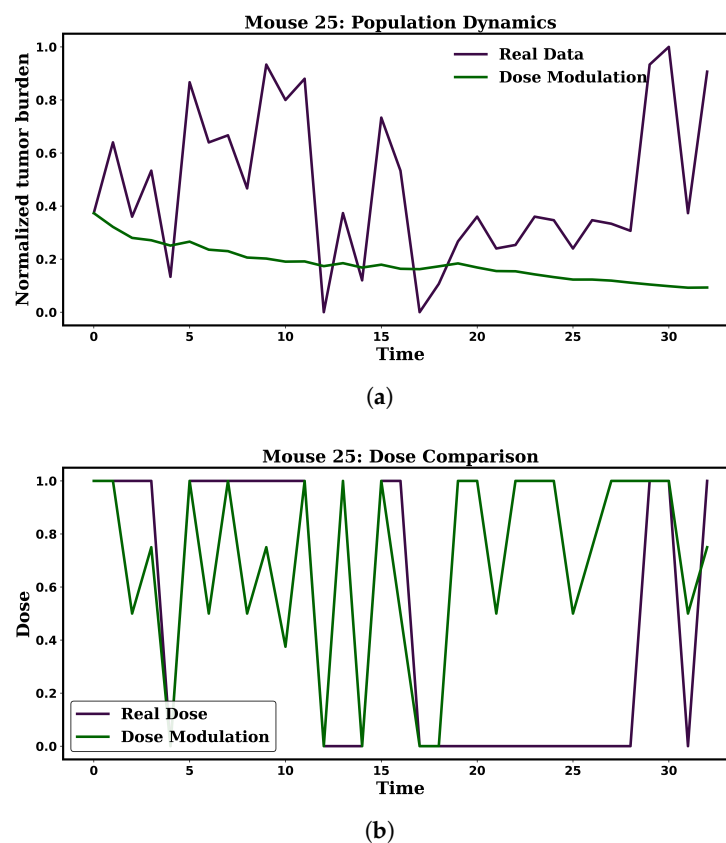


Figure 7. Comparison of (a) population dynamics and (b) treatment dose rates for mouse 25 under intermittent adaptive therapy (real data) and Dose Modulation adaptive therapy.

3.5. Application of Reinforcement Learning for Adaptive Cancer Therapy

Recent advances in artificial intelligence have facilitated the emergence of reinforcement learning (RL) frameworks capable of optimizing therapeutic strategies in complex and dynamic biological systems. In the context of oncology, RL offers a principled approach to designing adaptive treatment schedules that respond to the evolving state of the

tumor and its microenvironment. In this study, we developed a biologically informed RL model to generate personalized chemotherapy protocols for tumor-bearing mice, with the overarching goal of minimizing tumor burden, delaying or preventing the emergence of drug resistance, and avoiding overtreatment, key objectives in precision oncology [17,36].

The RL agent was trained on tumor burden and dosing data, enabling it to learn from the temporal interplay between treatment intensity, tumor regression, and resistance dynamics. The environment was modeled as a discrete-time system, where at each time step the agent observed the current state $s = (X, u)$, with X representing the tumor burden and u denoting the level of resistance. Based on this state, the agent selected an action $a \in \{0, 1, 2, 3\}$ corresponding to no dose, low dose, medium dose, or high dose administration, respectively.

A central component of the RL framework is the reward function, which encodes the therapeutic objectives and guides policy optimization. In our formulation, the reward at each time step was defined as

$$\text{reward} = C_1 \cdot (x_{\text{prev}} - x) - C_2 \cdot D - C_3 \cdot u, \quad (14)$$

where $(x_{\text{prev}} - x)$ quantifies the reduction in tumor burden relative to the previous time step, D is the administered drug dose, and u is the resistance level. The coefficients C_1 , C_2 , and C_3 are tunable parameters that determine the relative importance of tumor shrinkage, dose minimization, and resistance suppression, respectively. Increasing C_1 prioritizes aggressive tumor reduction, while larger C_2 values penalize excessive drug usage to mitigate toxicity. Similarly, higher C_3 values place greater emphasis on controlling resistance evolution, potentially favoring strategies that maintain a stable, treatment-sensitive tumor subpopulation. By adjusting these weights, the framework can be tailored to different clinical priorities, such as maximizing tumor control in aggressive disease or minimizing toxicity in frail patients.

The agent was trained using tabular Q-learning, a RL algorithm, with learning rate $\alpha = 0.1$, discount factor $\gamma = 0.9$, and exploration rate $\epsilon = 0.1$. The Q-table was updated iteratively according to the Bellman equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right], \quad (15)$$

where r_t is the reward obtained at time t . The discount factor γ controls the trade-off between immediate and long-term gains, with higher values encouraging the agent to adopt strategies that yield sustained tumor control rather than short-lived regressions.

Once trained, the reinforcement learning (RL) agent was applied to an unseen test subject (Mouse 25) within a simulated environment. The simulation was constructed using subject-specific parameters estimated from experimental tumor volume data. To evaluate the agent's performance, its dosing schedule was compared against the actual intermittent adaptive therapy protocol recorded for Mouse 25 in the experimental dataset (Figure 8). The RL-derived schedule achieved a reduction in tumor burden while lowering cumulative drug exposure, demonstrating its ability to balance efficacy with toxicity constraints. Notably, the agent exhibited adaptive dosing behavior—intensifying treatment in response to rising tumor burden and reducing it when resistance levels were low. This behavior, which aligns with clinical intuition, emerges directly from the structure of the reward function used during training. Its performance should be interpreted within the context of the simulation framework and the subject-specific parameter estimates embedded in the model.

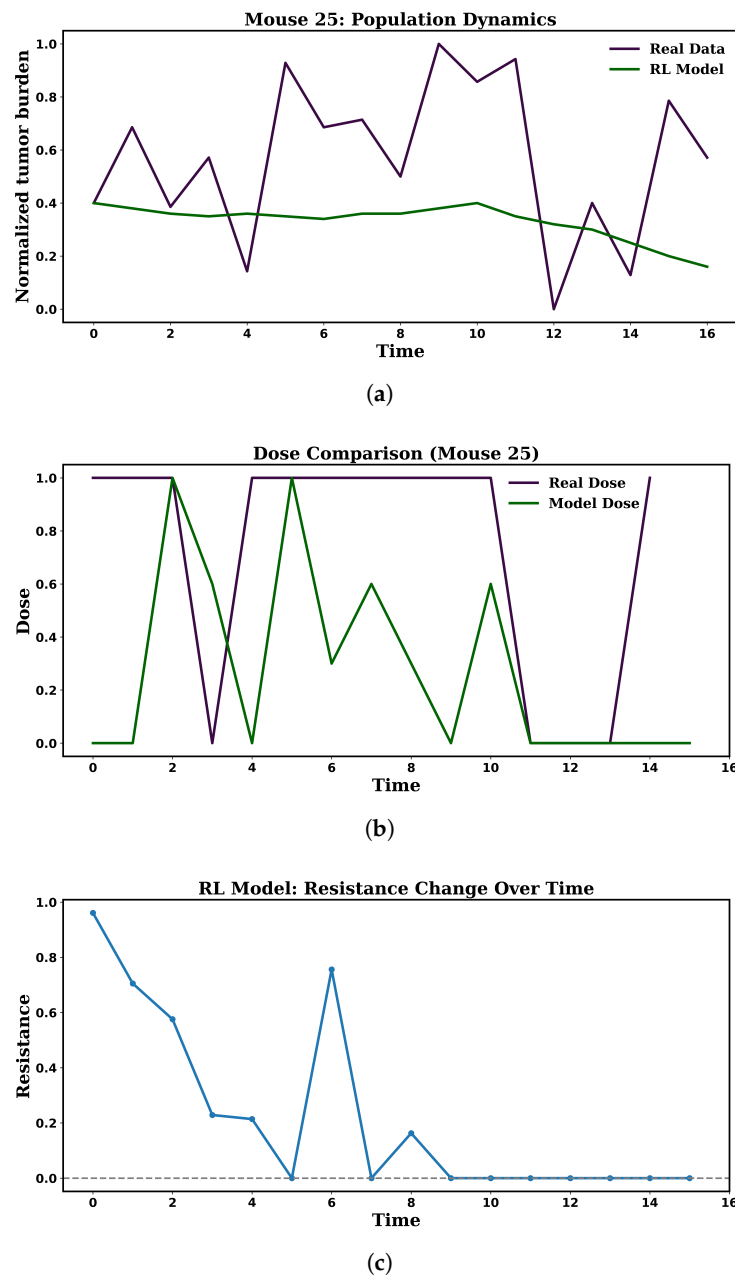


Figure 8. Illustration of (a) population dynamics, (b) treatment dose rates for mouse 25 under experimental intermittent adaptive therapy and simulated RL-based schedule ($C_1 = 2$, $C_2 = 0.1$, $C_3 = 0.1$), and (c) resistance level trajectories generated by the simulation.

A key strength of our approach lies in the integration of tumor growth and resistance models within the RL framework. By embedding differential equation-based dynamics into the environment, the agent learns policies that are not only data-driven but also biologically interpretable and generalizable across individuals and treatment contexts [37]. This hybrid modeling paradigm enhances clinical relevance, offering a pathway toward evolution-aware, patient-specific cancer therapies that adapt in real time to the shifting landscape of tumor ecology.

After estimating subject-specific parameters for each mouse, it becomes feasible to simulate and assess the effectiveness of various therapeutic strategies in a case-specific manner. Table 4 provides a quantitative comparison of three treatment approaches applied to Mouse 25 over 16 time units. The RL-derived schedule demonstrates more effective performance, achieving the lowest final tumor burden and resistance level while requiring

the least cumulative drug exposure. In contrast, the dose modulation strategy effectively reduces tumor burden but entails a higher overall drug consumption.

Table 4. Quantitative comparison of treatment strategies for Mouse 25 over 16 time units. Metrics include Final Tumor Burden (normalized), Total Cumulative Dose, and Final Resistance Level.

Model	Final Tumor Burden (Normalized)	Total Cumulative Dose	Final Resistance Level
Intermittent Dose (Real Data)	0.65	12.0	0.40
Dose Modulation	0.19	7.5	0.05
RL Model	0.16	4.6	0.001

3.6. Extension from Single-Agent to Combination Therapy Modeling

Combination chemotherapy, in which two or more cytotoxic or cytostatic agents are administered in a coordinated regimen, is a cornerstone of modern oncology. Such protocols aim to exploit complementary mechanisms of action, reduce the likelihood of cross-resistance, and achieve synergistic tumor suppression while maintaining acceptable toxicity profiles [38]. By targeting distinct molecular pathways or cell cycle phases, combination regimens can enhance therapeutic efficacy compared to single-agent protocols, particularly in heterogeneous tumors where subpopulations may exhibit differential drug sensitivities [38,39]. Clinically, combinations such as Gemcitabine with Capecitabine have been investigated in several malignancies, including pancreatic and breast cancers, with evidence of improved progression-free survival in selected patient cohorts [40,41].

Building on this rationale, we extended our reinforcement learning based tumor dynamics framework originally developed for single agent treatment protocols to capture the dynamics of dual drug interactions. The baseline datasets for the murine experiments comprised tumor volume measurements obtained under regimens of Gemcitabine and Capecitabine. Using these data as a reference, we developed a biologically informed simulation environment that integrates the pharmacodynamic effects of both Gemcitabine (D_G) and Capecitabine (D_C), while enforcing clinically relevant constraints, including an upper bound on the total normalized dose administered at each time step ($D_G + D_C < 1$). This formulation enables the model to explore adaptive dosing strategies that balance efficacy and toxicity within a realistic therapeutic window.

The reinforcement learning agent was trained using the Q-learning algorithm to adaptively allocate doses between the two therapeutic agents, with the overarching objective of maximizing tumor regression while simultaneously minimizing cumulative toxicity and the emergence of drug resistance. The reward function guiding the agent learning process was formulated as:

$$\text{reward} = (x_{\text{prev}} - x) - \lambda_{\text{dose}} \cdot (D_G + D_C) - \beta_{\text{final}} \cdot u, \quad (16)$$

where $(x_{\text{prev}} - x)$ quantifies the reduction in tumor burden between consecutive time steps, $(D_G + D_C)$ denotes the total normalized dose of Gemcitabine and Capecitabine administered at that step, and u represents the current level of treatment resistance within the tumor population. The coefficient λ_{dose} penalizes excessive drug administration to limit cumulative toxicity, while β_{final} penalizes high resistance levels to encourage strategies that maintain treatment sensitivity. In this study, we set $\lambda_{\text{dose}} = 0.08$ and $\beta_{\text{final}} = 0.15$, parameter values selected to balance the competing clinical priorities of maximizing therapeutic efficacy and minimizing adverse effects.

When applied to Mouse 25, the learned combination policy achieved a consistently lower tumor burden compared to the single-agent trajectory, while maintaining a smoother

dosing profile and reducing reliance on high-dose pulses (Figure 9). Importantly, the total normalized dose budget was identical to that of the single-agent protocol, indicating that the observed improvement was due to dose diversification rather than escalation. This aligns with prior computational and experimental studies showing that alternating or mixing agents with distinct mechanisms can suppress resistant clones more effectively than monotherapy [42,43].

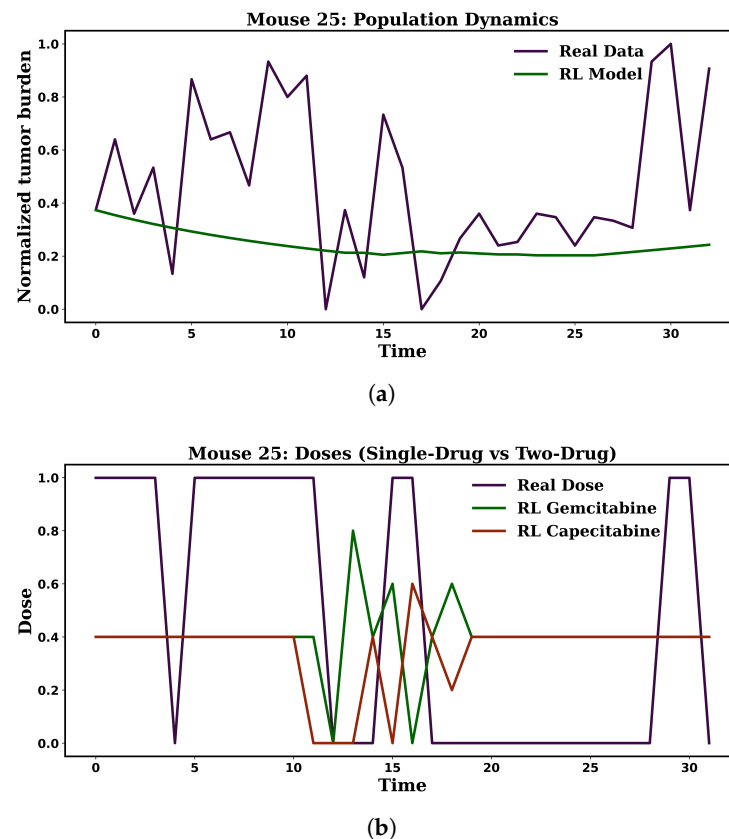


Figure 9. Comparison of (a) population dynamics and (b) treatment dose rates for mouse 25 under intermittent adaptive therapy protocol and RL (two drugs).

These results illustrate how reinforcement learning can generalize empirical single-agent data into optimized multi-agent treatment strategies, offering a pathway toward personalized, evolution-aware cancer therapy. By explicitly modeling pharmacodynamic constraints and tumor heterogeneity, the framework can identify dosing schedules that exploit drug complementarity, potentially improving long-term disease control without increasing cumulative toxicity.

4. Discussion and Conclusions

In this study, we developed and systematically evaluated an integrative modeling framework for personalized cancer therapy that combines evolutionary game theory, tumor growth modeling, and reinforcement learning (RL). The central objective was to capture the ecological and evolutionary dynamics of heterogeneous tumor cell populations—comprising both treatment-sensitive and treatment-resistant phenotypes—and to leverage these insights for the design of adaptive, evolution-aware treatment strategies.

At the core of our approach lies an Evolutionary Stackelberg Game formulation, in which the physician is modeled as a rational leader and the tumor cell population as a responsive follower. This leader–follower structure enables the anticipation of tumor evolutionary responses to treatment, allowing the physician’s strategy to be optimized in a forward-looking manner. The tumor’s ecological dynamics were described using a fitness-

generating function (G-function) that incorporates logistic growth, treatment-induced mortality, and resistance-associated fitness costs.

A key strength of this framework is the incorporation of both treatment-specific and subject-specific parameters, estimated directly from experimental data using the Python-based GEKKO optimization package. This calibration step ensures that the model reflects individual variability in tumor growth kinetics and drug response, thereby enhancing predictive accuracy. Using tumor volume data from 70 breast cancer-bearing mice treated with Capecitabine or Gemcitabine, we parameterized the model to simulate three clinically relevant protocols: MTD, Dose Modulation Adaptive Therapy, and Intermittent Adaptive Therapy.

Comparative analysis against classical tumor growth models—Exponential, Logistic, and Gompertz—highlighted the limitations of these traditional approaches. While classical models have been widely used in oncology modeling [44,45], they often fail to capture non-monotonic behaviors such as treatment-induced regressions, rebounds, and phenotypic switching, which are increasingly recognized in both preclinical and clinical settings [3,46]. In contrast, our game-theoretic model aligned more closely with experimental observations, particularly in scenarios involving adaptive or intermittent dosing, underscoring the value of explicitly modeling ecological and evolutionary feedbacks.

Our use of reinforcement learning builds on recent efforts to apply AI-driven optimization to cancer therapy [37,47]. For single-agent therapy, an RL agent trained via Q-learning learned to balance tumor regression against cumulative drug exposure and resistance suppression, as encoded in a multi-term reward function. This approach was then extended to combination therapy by incorporating two drugs—Gemcitabine and Capecitabine—into a pharmacodynamically constrained environment. The RL agent dynamically allocated doses between the two agents, achieving superior tumor control compared to single-agent protocols without exceeding the total dose budget. These results support the hypothesis proposed by Gatenby et al. [3] and further explored by West et al. [48] that dose diversification, rather than escalation, can yield more effective and biologically efficient treatment schedules.

We also examined the spatial dimension of tumor ecology by incorporating reaction-diffusion equations to model the dispersal of sensitive and resistant cells. This builds on prior work demonstrating the role of spatial heterogeneity in resistance evolution [16,29]. In our spatially explicit framework, drug-sensitive cells were conceptualized as cooperators, often exhibiting slower migration and reliance on shared resources, while resistant cells acted as defectors, characterized by higher motility and opportunistic exploitation of spatial niches.

Overall, our findings demonstrate that integrating evolutionary game theory, subject-specific parameter estimation, reinforcement learning, and spatial ecology yields a powerful framework for optimizing cancer therapy. The Evolutionary Stackelberg Game model consistently outperformed classical growth models in reproducing complex tumor dynamics and provided a principled basis for comparing treatment protocols. RL-based adaptive scheduling proved effective for both single- and multi-drug regimens, while spatial modeling illuminated the critical role of phenotypic diffusion in resistance evolution.

This work underscores the importance of explicitly accounting for ecological and evolutionary processes in treatment design. By anticipating tumor adaptation, exploiting competitive interactions between sensitive and resistant cells, and tailoring strategies to individual patient and treatment parameters, such integrative models have the potential to improve long-term tumor control, reduce cumulative toxicity, and ultimately enhance patient outcomes. Future research should focus on validating these approaches in larger,

clinically diverse datasets and extending the framework to incorporate immune–tumor interactions, multi-scale modeling, and real-time treatment adaptation.

Author Contributions: F.T. conducted the theoretical modeling and drafted the manuscript. J.S.S. contributed to the mathematical analysis. D.M. and M.H.M. reviewed and edited the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study are publicly available and were obtained from the article “Testing Adaptive Therapy Protocols Using Gemcitabine and Capecitabine in a Preclinical Model of Endocrine-Resistant Breast Cancer” (*Cancers* **2024**, *16*, 257) [13]. All relevant information and datasets can be accessed through the original publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Crawford, S. Is it time for a new paradigm for systemic cancer treatment? Lessons from a century of cancer chemotherapy. *Front. Pharmacol.* **2013**, *4*, 68. [\[CrossRef\]](#)
2. Gatenby, R.A.; Frieden, B.R. Inducing catastrophe in malignant growth. *Math. Med. Biol. J. IMA* **2008**, *25*, 267–283. [\[CrossRef\]](#)
3. Gatenby, R.A.; Silva, A.S.; Gillies, R.J.; Frieden, B.R. Adaptive therapy. *Cancer Res.* **2009**, *69*, 4894–4903. [\[CrossRef\]](#)
4. Belkhir, S.; Thomas, F.; Roche, B. Darwinian approaches for cancer treatment: Benefits of mathematical modeling. *Cancers* **2021**, *13*, 4448. [\[CrossRef\]](#)
5. Deris, A.; Sohrabi-Haghighat, M. Analysis of cancerous tumor growth by the competitive model based on the evolutionary game theory. *Int. J. Nonlinear Anal. Appl.* **2023**, *14*, 1903–1910.
6. Hajdowska, K.; Swierniak, A.; Borys, D. Modelling Changes in Genetic Heterogeneity Using Games with Resources. *Comput. Methods Programs Biomed.* **2025**, *270*, 108916. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Bayer, P.; Gatenby, R.A.; McDonald, P.H.; Duckett, D.R.; Staňková, K.; Brown, J.S. Coordination games in cancer. *PLoS ONE* **2022**, *17*, e0261578. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Tavakoli, F.; Sartakhti, J.S.; Manshaei, M.H.; Basanta, D. Cancer immunoediting: A game theoretical approach. *In Silico Biol.* **2020**, *14*, 1–12. [\[CrossRef\]](#)
9. Maynard Smith, J. The theory of games and the evolution of animal conflicts. *J. Theor. Biol.* **1974**, *47*, 209–221. [\[CrossRef\]](#)
10. Romano, C.; Borri, A.; Di Benedetto, M.D. Stackelberg evolutionary games for cancer modeling and treatment. In Proceedings of the 2024 IEEE 63rd Conference on Decision and Control (CDC), Milan, Italy, 16–19 December 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 7044–7049.
11. Salvioli, M.; Garjani, H.; Satouri, M.; Broom, M.; Viossat, Y.; Brown, J.S.; Dubbeldam, J.; Staňková, K. Stackelberg Evolutionary Games of Cancer Treatment: What Treatment Strategy to Choose if Cancer Can be Stabilized? *Dyn. Games Appl.* **2024**, *15*, 1750–1769. [\[CrossRef\]](#)
12. Reed, D.R.; Metts, J.; Pressley, M.; Fridley, B.L.; Hayashi, M.; Isakoff, M.S.; Loeb, D.M.; Mankanji, R.; Roberts, R.D.; Trucco, M.; et al. An evolutionary framework for treating pediatric sarcomas. *Cancer* **2020**, *126*, 2577. [\[CrossRef\]](#)
13. Seyed, S.; Teo, R.; Foster, L.; Saha, D.; Mina, L.; Northfelt, D.; Anderson, K.S.; Shibata, D.; Gatenby, R.; Cisneros, L.H.; et al. Testing adaptive therapy protocols using gemcitabine and capecitabine in a preclinical model of endocrine-resistant breast cancer. *Cancers* **2024**, *16*, 257. [\[CrossRef\]](#)
14. Stein, J.; Salvioli, M.; Garjani, H.; Dubbeldam, J.; Viossat, Y.; Brown, J.S.; Staňková, K. Stackelberg evolutionary game theory: How to manage evolving systems. *Philos. Trans. R. Soc. B* **2023**, *378*, 20210495. [\[CrossRef\]](#)
15. Pressley, M.; Salvioli, M.; Lewis, D.B.; Richards, C.L.; Brown, J.S.; Staňková, K. Evolutionary dynamics of treatment-induced resistance in cancer informs understanding of rapid evolution in natural systems. *Front. Ecol. Evol.* **2021**, *9*, 681121. [\[CrossRef\]](#)
16. Gluzman, M.; Scott, J.G.; Vladimirsky, A. Optimizing adaptive cancer therapy: Dynamic programming and evolutionary game theory. *Proc. R. Soc. B* **2020**, *287*, 20192454. [\[CrossRef\]](#)
17. Zhang, J.; Cunningham, J.J.; Brown, J.S.; Gatenby, R.A. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nat. Commun.* **2017**, *8*, 1816. [\[CrossRef\]](#)
18. Fu, F.; Nowak, M.A.; Bonhoeffer, S. Spatial heterogeneity in drug concentrations can facilitate the emergence of resistance to cancer therapy. *PLoS Comput. Biol.* **2015**, *11*, e1004142. [\[CrossRef\]](#)
19. Martinez, V.A.; Laleh, N.; Salvioli, M.; Thuijsman, F.; Brown, J.; Cavill, R.; Kather, J.; Staňková, K. Improving mathematical models of cancer by including resistance to therapy: A study in non-small cell lung cancer. *bioRxiv* **2021**. [\[CrossRef\]](#)

20. Ghaffari Laleh, N.; Lo, W.; Tran, H.; Frieboes, H.B.; Cristini, V.; Lowengrub, J.S.; Enderling, H.; Scott, J.G. Classical mathematical models for prediction of response to chemotherapy and immunotherapy. *PLoS Comput. Biol.* **2022**, *18*, e1009822. [\[CrossRef\]](#)
21. Benzekry, S.; Lamont, C.; Beheshti, A.; Tracz, A.; Ebos, J.M.L.; Hlatky, L.; Hahnfeldt, P. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput. Biol.* **2014**, *10*, e1003800. [\[CrossRef\]](#)
22. Beal, L.D.; Hill, D.J.; Martin, R.; Hedengren, J.D. GEKKO optimization suite. *Processes* **2018**, *6*, 106. [\[CrossRef\]](#)
23. Tian, H.; Guo, Y.; Wang, X. Algorithm of Stackelberg game with multiple leaders followers based on evolutionary game theory. *J. Syst. Eng.* **2005**, *20*, 303.
24. Salvioli, M.; Dubbeldam, J.; Staňková, K.; Brown, J.S. Fisheries management as a Stackelberg Evolutionary Game: Finding an evolutionarily enlightened strategy. *PLoS ONE* **2021**, *16*, e0245255. [\[CrossRef\]](#)
25. Romano, C.; Di Benedetto, M.D.; Borri, A. Stackelberg Evolutionary Games With Modulated Leadership: A Three-Agent Framework for Tumor Immune Dynamics. *IEEE Control. Syst. Lett.* **2025**, *9*, 468–473. [\[CrossRef\]](#)
26. Gatenby, R.A.; Vincent, T.L. An evolutionary model of carcinogenesis. *Cancer Res.* **2003**, *63*, 6212–6220.
27. Basanta, D.; Hatzikirou, H.; Deutsch, A. Studying the emergence of invasiveness in tumours using game theory. *Eur. Phys. J. B* **2008**, *63*, 393–397. [\[CrossRef\]](#)
28. Archetti, M. Evolutionary dynamics of the Warburg effect: Glycolysis as a collective action problem among cancer cells. *J. Theor. Biol.* **2014**, *341*, 1–8. [\[CrossRef\]](#)
29. Kaznatcheev, A.; Peacock, J.; Basanta, D.; Marusyk, A.; Scott, J.G. Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nat. Ecol. Evol.* **2019**, *3*, 450–456. [\[CrossRef\]](#)
30. Romano, C.; Di Benedetto, M.D.; Borri, A. Evolution-Informed Modeling and Control of Tumor Growth. *IEEE Trans. Autom. Sci. Eng.* **2025**, early access. [\[CrossRef\]](#)
31. Marusyk, A.; Almendro, V.; Polyak, K. Intratumor heterogeneity: A looking glass for cancer. *Nat. Rev. Cancer* **2012**, *12*, 323–334. [\[CrossRef\]](#)
32. Gatenby, R.A.; Gillies, R.J. A microenvironmental model of carcinogenesis. *Nat. Rev. Cancer* **2008**, *8*, 56–61. [\[CrossRef\]](#)
33. Sartakhti, J.S.; Manshaei, M.H.; Sadeghi, M. MMP–TIMP interactions in cancer invasion: An evolutionary game-theoretical framework. *J. Theor. Biol.* **2017**, *412*, 17–26. [\[CrossRef\]](#)
34. Zeltz, C.; Primac, I.; Erusappan, P.; Alam, J.; Noel, A.; Gullberg, D. Cancer-associated fibroblasts in desmoplastic tumors: Emerging role of integrins. In *Seminars in Cancer Biology*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 62, pp. 166–181.
35. Sartakhti, J.S.; Manshaei, M.H.; Archetti, M. Game theory of tumor–stroma interactions in multiple myeloma: Effect of nonlinear benefits. *Games* **2018**, *9*, 32. [\[CrossRef\]](#)
36. Gatenby, R.A.; Brown, J.S.; Vincent, T.L. The evolution and ecology of resistance in cancer therapy. *Cold Spring Harb. Perspect. Med.* **2020**, *10*, a040972. [\[CrossRef\]](#)
37. Yang, C.-Y.; Shiranthika, C.; Wang, C.-Y.; Chen, K.-W.; Sumathipala, S. Reinforcement learning strategies in cancer chemotherapy treatments: A review. *Comput. Methods Programs Biomed.* **2023**, *229*, 107280. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Al-Lazikani, B.; Banerji, U.; Workman, P. Combination cancer therapy: A road map toward the next generation. *Nat. Rev. Cancer* **2017**, *17*, 293–299.
39. Greco, F.; Vicent, M.J. Combination strategies for cancer therapy: The next generation. *Ther. Deliv.* **2015**, *6*, 165–181.
40. Herrmann, R.; Bodoky, G.; Ruhstaller, T.; Glimelius, B.; Bajetta, E.; Schüller, J.; Saletti, P.; Bauer, J.; Figer, A.; Pestalozzi, B.; et al. Gemcitabine plus capecitabine compared with gemcitabine alone in advanced pancreatic cancer: A randomized, multicenter, phase III trial of the Swiss Group for Clinical Cancer Research (SAKK) and the Central European Cooperative Oncology Group (CECOG). *J. Clin. Oncol.* **2007**, *25*, 2212–2217. [\[CrossRef\]](#)
41. Scheithauer, W.; Schüller, J.; Ulrich-Pur, H.; Schmid, K.; Raderer, M.; Haider, K.; Kwasny, W.; Depisch, D.; Schneeweiss, B.; Lang, F.; et al. Randomised multicentre phase II trial of two different schedules of capecitabine plus gemcitabine in advanced pancreatic cancer. *Br. J. Cancer* **2003**, *89*, 597–602.
42. Geng, Z.; Wang, Y.; Wang, Z.; Wang, Y. Adaptive chemotherapy scheduling via reinforcement learning. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 1430–1441.
43. Komorowski, M.; Celi, L.A.; Badawi, O.; Gordon, A.C.; Faisal, A.A. Reinforcement learning for optimal control of cancer treatment. *PLoS Comput. Biol.* **2018**, *14*, e1005964.
44. Simeoni, M.; Magni, P.; Cammia, C.; De Nicolao, G.; Croci, V.; Pesenti, E.; Germani, M.; Poggese, I. Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Res.* **2004**, *64*, 1094–1101. [\[CrossRef\]](#)
45. Gerlowski, L.E.; Jain, R.K. Microvascular permeability of normal and neoplastic tissues. *Microvasc. Res.* **1986**, *31*, 288–305. [\[CrossRef\]](#)
46. Enriquez-Navas, P.M.; Kam, Y.; Das, S.; Hassan, S.; Silva, A.S.; Foroutan, P.; Ruiz, E.; Martinez, G.; Minton, S.; Gillies, R.J.; et al. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Sci. Transl. Med.* **2016**, *8*, 327ra24. [\[CrossRef\]](#) [\[PubMed\]](#)

47. Zhao, Y.; Wang, Z.; Li, Y.; Zhang, J. Reinforcement learning for personalized cancer treatment. *PLoS Comput. Biol.* **2020**, *16*, e1008322. [[CrossRef](#)]
48. West, J.; You, L.; Zhang, J.; Gatenby, R.A.; Brown, J.S.; Newton, P.K. Towards multidrug adaptive therapy. *Cancer Res.* **2020**, *80*, 1578–1589. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.