

Implementation of a Mathematical Model to Predict Radiotherapy Treatment Response of Non-Small Cell Lung Cancer Patients

R. Su^{1,2,*} and G. Pang^{3,4}

¹*Engineering Physics, The University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada*

²*Odette Cancer Centre, Ontario M4N 3M5, Canada*

³*Department of Physics, Ryerson University, Toronto, Ontario M5B 2k3, Canada*

⁴*Odette Cancer Centre, Department of Radiation Oncology,
University of Toronto, Ontario M4N 3M5, Canada*

(Dated: April 6, 2021)

We report on the development of a program that implements a mathematical model to describe radiation therapy treatment response of patients with non-small cell lung cancer proposed in an article by Geng et al. published in 2017. The program can predict the effect of well-separated radiation fractions on tumor growth and the Kaplan-Meier survival curve for an arbitrary patient population using a Monte Carlo method. While our program partially reproduces Geng et al.'s results well, further improvements to the source code and conceptual clarification is required.

I. INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths, and the most prevalent of all cancer types [1]. While the 5-year relative survival rate of non-small cell lung cancer patients has improved by 8.7% since 1975, it remains at 25.1% as of 2015 [2]. To increase the survival rate of cancer patients, clinicians can use mathematical modelling to optimize treatment scheduling and develop hypotheses for testing in clinical trials.

While many papers have proposed new mathematical models for lung cancer, few describe the implementation of the models in detail. The goal of this paper is to illustrate our process to implement a program in `Python` based on a predictive model for non-small lung cancer growth and treatment response described in an article by Geng et al. published in 2017 [3]. We also hope to provide some reasoning to the formulation of the mathematical model.

Non-small cell lung cancer accounts of 85% of all lung cancer cases [4]. It is less sensitive to chemotherapy and radiation therapy and takes longer to spread to other organs compared with small cell lung cancer [5].

Accurate cancer staging is important for both the clinician and the patient. Cancer staging is the process where clinician assigns a standardized metric to the state of growth and spread of the patient's tumor, which translates into the patient's prognostic information and eligibility for clinical trials. We will refer to the tumor, node, metastasis (TNM) staging system in this paper. The TNM system assesses the size of the primary tumor (T), the degree of lymph node involvement (N), and the presence or absence of metastasis (M) [6]. Once these three categories are assessed, a stage number of 0, I, II,

III, and IV in an order of increasing severity is assigned to the patient.

The preferred treatment method for lung cancer is surgical resection [4, 5]. The feasibility of surgical resection depends on whether the tumor remains at its primary site or has metastasized, and whether the patient can tolerate the surgery. Since patients are still at risk of relapse after surgery, additional chemotherapy and radiotherapy is used. For locally advanced unresectable tumors, which makes up 40% of non-small cell lung cancer patients [3], combined chemotherapy and radiotherapy is used to extend patient survival. This mixed combination of different treatment methods is known as adjuvant therapy.

Each patient responds to adjuvant therapy differently. If clinicians can accurately predict the effect of treatment for any patient, then they can develop treatment regimens that increase patient survival. Mathematical modelling hopes to fill this role.

II. METHODS

We encourage the reader to refer to the source code while reading this paper:

github.com/sillyPhotons/NSCLC_Modelling

A. Mathematical Modelling

Geng et al. proposed an ordinary differential equation to model non-small cell lung cancer growth that states the instantaneous rate of change in the number of living tumor cells is the sum of three terms: the rate of constrained tumor growth; the effect of radiotherapy; the effect of chemotherapy. Each term of the equation includes multiple parameters that needs to be estimated before the equation can be solved. In this work, we will disregard the chemotherapy portion of the equation.

While Geng et al. state their equation in terms of tumor cell number, N , we will state them in terms of

*On undergraduate co-op placement at Odette Cancer Centre, supervised by G. Pang from January 3 to May 1, 2020. Correspondance can be addressed to R. Su at ruihengsu@alumni.ubc.ca

tumor volume, V . The two formulations differ only by a factor of tumor cell density.

1. Constrained Tumor Growth

The first term in the equation proposed describes constrained tumor growth.

Let $V(t)$ be the tumor volume a patient at time t after diagnosis, ρ be a patient specific tumor growth rate parameter, and K be the maximum tumor volume known as the carrying capacity. Geng et al. expects untreated tumor growth to follow Gompertzian kinetics - tumor growth rate decreases exponentially with time.

$$\frac{dV(t)}{dt} = \rho V(t) \log \left(\frac{K}{V(t)} \right) \quad \text{Tumor Growth} \quad (1)$$

Given an initial tumor diameter, we can make the assumption of a spherical tumor to state the initial value problem,

$$\frac{dV(t)}{dt} = \rho V(t) \log \left(\frac{K}{V(t)} \right) \quad V(t_0) = V_0$$

Using the separation of variables,

$$\int \frac{1}{V(t)} \left[\log \left(\frac{K}{V(t)} \right) \right]^{-1} dV(t) = \int \rho dt + C$$

$$u = \log \left(\frac{K}{V(t)} \right) \quad du = \frac{-1}{V} dV(t) \quad \text{Apply } u\text{-substitution}$$

We obtain the analytic solution to equation (1).

$$V(t) = K \exp \left[-\log \left(\frac{K}{V_0} \right) e^{-\rho t} \right] \quad (2)$$

Examining equation (2), in the limit as t approaches infinity, the tumor volume, V , approaches the carrying capacity, K . This describes the clinically observed pattern of decreasing growth rate with increasing tumor size [3, 7, 8]. Setting the second derivative of equation (2) to zero, we find that y-coordinate of the inflection point is precisely K/e , or 36.8% of the carrying capacity. This makes equation (2) suitable for modelling processes that start to grow slower upon reaching approximately 1/3 of its maximum size as seen in figure 1.

2. Radiation Therapy Effect

The second term in equation proposed describes the effect of radiotherapy on tumor growth.

Radiation therapy, or radiotherapy, is the use of focused ionizing radiation to kill tumor cells. High energy radiation causes the excitation of electrons at the tumor site through the photoelectric effect, Compton scattering, and pair production; producing free radicals that react with nearby molecules to induce chemical damage. The

primary contribution to cell death is irreparable damage to DNA strands.

Historically, radiotherapy for non-small cell lung cancer is usually comprised of a total dose 60 to 70 Gy delivered using conventional fractionation - 1.8 to 2.0 Gy fractions per day [5]. Other fractionation schemes of clinical interest include hyperfractionation and accelerated fractionation [3, 5]. Hyperfractionation involves the patient receiving a higher daily dose fraction, taking a shorter time to deliver the prescribed dose compared to conventional fractionation. Accelerated fractionation involves the patient receiving the same dose fraction as conventional fractionation, but more than once daily.

The purpose of delivering radiation in separated fractions is to minimize harm to patients by allowing healthy cells time to repair. But having too much time between fractions can cause tumor cells to repopulate, decreasing the effectiveness of radiotherapy and increasing total dose needed to achieve the intended effect.

The most widely used model for radiation effect modelling is the linear-quadratic model [3, 7, 9, 10], which is well validated from 1 to 10 Gy per fraction and can be used up to 18 Gy per fraction [10]. Below 1 Gy per fraction, some cells *in vitro* and *in vivo* display increased sensitivity to radiation (radiosensitivity), deviating from the linear-quadratic model [9].

Geng et al. used the linear-quadratic model to predict the effects radiation fractions to tumor cell growth, assuming all fractions are well-separated and instantaneously delivered. While the general form of the equation includes the Lea-Catcheside time factor in the quadratic dose term, the factor reduces to 1 under this assumption [7, 10].

Let $\delta(t)$ be the Dirac delta function, t_R be the time of radiation, α be a patient specific radiosensitivity param-

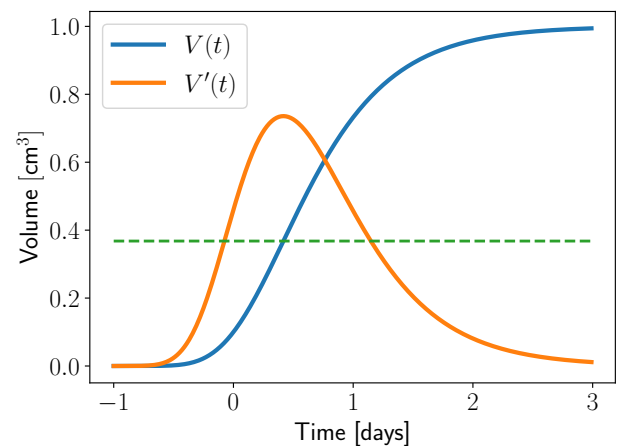


FIG. 1: Example plot of equation (2) and its derivative with $K = 1$, $\rho = 0.1$, $V_0 = 0.1$. We observe that the inflection point occurs at $y = 1/e$ and $V(t)$ is asymmetric about its inflection point.

eter with units Gy^{-1} , β be a constant with units Gy^{-2} , and d be a constant representing the dose per fraction. The radiotherapy term is given by

$$\frac{dV(t)}{dt} = -\delta(t - t_R)V(t)(\alpha d + \beta d^2) \quad \text{Radiotherapy} \quad (3)$$

Geng et al. assumes that $\alpha/\beta = 10 \text{ Gy}$ for tumor tissues.

Geng et al.'s original article did not contain the Dirac delta function, $\delta(t - t_R)$ with units days^{-1} ; equation (3) is the formulation given in a later correction [11].

3. Discrete-Time Model

The sum of equation (1) and equation (3) results in Geng et al.'s model for the instantaneous rate of change in tumor volume of a non-small cell lung cancer patient treated with radiation therapy.

$$\frac{dV(t)}{dt} = \rho V(t) \log\left(\frac{K}{V(t)}\right) - \delta(t - t_R)V(t)(\alpha d + \beta d^2) \quad (4)$$

While Geng et al. solved equation (4) in `Python` using the `odeint` function of the `SciPy` library [11], we solved equation (4) via time stepping using discrete-time models.

But there is one problem. Equation (4) cannot be evaluated in its current form due to the Dirac delta function, which is undefined at the point of impulse, t_R .

We propose a formulation without the delta function. For n dose fractions delivered at times t_1, \dots, t_n , let δ_{t,t_R} be the Kronecker delta function, where

$$\delta_{t,t_R} = \begin{cases} 1 & \text{when } t = t_R \\ 0 & \text{when } t \neq t_R \end{cases}$$

We define a function, $D(t)$, where

$$D(t) = d \sum_{R=1}^n \delta_{t,t_R} \quad (5)$$

We will use the following redefinition of equation (4) for the remainder of this paper.

$$\frac{dV(t)}{dt} = \rho V(t) \log\left(\frac{K}{V(t)}\right) - V(t) (\alpha D(t) + \beta D(t)^2) \quad (6)$$

Given the initial tumor volume, V_0 , and the values of ρ, K, α, β , and d , the simplest discrete-time mode we can formulate is based on the forward Euler's method, relying on successive linear approximations. Let $dV(t)/dt = f(t, V)$, and Δt be the step size, the model is given by

$$V_{n+1} = f(t_n, V_n)\Delta t + V_n \quad (7)$$

The forward Euler's method model only requires a single function evaluation per time step. But it lacks in

accuracy for non-smooth functions or when the step size Δt is too large. It can be shown that the forward Euler's method is a first order method - halving step size reduces the error by a factor of 2. Figure 2A visualizes a single time step using the forward Euler's method.

We implemented another model based on the fourth order Runge-Kutta method, which offers better accuracy (halving step size reduces error by factor of 2^4) at the price of requiring four function evaluations per time step. Let $dV/dt = f(t, V)$, then

$$\begin{aligned} S_1 &= f(t_n, V_n) \\ S_2 &= f(t_n + \frac{\Delta t}{2}, V_n + \Delta t \frac{S_1}{2}) \\ S_3 &= f(t_n + \frac{\Delta t}{2}, V_n + \Delta t \frac{S_2}{2}) \\ S_4 &= f(t_n + \Delta t, V_n + \Delta t S_3) \\ V_{n+1} &= V_n + \frac{\Delta t}{6}(S_1 + 2S_2 + 2S_3 + S_4) \end{aligned} \quad (8)$$

We visualized a single time step of this model in figure 2B. We can see the difference in accuracy between the forward Euler's method model and the Runge-Kutta method model by comparing A and B of figure 2.

The final discrete-time model we implemented was used by Geng et al., involving an exponential [11].

$$V_{n+1} = V_n e^{\rho \Delta t \log\left(\frac{K}{V_n}\right) - (\alpha D(t) + \beta D(t)^2)} \quad (9)$$

B. Parameter Estimation

Since the parameters ρ and α are patient specific - we expect their values to vary from patient to patient - a reasonable assumption to make is that α and ρ value are normally distributed across all patients.

To acquire values for ρ and α requires estimating the mean and standard deviation of their normal distributions. But as cells with faster growth rate have shown to be more sensitive to the harming effects of ionizing radiation [3], a linear correlation coefficient, ψ , between the distributions of α and ρ can be introduced. Let α_μ, ρ_μ be the mean and $\alpha_\sigma, \rho_\sigma$, be the standard deviation of α and ρ distributions. By definition, ψ is the ratio of the covariance of α and ρ and the product of α_σ and ρ_σ .

To draw correlated (α, ρ) samples, we pass an array (α_μ, ρ_μ) , and a covariance matrix \mathbf{C} , where

$$\mathbf{C} = \begin{bmatrix} \alpha_\sigma^2 & (\alpha_\sigma \rho_\sigma) / \psi \\ (\alpha_\sigma \rho_\sigma) / \psi & \rho_\sigma^2 \end{bmatrix}$$

to the `multivariate.normal` function of the `Numpy` library in `Python`.

Similarly, we can interpret the initial tumor diameter for a patient to be a sample from the diameter distribution of the stage that the patient belongs in. From clinical observations, Geng et al. assume these distribution are log-normal distributions.

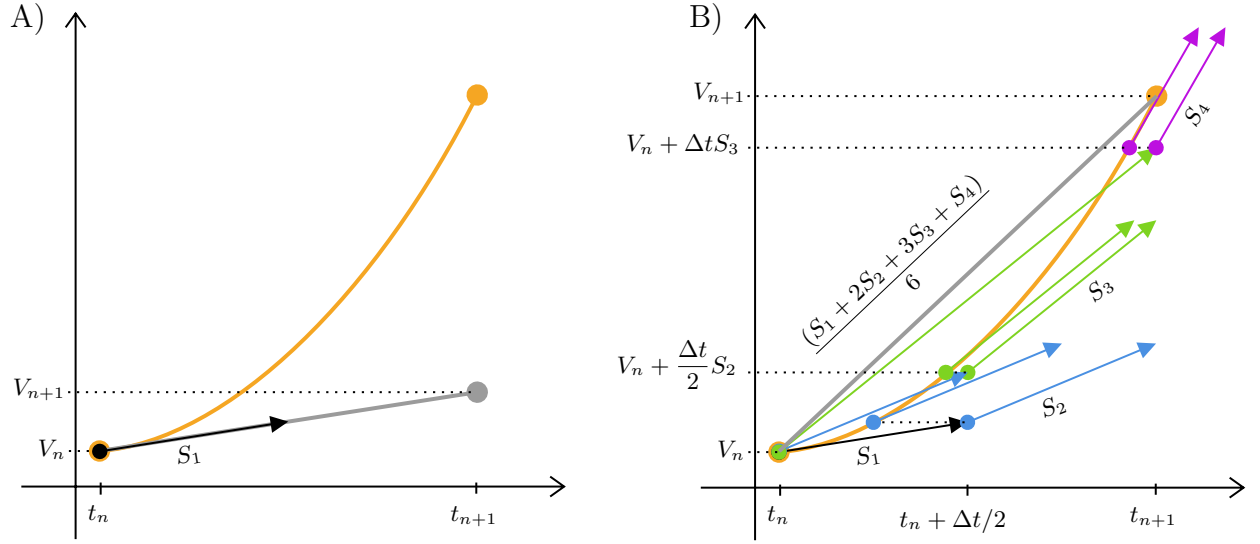


FIG. 2: A): One time step of the forward Euler's method. B): One time step of the fourth order Runge-Kutta method. The grey line has a slope equal to the weighted average of S_1, S_2, S_3, S_4 . The black arrow has slope S_1 . The blue arrows have slope S_2 . The green arrows have slope S_3 . The purple arrows have slope S_4 . A single step using the fourth order Runge-Kutta method can be much more accurate than the forward Euler's method, but requires more function evaluations.

| | Name | Variable | Distribution | Constraints | Units |
|--------------|--|--|--------------|---------------------------------------|--------------------|
| Optimized | Stage I volume distribution | $V_0^I : V_\mu^I, V_\sigma^I$ | Lognormal | $> 0.3 \text{ cm}, < 5.0 \text{ cm}$ | cm^3 |
| | Stage II volume distribution | $V_0^{II} : V_\mu^{II}, V_\sigma^{II}$ | | $> 0.3 \text{ cm}, < 13.0 \text{ cm}$ | |
| | Stage IIIA volume distribution | $V_0^{IIIA} : V_\mu^{IIIA}, V_\sigma^{IIIA}$ | | $> 0.3 \text{ cm}, < 13.0 \text{ cm}$ | |
| | Stage IIIB volume distribution | $V_0^{IIIB} : V_\mu^{IIIB}, V_\sigma^{IIIB}$ | | $> 0.3 \text{ cm}, < 13.0 \text{ cm}$ | |
| | Stage IV volume distribution | $V_0^{IV} : V_\mu^{IV}, V_\sigma^{IV}$ | | $> 0.3 \text{ cm}, < 13.0 \text{ cm}$ | |
| | Gompertz growth rate | $\rho : \rho_\mu, \rho_\sigma$ | Normal | $> 0 \text{ days}^{-1}$ | days^{-1} |
| | Radiosensitivity | $\alpha : \alpha_\mu, \alpha_\sigma$ | | $> 0 \text{ Gy}^{-1}$ | Gy^{-1} |
| | Carrying capacity | K | Uniform | $> 0 \text{ cm}$ | cm^3 |
| | α, ρ correlation coefficient | ψ | | > 0 | 1 |
| User-Defined | α / β ratio | α / β | Constant | $= 10 \text{ Gy}$ | Gy |
| | Dose per fraction | d | | $> 0 \text{ Gy}$ | |
| | Total dose | | | $> 0 \text{ Gy}$ | |
| | Treatment days | | | $> 0 \text{ days}$ | days |
| | Rest days | | | $\geq 0 \text{ days}$ | |
| | Death diameter | | | $= 13 \text{ cm}$ | cm |

TABLE I: Summary of required parameters to solve the tumor growth model radiotherapy treatment effect for a single patient. The constraints for the volume distributions are given in terms of diameter. In the actual implementation, samples of a tumor diameter distribution is converted to volume by assuming a spherical tumor. Geng et al. assume the minimum detectable tumor diameter is 0.3 cm.

To obtain a value for the initial tumor diameter for an arbitrary patient in a specific cancer stage then requires estimating the diameter distribution for that stage. A tumor volume can be specified by the tumor diameter, assuming a spherical tumor. We specified the log-normal

tumor volume distribution for each cancer stage by the mean, $V_\mu^I, \dots, V_\mu^{IV}$, and the standard deviation, $V_\sigma^I, \dots, V_\sigma^{IV}$ its underlying normal distribution in terms of tumor diameter.

The parameter estimation process involves fitting clin-

ical data to obtain estimates for ρ_μ , ρ_σ , K , α_μ , α_σ , ψ , $V_\mu^{I,\dots,IV}$, and $V_\sigma^{I,\dots,IV}$ for a total of 16 parameters.

Geng et al. fit Kaplan-Meier survival curves generated using a Monte Carlo method based on equation (6) to clinical survival curves obtained from the California Cancer Registry [12], and the RTOG-8808 clinical trial [13].

The California Cancer Registry data set contains survival curves of 22945 (Geng et al. mistakenly cites 23954 patients) untreated stage I, II, IIIA, IIIB, and IV non-small cell lung cancer patients. The RTOG-8808 trial data set contains the survival curve of a population of 9 stage I, 67 stage IIIA, and 76 stage IIIB patients that received conventional radiotherapy with 2 Gy fractions for 5 days a week to 60 Gy. But since these data sets were inaccessible during this project, we used an online data extraction tool (WebPlotDigitizer) [14] to estimate the data - spanning 59 months after diagnosis - from figures 3 and 5 from Geng et al.'s paper instead. The survival curves from the two data sets are shown in figures 3 and 4.

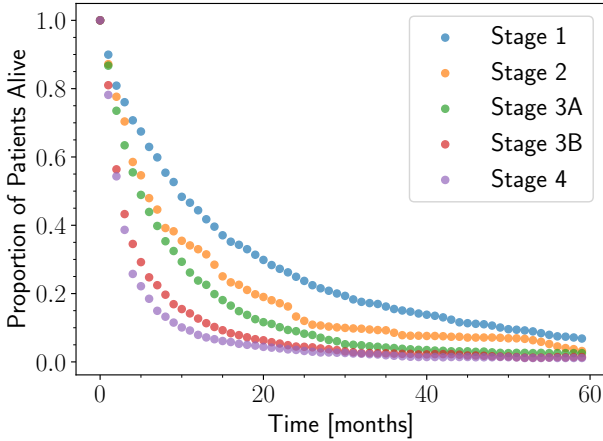


FIG. 3: Plot of the survival curves for untreated non-small cell lung cancer patients estimated from figure 3 of Geng et al.'s paper. We can see that the initial slopes of the curve increases in magnitude with cancer stage.

1. Parameter Optimization

Since a 16-dimensional parameter space cannot be efficiently searched, Geng et al. proposed a multi-stepped estimation process, as shown in figure 5.

Geng et al. define the cost function for minimization as the of sum of the variance between a survival curve from clinical data and a survival curve generated from our discrete-time model using a Monte-Carlo method. We used the `minimize` function from the `lmfit` library [15] to automate the minimization process shown in figure 6A using Powell's method.

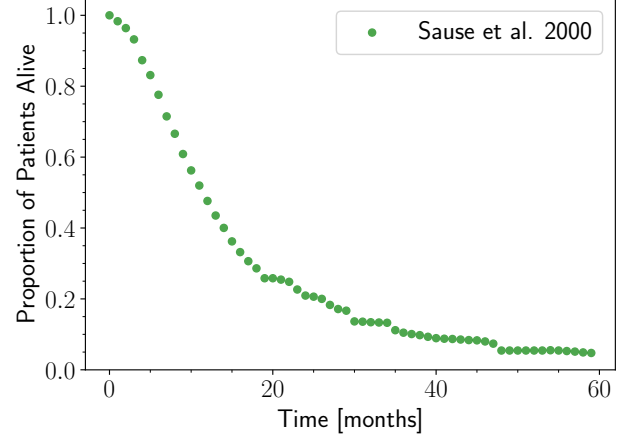


FIG. 4: Plot of the survival curve for non-small cell lung cancer patients from the standard radiotherapy arm in the RTOG-8808 clinical trial estimated from figure 5 of Geng et al.'s paper.

In **Python**, we define our cost function to take: a variable number of parameters; x - and y -coordinates arrays of the survival curve data; a function object that implements one of three discrete-time model described (functions are objects in **Python**). As opposed to returning a sum of residuals as done by Geng et al., our cost function returns an array of residuals between the model and data to be minimized, as required by the `lmfit` documentation. Let \mathbf{SF} be a 1 by n vector, where n corresponds to the number of time steps between the initial and final time of the simulated time interval, and the i^{th} element represents the proportion of patients alive after i time steps. Our cost function maps the values of a set of parameter values to a vector of residuals between \mathbf{SF} of the model and data.

$$\mathbf{SF}_{\text{model}} - \mathbf{SF}_{\text{data}} \quad (10)$$

2. Monte Carlo Patient Population

Figure 6A shows that parameter optimization requires generating a Monte Carlo patient population to replicate the initial tumor characteristics of patients in the data. The parameters we generate for each patient is summarized in table I.

3. Kaplan-Meier Survival Curve Generation

The procedure to generate a Kaplan-Meier survival curve is shown in figure 6B. After generating Monte Carlo patient population and specifying magnitude of each time step in units of days, we use a discrete-time model to observe the tumor growth for each patient in the population. Following each time step, we check whether the

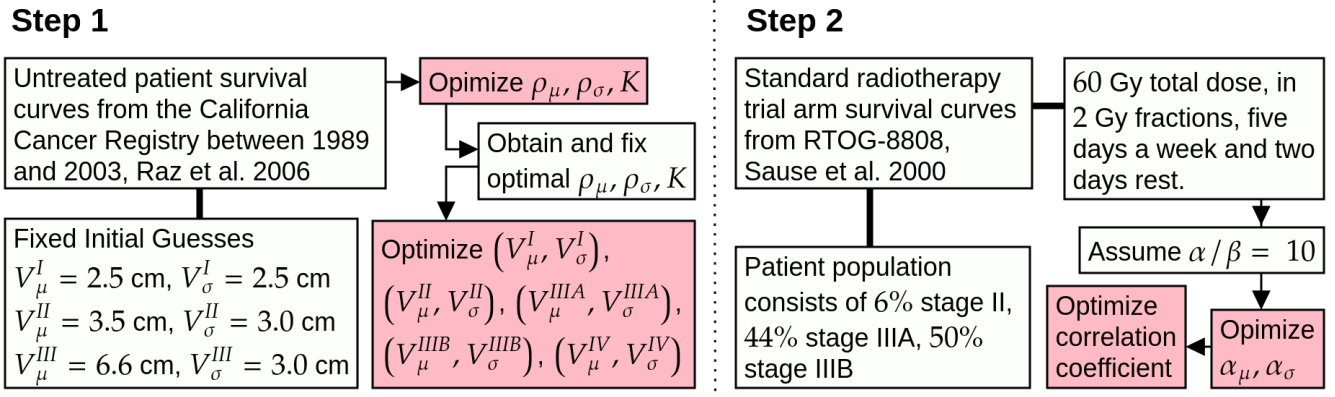


FIG. 5: In step 1, Geng et al. optimized the tumor growth parameters using survival curves of untreated patients and predefined volume distributions for each stage. Geng et al. then optimized the volume distributions for each stage using the optimized tumor growth parameters. In step 2, Geng et al. optimized the radiosensitivity parameter, α , by replicating the patient population and radiotherapy fractionation used by the standard radiotherapy arm of the RTOG-8808 trial, and assuming that $\alpha/\beta = 10$. Geng et al. optimized the linear correlation coefficient between α and ρ last, holding all optimized parameters constant.

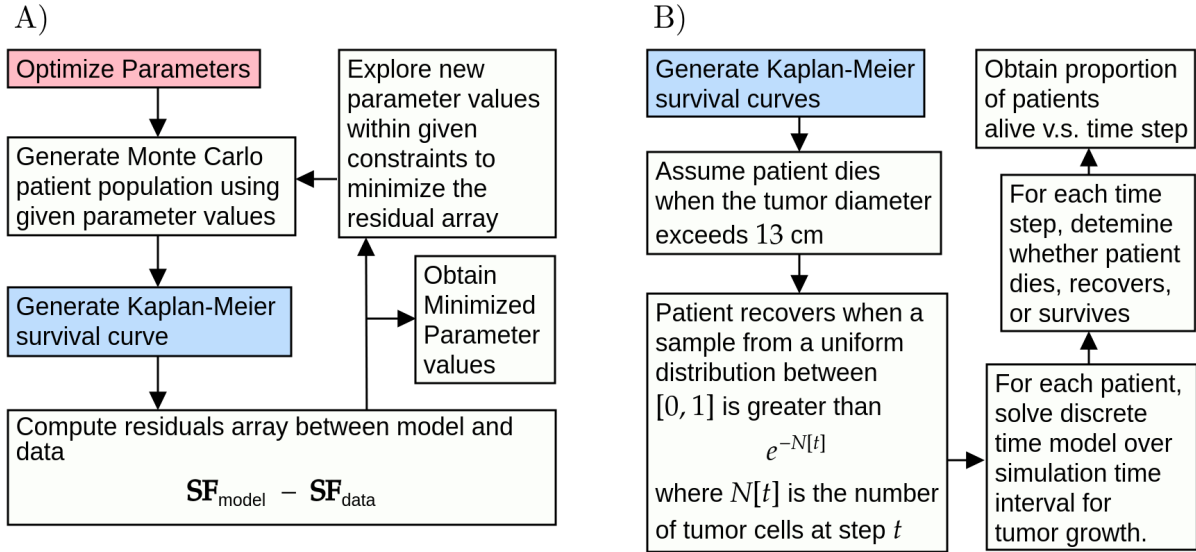


FIG. 6: A): The parameter optimization process. The `minimize` function recursively explores the parameter space using Powell's method. B): The Kaplan-Meier survival curve generation process. The key assumption of is that the patient is dies when the tumor diameter exceeds 13 cm. While Geng et al. defines the death condition in terms of diameter, the program converts this to tumor volume, assuming a spherical tumor.

current tumor diameter exceeds the death diameter followed by whether the patient recovers.

For consistency with Geng et al.'s work, we used a step size of $\Delta t = 0.01$ days, and 10000 patients per Monte Carlo population [11]. But such a fine step size and large number of patients can cause a significant increase in run-time for a sequential implementation, as the discrete time model must be evolved over the simulation interval for every patient.

Our implementation highlights the power of parallel programming to reduce run-time. There are numerous options for parallelism in Python (e.g. `threading`,

`multiprocessing`), but most are syntactically difficult to use.

The `ray` library [16] offers a simple and scalable method to enable python functions to be run in parallel. Using `ray`, our program can solve the discrete-time model for multiple patients at a time during in each parameter optimization iteration.

| | Name | Distribution | Mean | Standard Deviation | Units | Source |
|-------------------|--|-------------------|------------------------------------|-----------------------------|--------------------|-------------------|
| Tumor Growth | Stage I volume distribution | Underlying Normal | $V_{\mu}^I = 0.207$ | $V_{\sigma}^I = 0.774$ | cm | Table 2 |
| | Stage II volume distribution | | $V_{\mu}^{II} = 1.261$ | $V_{\sigma}^{II} = 0.694$ | | |
| | Stage IIIA volume distribution | | $V_{\mu}^{IIIA} = 1.621$ | $V_{\sigma}^{IIIA} = 0.462$ | | |
| | Stage IIIB volume distribution | | $V_{\mu}^{IIIB} = 2.168$ | $V_{\sigma}^{IIIB} = 0.215$ | | |
| | Stage IV volume distribution | | $V_{\mu}^{IV} = 2.270$ | $V_{\sigma}^{IV} = 0.298$ | | |
| | Gompertz growth rate | Normal | $\rho_{\mu} = 7.00 \times 10^{-5}$ | $\rho_{\sigma} = 0.00723$ | days ⁻¹ | Table 3 |
| | Carrying capacity | Constant | $K = 30$ | | cm | |
| | Death diameter | | 13 | | | |
| Radiation Therapy | Radiosensitivity | Normal | $\alpha_{\mu} = 0.0398$ | $\alpha_{\sigma} = 0.168$ | Gy ⁻¹ | Page 1 |
| | α, ρ correlation coefficient | Constant | $\psi = 0.87$ | | 1 | |
| | α / β ratio | | $\alpha / \beta = 10$ | | | Sause et al. 2000 |
| | Dose per fraction | | $d = 2$ | | Gy | |
| | Total dose | | 60 | | | |
| | Treatment days | | 5 | | days | |
| | Rest days | 2 | | | | |

TABLE II: Summary of the numerical parameter values used to reproduce Geng et al.’s results. We source these values from: tables 3 and 2, and page 1 of Geng et al.’s paper; and the description of the standard radiotherapy arm in Sause et al.’s paper on the RTOG-8808 trial.

C. Code Verification

We applied test-driven development to minimize syntactic and conceptual errors. We ensured that the specifications of each function or class is met prior to run-time through unit tests written using the `unittest` library.

We reasoned that our program is conceptually correct if the following conditions were met: parameters estimated using reported parameter values in Geng et al.’s paper as initial guesses, with a step size > 0.01 days and at least 1000 patients per Monte Carlo population does not deviate from the initial value by over 15%; replications of figures 3, 4A, 5B, 6B in Geng et al.’s paper have negligible difference upon visual inspection.

To reproduce Geng et al.’s results, we used the parameter values reported in tables 2 and 3 of their paper for our program: table 2 gives the optimized mean and median of the log-normal volume distributions for each cancer stage in columns 3 and 4; table 3 gives the patient death diameter and optimized $\rho_\mu, \rho_\sigma, \alpha_\mu, \alpha_\sigma, K$, and ψ . Table 3 also provides some numerical values on the volume distributions of each stage. But the exact meaning of these values were not clearly stated. We found that survival curves generated by assuming that these values referred to $V_\mu^{I, \dots, IV}, V_\sigma^{I, \dots, IV}$ did not match the data for untreated patients. We believe values given table 1 and 3 refers to the underlying normal distribution while table 2 refers to actual log-normal distribution - this remains to be verified.

The mean and median of a log-normal are related to the mean and standard deviation of its underlying normal distribution (μ, σ) by the following equalities:

$$\begin{aligned}\mu &= \log(\text{median}) \\ \sigma &= \sqrt{2|\log(\text{mean}) - \mu|}\end{aligned}$$

We summarize the numerical values used for prediction in table II.

D. Model Prediction

After verifying the correctness of our program, we can use it to predict the Kaplan-Meier survival curve of a non-small cell lung cancer population made of stage I, II, IIIA, IIIB, and IV patients in user-defined proportions, or to simulate the effect of well-separated radiation fractions on tumor growth for a single patient.

III. RESULTS

We verify the ability of our model to estimate parameter values, reproduce plots, and predict radiotherapy treatment response.

A. Parameter Estimation

We were unable to follow the complete parameter estimation process outlined in figure 5 when each optimization is configured to use a 10000 patient population and a step size $\Delta t = 0.01$ days. This is because of a lack of virtual memory in the quad-core computer used, as the current program creates multiple large 2-dimensional arrays whose number of elements is proportional to the number of patients and inversely proportional to the step size. But even if the lack of memory is not an issue, we expect a lengthy run-time as the device used can only handle four tasks in parallel.

We argued that our model's ability to estimate parameters can be demonstrated if setting initial guesses to reported values does not deviate from the reported value by more than 15% for a simulated population of ≤ 10000 patients and a step size > 0.01 days.

Setting the number of Monte Carlo patients to 1000, and the step size of $\Delta t = 1$ day, we optimized the following using reported optimal values given table II: the initial tumor volume distributions; α with $\psi = 0$; ψ .

| Parameter | Value | Percent Difference |
|---------------------------------|------------|---------------------------|
| V_μ^I, V_σ^I | 0.59, 4.22 | $\leq 96\%$, 138% |
| $V_\mu^{II}, V_\sigma^{II}$ | 1.21, 2.52 | $\leq 4\%$, $\leq 114\%$ |
| $V_\mu^{IIIA}, V_\sigma^{IIIA}$ | 2.86, 1.35 | $\leq 55\%$, $\leq 98\%$ |
| $V_\mu^{IIIB}, V_\sigma^{IIIB}$ | 2.22, 0.33 | $\leq 3\%$, $\leq 8\%$ |
| $V_\mu^{IV}, V_\sigma^{IV}$ | 2.27, 0.30 | $\leq 1\%$, $\leq 1\%$ |
| $\alpha_\mu, \alpha_\sigma$ | 0.03, 0.15 | $\leq 26\%$, $\leq 14\%$ |
| ψ | 0.74 | $\leq 16\%$ |

TABLE III: The mean values of the estimated parameters. We did not report the uncertainties in these values as we intend for this to serve as a first demonstration of our program to estimate parameter values.

We find that the mean and standard deviation of the underlying normal distribution for the volume distributions of stage I, II, IIIA exceeds the 15% limit proposed. We believe that this is because of both the error in data extraction, and that the initial values of $V_\mu^I, V_\sigma^{IIIB}, V_\sigma^{IV}$ seen in table II were less than 0.3 cm, the minimum detectable tumor diameter.

This raises whether our method of volume sampling is correct. A sign of correctness is a good match between a histogram produced from samples of the stage I volume distribution specified in table II and reference histogram in figure 4A of Geng et al.'s paper.

We find our histogram of 200000 samples from the stage I volume distribution with constraints $0.3 < V < 5.0$ has less entries in the first bin and more entries in the second bin than the reference histogram. We also

find that if we sample the same stage 1 distribution using a constraint of $0 < V < 0.5$, the 1 first bin increases in the number of entries, but the rest of the bins remain relatively unchanged. We show this in figure 7.

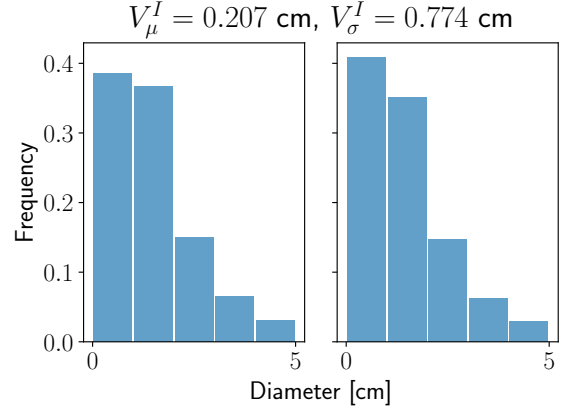


FIG. 7: Histogram of samples from the log-normal initial volume distribution for stage I patients. The histogram on the left is constrained to > 0.3 cm, whereas the histogram on the right is constrained to > 0 cm.

B. Model Reproduction

Our program performed well in reproducing figures 3, 5B, and 6B of Geng et al.'s paper.

For the generation of Kaplan-Meier survival curves, we found that the choice of discrete-time model causes no visible difference in the shape and fit to data for a sufficiently large Monte Carlo population and a step size of $\Delta t = 1$ day.

Figures 8, 9, 10, 11, and 12 show the survival curves of untreated non-small cell lung cancer generated using the optimal parameter values Geng et al. reported with survival curves. They correspond to figure 3 in Geng et al.'s paper.

Figure 13 shows the survival curves generated using reported optimal parameters with and without incorporating a correlation between tumor radiosensitivity and growth rate. Our model replicates figure 5B of Geng et al.'s paper well despite minor differences, which we attribute to inaccuracies from data extraction.

We simulated the response of a patient 217 days after diagnosis and standard fractionation radiotherapy after a 14 day delay, using the following parameter values: $V_0 = 8$ cm, $\rho = 0.008$ days $^{-1}$, $K = 30$ cm, $\alpha = 0.3$ Gy $^{-1}$, $\beta = 0.03$ Gy $^{-2}$. Geng et al. also used these values, in addition to chemotherapy parameters to predict patient response to sequential vs concurrent chemotherapy and radiotherapy in figure 8 of their paper.

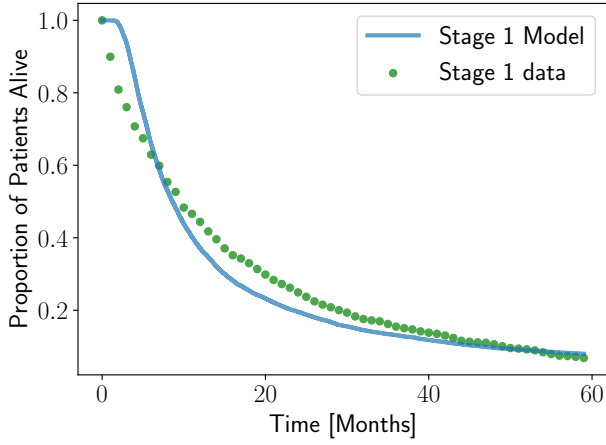


FIG. 8: Plot of untreated stage 1 patient survival curve and model survival curve.

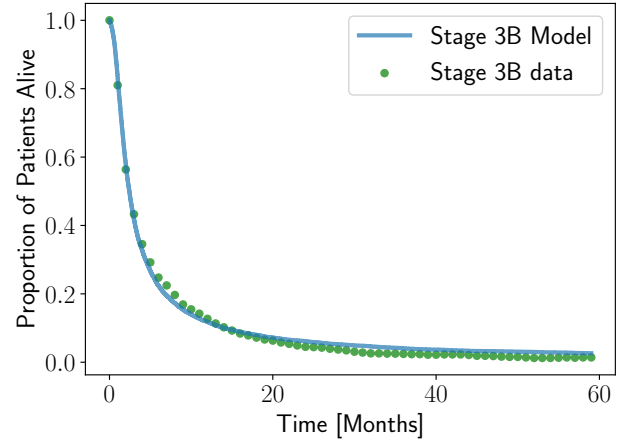


FIG. 11: Plot of untreated stage 3B patient survival curve and model survival curve.

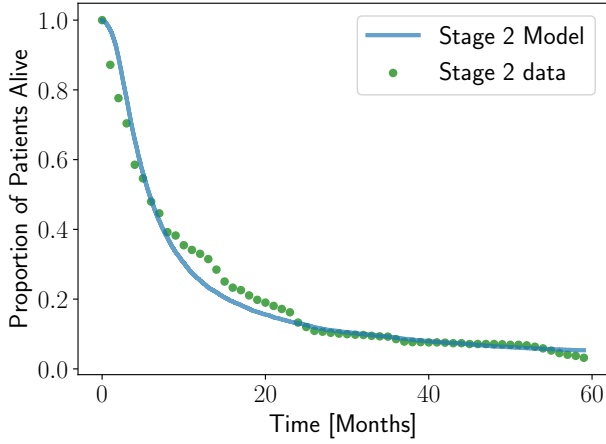


FIG. 9: Plot of untreated stage 2 patient survival curve and model survival curve.

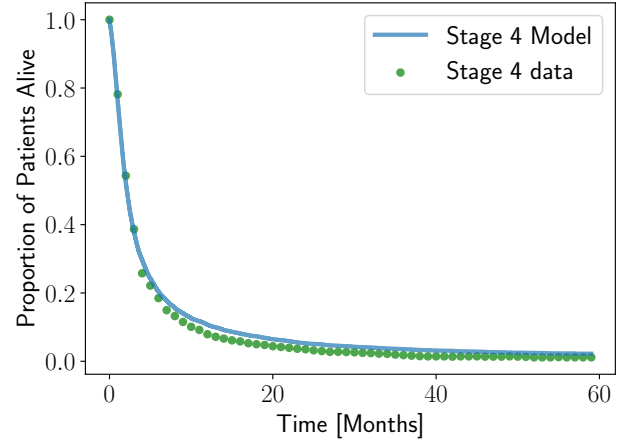


FIG. 12: Plot of untreated stage 4 patient survival curve and model survival curve.

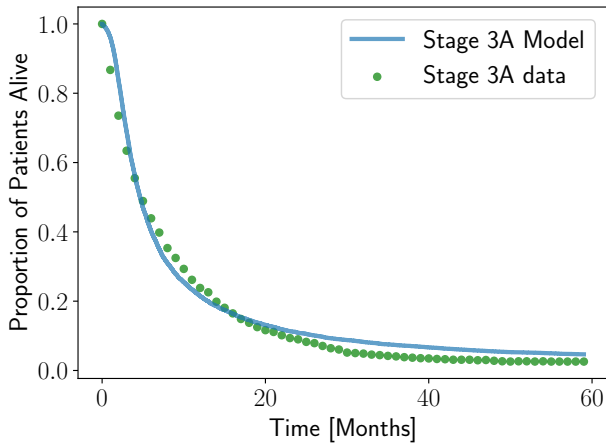


FIG. 10: Plot of untreated stage 3A patient survival curve and model survival curve.

We plot the response predicted by the three discrete-time models we described at three different step sizes in figures 14, 15, and 16. When $\Delta t = 1$, we see that the forward Euler's method model predicts a much stronger radiotherapy response than the other two models. We expected this as we have shown that the Euler's method model is a less accurate model than the 4th order Runge-Kutta method model. For a smaller step size, we find that the response curve of Euler's method model gets closer to the curve of Geng et al.'s model, yet the Runge-Kutta method model curve moves away from Geng et al.'s model, predicting a much weaker response. This contradicts with our expectation that for a finer step size, all three will predict a common response curve. Out of all three models, Geng et al.'s model was the most consistent.

Interestingly, the untreated growth curves that the

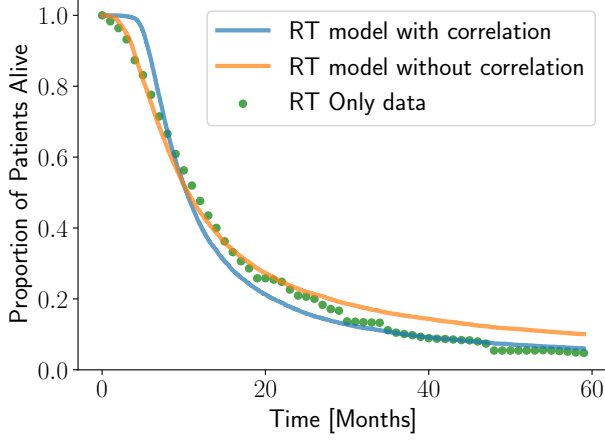


FIG. 13: Plot of the RTOG-8808 standard radiotherapy trial arm survival curve with survival curves generated with and without a correlation between α and ρ distributions. The plot shows a better fit of the model with correlation to data for larger times, as does Geng et al.'s plot.

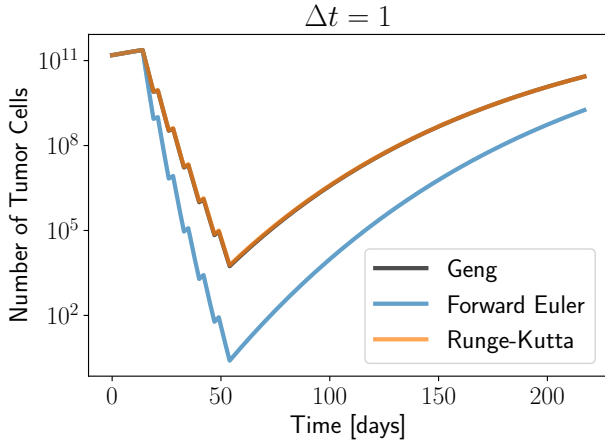


FIG. 14: Patient response predicted by three different discrete-time models at $\Delta t = 1$ days. Both Geng et al.'s and the 4th order Runge-Kutta method model predict the same response whereas the forward Euler's method predicts a stronger radiotherapy response.

three models predict were relatively invariant to step size, suggesting that the difference in treatment response seen is primarily due to the radiotherapy term.

C. Model Prediction

To show the ability of our program to predict treatment responses, we simulated the effects of a 60 Gy total, three 2 Gy fractions per week scheme and a 120 Gy total, five 2 Gy fractions per week scheme for the same patient configuration as in figure 14, 15, and 16, and compare it

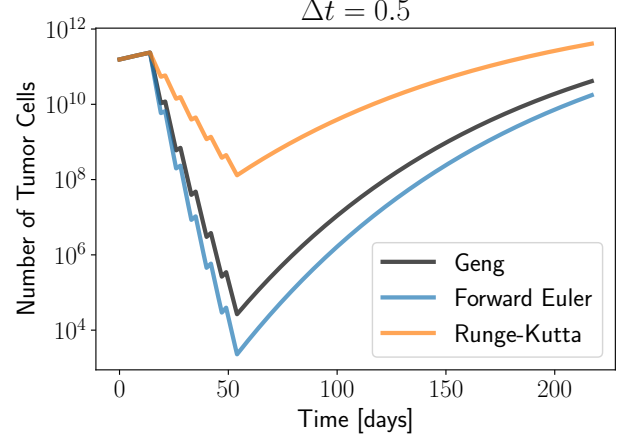


FIG. 15: Patient response predicted by three different discrete-time models at $\Delta t = 0.5$ days. Geng et al.'s and the forward Euler's method model predicts a similar response, whereas the 4th order Runge-Kutta method model predicts a much weaker radiotherapy response.

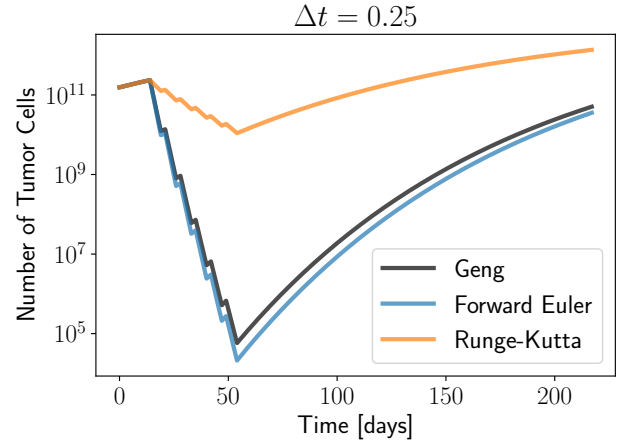


FIG. 16: Patient response predicted by three different discrete-time models at $\Delta t = 0.25$ days. Geng et al.'s and the forward Euler's method model predicts a similar response, whereas the 4th order Runge-Kutta method model predicts a much weaker radiotherapy response.

to the standard 60 Gy total, five 2 Gy fraction per week scheme.

Using Geng et al.'s model and $\Delta t = 1$ days, we plot the results in figure 17.

We find that the (3, 2, 60) scheme resulted in a lower tumor cell count than (60, 5, 2) scheme at day 200, but the (60, 5, 2) scheme takes shorter time to complete and is more effective for taking the tumor cell number to lower count quickly. The (120, 5, 2) scheme resulted in patient recovery as seen in the termination of the curve.

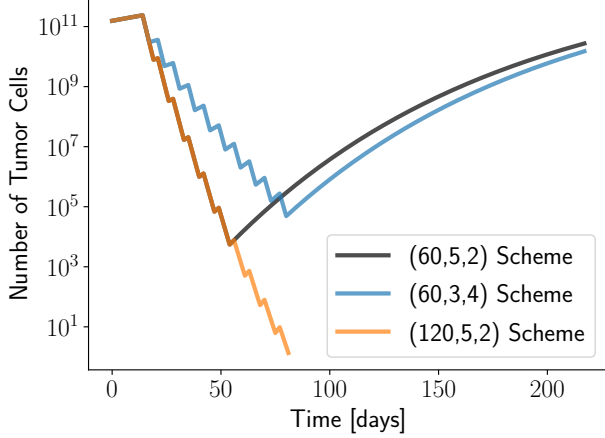


FIG. 17: The treatment response of a patient to three different fractionation schemes. The ordered triple in the legend describes the scheme used to each response curve corresponding to total dose, consecutive days of 2 Gy fractions, and consecutive days of rest.

IV. DISCUSSION

We have illustrated of how to implement the mathematical model proposed by Geng et al. in `Python`, from parameter estimation to predicting radiotherapy treatment response. But our program requires further development and verification both programmatically and conceptually.

The combinations of step size and number of patients per Monte-Carlo population that can be used with the current program is limited by the virtual memory available in the computer used. This signals the need to improve the source code, perhaps by creating multiple smaller arrays to be quickly used and discarded to make room for other arrays, even though this may lead to a slower performance.

We also have the option of translating the source code to a faster language such as `Julia`, which is syntactically simple like `Python`. Once the source code has been made to be more efficient, we can use a computer that for more parallelization to further shorten the time needed for parameter estimation. But there are also conceptual issues that needs to be addressed.

In Geng et al.’s paper, we are unsure of how volume distributions values given in table 3 are used and relates to the mean and median values given in table 2. We need more clarification on how Geng et al. have sampled the volume distributions or plotted their histograms since figure 7 did not match figure 4A of their paper. Geng et al. also states that they implemented a ”1.48% survival reduction according to life expectancy tables from the SEER data to account for unexpected natural death events”, but did not explain how, or provide a citation to the SEER database. And Geng et al. also mentions

a uniformly sampled delay of 2 to 3 weeks between diagnosis and the start of treatment. We are unsure of where this delay is used.

Our program successfully replicates figures 3, 5B, and 6B of Geng et al.’s paper with minor differences. We found that accuracy of the forward Euler’s method model improves with smaller step size. But we were surprised by the behaviour of the Runge-Kutta model. Future works can look to explaining this difference and obtaining the actual data sets used to match Geng et al.’s results directly.

The value of mathematically modelling lies in providing accurate predictions. We acknowledge that there are other models to describe tumor growth, each with different assumptions [7, 8, 17]. Murphy et al. showed that failure to consider these assumptions can be severe: a 6-fold difference in the predicted amount of chemotherapy needed to suppress a sample tumor growth [18]. A verification on the validity of the Gompertz model to describe non-small cell cancer growth (over a 60-month period) is needed. Other growth models should also be considered since the model that best fits experimental data might not be the best model to predict future growth.

We have only partially implemented Geng et al.’s model. For future works, we can implement the portion of the model to account for chemotherapy. After implementing the entire model, we can use our models to optimize treatment regimens, or develop a GUI to enable users without coding experience to use our model.

V. CONCLUSION

We have illustrated our process to implement a program in `Python` that can predict the Kaplan-Meier survival curve of an arbitrary non-small cell lung cancer population and patient response to well-separated radiotherapy fractions following Geng et al.’s paper. We reasoned about the proposed differential equation and derived two discrete-time models besides the one provided by Geng et al.. While our model reproduces some of Geng et al.’s results well, further conceptual validations and code improvements in parameter estimation is required before it can be applied clinically to optimize treatment scheduling.

-
- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, *CA: A Cancer Journal for Clinicians* **68**, 394 (2018).
 - [2] N. Howlader, A. Noone, M. Krapcho, D. Miller, A. Brest, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D. Lewis, H. Chen, E. Feuer, and K. Cronin, *SEER Cancer Statistics Review*, (1975-2016. November 2018. April 2019).
 - [3] C. Geng, H. Paganetti, and C. Grassberger, *Scientific reports* **7**, 1 (2017).
 - [4] J. R. Molina, P. Yang, S. D. Cassivi, S. E. Schild, and A. A. Adjei, in *Mayo Clinic Proceedings*, Vol. 83 (Elsevier, 2008) pp. 584-594.
 - [5] PDQ Adult Treatment Editorial Board, “Non-small cell lung cancer treatment (pdq): Health professional version,” (2020).
 - [6] S. Tsim, C. O’Dowd, R. Milroy, and S. Davidson, *Respiratory medicine* **104**, 1767 (2010).
 - [7] C. Grassberger and H. Paganetti, *Physics in Medicine & Biology* **61**, R344 (2016).
 - [8] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. Ebos, L. Hlatky, and P. Hahnfeldt, *PLoS Comput Biol* **10**, e1003800 (2014).
 - [9] P. Mayles, A. Nahum, J. Rosenwald, and N. Papanikolaou, *Medical Physics* **35**, 4281 (2008).
 - [10] D. J. Brenner, *Seminars in radiation oncology* **18**, 234 (2008).
 - [11] G. Changran, H. Paganetti, and C. Grassberger, *Scientific Reports* **8**, 12631 (2018).
 - [12] D. Raz, J. Zell, S.-H. Ou, D. Gandara, H. Anton-Culver, and D. Jablons, *Chest* **132**, 193 (2007).
 - [13] W. Sause, P. Kolesar, S. Taylor, D. Johnson, R. Livingston, R. Komaki, B. Emami, W. Curran, R. Byhardt, R. Dar, and A. III, *Cancer/Radiothérapie* **5**, 206 (2001).
 - [14] F. Marin, A. Rohatgi, and S. Charlot, (2017).
 - [15] M. Newville, T. Stensitzki, D. B. Allen, and A. Ingargiola, “LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python,” (2014).
 - [16] P. Moritz, R. Nishihara, S. Wang, A. Tumanov, R. Liaw, E. Liang, M. Elibol, Z. Yang, W. Paul, M. I. Jordan, and I. Stoica, [1712.05889v2](https://doi.org/10.1712.05889v2).
 - [17] L. von Bertalanffy, *The Quarterly Review of Biology* **32**, 217 (1957), pMID: 13485376, <https://doi.org/10.1086/401873>.
 - [18] H. Murphy, H. Jaafari, and H. M. Dobrovolsky, *BMC cancer* **16**, 163 (2016).