Article

Bayesian Information-Theoretic Calibration of Radiotherapy Sensitivity Parameters for Informing Effective Scanning Protocols in Cancer

Heyrim Cho [1,†], **Allison L. Lewis [2,*,†]** and **Kathleen M. Storey [3,†]**

1   Department of Mathematics, University of California, Riverside, CA 92521, USA; heyrimc@ucr.edu
2   Department of Mathematics, Lafayette College, PA 18042, USA
3   Department of Mathematics, University of Michigan, MI 48109, USA; storeyk@umich.edu
*   Correspondence: lewisall@lafayette.edu
†   These authors contributed equally to this work.

check for updates

Abstract: With new advancements in technology, it is now possible to collect data for a variety of different metrics describing tumor growth, including tumor volume, composition, and vascularity, among others. For any proposed model of tumor growth and treatment, we observe large variability among individual patients' parameter values, particularly those relating to treatment response; thus, exploiting the use of these various metrics for model calibration can be helpful to infer such patient-specific parameters both accurately and early, so that treatment protocols can be adjusted mid-course for maximum efficacy. However, taking measurements can be costly and invasive, limiting clinicians to a sparse collection schedule. As such, the determination of optimal times and metrics for which to collect data in order to best inform proper treatment protocols could be of great assistance to clinicians. In this investigation, we employ a Bayesian information-theoretic calibration protocol for experimental design in order to identify the optimal times at which to collect data for informing treatment parameters. Within this procedure, data collection times are chosen sequentially to maximize the reduction in parameter uncertainty with each added measurement, ensuring that a budget of n high-fidelity experimental measurements results in maximum information gain about the low-fidelity model parameter values. In addition to investigating the optimal temporal pattern for data collection, we also develop a framework for deciding which metrics should be utilized at each data collection point. We illustrate this framework with a variety of toy examples, each utilizing a radiotherapy treatment regimen. For each scenario, we analyze the dependence of the predictive power of the low-fidelity model upon the measurement budget.

Keywords: mutual information; bayesian experimental design; tumor growth models; radiotherapy treatment; scanning protocol

1. Introduction

Mathematical and computational approaches have long been developed for a better understanding of cancer [1–5]. Complex mechanisms that control carcinogenesis, tumor progression, and dynamics under treatments have been studied through various mathematical models [1,4]. Such quantitative models are utilized to predict cancer dynamics and therapy response, improve early detection, and design therapeutic protocols [5]. Moreover, advances in technology have made it possible to collect considerable amounts of detailed data, including genetic, cellular level, and tissue-scale imaging, describing various aspects of cancer progression in patients [6,7]. Incorporating the available data into an appropriate mathematical model can enhance the effectiveness of personalized treatments. However,

practically speaking, an abundance of data may not be available in a clinical setting, particularly in the temporal domain. Specifically, data can often only be collected at sparse times, due to physical and financial burdens to patients. Therefore, with a limited number of total scans, it is important to decide when to collect the data to most accurately calibrate the model for maximum predictive power. Moreover, another important purpose of data collection is to aid in making an accurate prediction as early as possible during the treatment so that any necessary adjustments—for example, altering the dosage or discontinuing the treatment—can be made as soon as possible. However, this goal is also prohibited by the limited availability of patient data in the temporal domain.

Many different types of mathematical models have been developed to simulate tumor growth and response to treatment. These range from phenomenological models composed of ordinary differential equations (ODEs) [8–10] to complex multi-scale models that combine representations of the tumor microenvironment at the subcellular, cellular, and tissue scales [11–14]. Typically the biological accuracy increases as the model complexity increases; however, achieving unique identifiability of model parameter values becomes more challenging. In [15], we present a framework for choosing an appropriate model and calibrating parameter values, given a set of data describing various tumor metrics over time. Here, we extend that work by developing a methodology to determine the optimal data collection times, in order to maximize the utility of the collected samples.

We focus on applying this methodology to models that simulate tumor response to radiotherapy. Radiotherapy is a common treatment modality applied to many cancer types, and there is a long tradition of mathematical modeling of radiotherapy response. Fractionated radiation dosing is typically modeled using the linear-quadratic (L-Q) model [16,17]. Several recent studies have applied the L-Q model to patient-specific data, in an effort to evaluate and predict individual responses to radiotherapy [15,18–21]. Our work extends this body of literature by suggesting optimal sampling times for tumors undergoing radiotherapy, so that the L-Q parameters can be efficiently calibrated and tuned for more accurate post-treatment predictions.

We utilize a Bayesian information-theoretic experimental design framework for choosing optimal design conditions at which to collect data for model calibration. This methodology is based on the concept of maximizing information gain—and thus reducing uncertainty—about low-fidelity model parameters, while minimizing the number of design conditions at which experimental data must be collected. Originally built around the concept of Shannon entropy [22], and utilizing the kth-Nearest-Neighbor (kNN) estimate of mutual information to make the procedure more computationally feasible [23], this framework was first introduced in [24–26]. The methodology was further extended in [27] to incorporate more robust Bayesian methods for supporting highly correlated and nonlinear parameter dependencies. Here, we further amend this framework to be suitable for a temporal data collection setting, emphasizing the minimization of uncertainty in our model parameters while also penalizing our algorithm for choosing points further out in time, which precludes the potential information gain that would result from data collection at intermediate time steps.

Throughout this investigation, we assume an idealized scenario in which clinicians have the ability to take informative measurements on any given day during treatment with limited measurement noise. We note that in general, the "best" choice of metric may vary depending on the type of cancer, the chosen treatment, and the cost of data collection, among other factors. In addition to measuring total tumor volume via 2D or 3D imaging techniques and estimating viable tumor fraction from scan cross-sections—both of which we simulate here in our test scenarios—other choices of metrics could include testing for certain tumor antigens (such as prostate-specific antigen, or PSA, in prostate cancer [28]), measuring white blood cell counts for leukemia and other blood cancers, or checking for genetic biomarkers for use in a diagnostic setting. To illustrate the proposed methodology, we'll assume a scenario in which we have access to tumor volume scans that are informative on a short time scale (i.e., there is not a significant lag in tumor volume response to the prescribed RT treatment). Though this may not be practical—see [29,30] for a discussion of the effect of radiation-induced inflammation on the accuracy of tumor volume measurements—we emphasize that the purpose of this investigation is

to demonstrate the feasibility of a mathematical framework for determining optimal times, which can be adapted to whichever model, treatment, and metric combination is suitable for the particular cancer to be analyzed. Additionally, we note that various noninvasive imaging techniques for early assessment of tumor response to treatments are being developed [31–34], which could mitigate some of these concerns.

In Section 2, we introduce the mathematical models for tumor growth and radiotherapy that will be used to illustrate our methodology throughout this work. Section 3 outlines the procedure for using mutual information to optimally choose design conditions at which to evaluate tumor data in order to maximize information gain about model parameters for accurate and timely calibration. This section includes the novel introduction of a score function for temporal data collection, which penalizes the user for skipping too many potential data evaluation times in pursuit of the largest possible information gain. In Section 4, we illustrate our proposed framework for several examples that showcase a wide variety of scenarios with respect to model complexity and data collection protocols. We summarize our findings and discuss plans for future investigation in Sections 4.5 and 5.

2. Mathematical Models Used for Testing

We begin by presenting two mathematical models that will be used throughout this work as low-fidelity models to be calibrated for clinical predictions, and another model that will be used to generate synthetic high-fidelity data. The first low-fidelity model is a one-compartment ODE model that tracks tumor volume over time—our simplest model for describing tumor growth. The second model is a two-compartment ODE model that incorporates a state variable for tracking the portion of tumor volume that is composed of necrotic tissue, thereby introducing the concept of tumor heterogeneity.

These two models will be calibrated using high-fidelity data generated by a more complex model, namely, a cellular automaton model. In addition to quantifying both the viable and necrotic cell populations, the cellular automaton model also tracks the cell division cycle, quiescent cells, and oxygen levels, allowing for more accurate reflection of the stochastic and heterogeneous nature of cancer growth in reality [14,15]. In all three cases, we incorporate treatment via radiation using the linear-quadratic model for radiotherapy [35,36], as outlined in Section 2.4.

2.1. The One-Compartment Ode Model

The one-compartment model describes the time evolution of the total tumor volume, $V(t)$, using a logistic growth model with growth rate λ and carrying capacity K:

$$\frac{dV}{dt} = \underbrace{\lambda V \left(1 - \frac{V}{K}\right)}_{\text{logistic growth}} - \underbrace{\eta V.}_{\text{natural cell death}} \tag{1}$$

We incorporate natural cell death via the term $-\eta V$. We note that, Equation (1) is not structurally identifiable [15], since there exist infinitely many pairs (λ, η) which yield the same net growth rate $\lambda - \eta$. Thus, it will be convenient to re-parameterize Equation (1) to obtain the simpler and parametrically-identifiable form

$$\frac{dV}{dt} = AV \left(1 - \frac{B}{A}V\right), \tag{2}$$

where $A = \lambda - \eta$ and $B = \frac{\lambda}{K}$. From this point forward, any reference to the one-compartment model is referring to the re-parameterized form, Equation (2).

Biologically speaking, as a tumor grows, regions at a distance from oxygen and nutrient sources (e.g., blood vessels for tumors growing *in vivo*) may undergo necrosis in response to sustained oxygen and/or nutrient deprivation. In this simple one-compartment model, such dead or necrotic cells are assumed to be removed from the tumor instantaneously; that is, we view the tumor as a homogeneous mass of proliferating, viable cells.

2.2. The Two-Compartment Ode Model

In order to account for some aspects of tumor heterogeneity, we next study a two-compartment model that tracks the time evolution of the viable tumor cell volume, $V(t)$, and the necrotic core volume, $N(t)$, originally developed in [37] and further analyzed in [15]. We consider this model in an effort to better represent reality, as it has been shown that the proportion of necrotic material has a significant impact on one's ability to estimate a tumor's response to radiotherapy [15,38,39]. We still assume that the population of proliferating (i.e., viable) cells, $V(t)$, grows logistically with growth rate λ and carrying capacity K, and that viable cells convert to necrotic cells at a constant rate η. In this second model, we assume that the natural death of viable cells results in the build-up of a necrotic core, denoted by $N(t)$, where the necrotic material then undergoes natural decay at a constant rate ζ. We note that as $\zeta \to \infty$, the two-compartment model converges to the one-compartment model behaviorally. Combining these processes, we arrive at the following ODE system for $V(t)$ and $N(t)$:

$$\frac{dV}{dt} = \lambda V \left(1 - \frac{V}{K} \right) - \eta V, \tag{3a}$$

$$\frac{dN}{dt} = \eta V - \zeta N. \tag{3b}$$

Throughout the following investigation, we refer to this system (3) as the two-compartment model.

2.3. The Cellular Automaton Model

In the absence of experimental data, we generate synthetic total tumor and necrotic volume data using a spatially-explicit, hybrid cellular automaton (CA) model to allow for illustration of our methodology. Throughout this investigation, the synthetic data from the CA model serves as the high-fidelity data in the high-to-low fidelity model calibration framework referred to in [27]. Our cellular automaton model is adapted from that developed in [14] and later expanded in [15]. The model incorporates spatially heterogeneous oxygen levels, a stochastic cell life cycle, and a heterogeneous cell population, including proliferating, quiescent, and necrotic cells. The cells are arranged on a discrete lattice representing a two-dimensional square cross-section of size 0.36×0.36 cm^2 through a three-dimensional cancer spheroid *in vitro*. We identify with each automaton $\mathbf{x} = (x,y)$ at time t a dynamical variable with a state and a neighborhood.

Each automaton can be occupied by a tumor cell in one of three states—proliferating, $\mathcal{P}$, quiescent, $\mathcal{Q}$, or necrotic, $\mathcal{N}$—or can be unoccupied and denoted as empty, $\mathcal{E}$. We note that all lattice cells have an associated oxygen level, regardless of whether they are currently occupied. This oxygen level, c, determines the state of the occupying cell using thresholds c_N and c_Q: if $c > c_Q$ then the cells proliferate, if $c_N < c < c_Q$ then the cells transition to quiescent cells with a halved oxygen consumption rate, and if $c \leq c_N$ then the cells are considered necrotic.

We model the single growth-rate-limiting nutrient, oxygen, explicitly via a reaction-diffusion equation. See [14,15] for a detailed description of the oxygen model. If a cell becomes necrotic due to low oxygen concentration or irradiation, which will be discussed in Section 2.4, the necrotic cells are lysed at rate p_{NR}. Lysis involves removing the necrotic cell and then shifting inward a chain of cells starting from the boundary of the spheroid to fill in the removed cell's site. Additionally, the CA mimics several other mechanisms observed in tumor spheroids. We incorporate the regulatory process known as contact inhibition of proliferation by reducing the division of cells with a large number of neighbors. We simulate cell-cell adhesion by shifting chains of cells outward after cell division. The details of these model features are described in [14,15].

We use the CA model to generate a series of synthetic spheroids that differ in their response to radiosensitivity. The parameter values that are used to generate data using the CA model are shown in Table A1 in the Appendix A. These parameters are estimated using experimental data from the prostate cancer cell line, PC3, in [14]. Note that the parameter values are listed with volumetric units;

we convert the units to a two-dimensional cross sectional area by assuming that the three-dimensional tumor takes on an ellipsoidal shape, with volume measured by $V = 0.5 * h * v^2$, where h is the length in the horizontal direction through the tumor center, and v is the length in the vertical direction.

2.4. Radiotherapy Treatment

Finally, we discuss the incorporation of a radiotherapy (RT) treatment protocol in all three models. We consider a typical tumor treatment regimen in which daily doses of 2 Gy are administered Monday through Friday for six consecutive weeks. We use the linear-quadratic model [35,36] to account for the effects of RT. This model assumes that the fraction of cells that survive exposure to a single administered dose d of RT is given by

$$\text{Survival fraction,} \quad SF = e^{-\alpha d - \beta d^2}, \tag{4}$$

where α and β represent tissue-specific radiosensitivity parameters that model single- and double-strand breaks of the DNA, respectively [40]. We note that the linear-quadratic model is a reasonable choice for fast-growing tumors, but it does not account for delayed effects from radiotherapy, thus making this model less applicable to slow-growing tumors [41].

For implementation in the one-compartment model, we assume that the effect of RT is instantaneous. That is, the irradiated cell fraction is removed immediately from the tumor volume, akin to the treatment of natural cell death in our model. Under these assumptions, the one-compartment model with RT treatment can be re-formulated as

$$\begin{cases} \frac{dV}{dt} = AV \left(1 - \frac{B}{A} V \right), & \text{for } t_i^+ < t < t_{i+1}^-, \\ V(t_i^+) = \exp(-\alpha d - \beta d^2) \, V(t_i^-), \end{cases} \tag{5}$$

where t_i (for $i = 1, 2, \ldots, n_R$) denote the times at which an RT dose is delivered, and $V(t_i^{\pm})$ denote the tumor volume just before and after radiotherapy is administered. We acknowledge that the assumption of exponential treatment response may overestimate radiation sensitivity, and our planned future work includes the comparison of different radiotherapy models [18,20,42,43] within our framework.

Because our two-compartment model allows for the existence of a necrotic core, our implementation of RT in the two-compartment model assumes that irradiated cells from the total tumor volume move directly into the necrotic compartment, and later decay naturally. This manifests visually as a delayed reaction to RT in the total tumor volume; it is the heterogeneous composition of the tumor that displays the immediate effects of RT.

We apply the linear-quadratic model to the cellular automaton model in a similar fashion. At each administration of RT, each living cell converts to a necrotic cell with probability $1 - e^{-\alpha d - \beta d^2}$, and then undergoes natural decay during a later iteration with probability p_{NR}, as detailed in Section 2.3. Again, we acknowledge that this may lead to an overestimate of the necrotic region, since radiation-induced cell death often occurs more quickly than hypoxia-induced necrosis. In future work, we plan to increase the accuracy of the CA model by scaling cells' radiosensitivity based on the local oxygen level.

To illustrate our framework for several scenarios with data displaying a variety of behavior with regard to treatment success, we produce three different sets of test data using the CA model. These three test scenarios are representative of a patient whose tumor is highly sensitive to radiotherapy (generated in the CA model using $\alpha = 0.14$, $\beta = 0.14$, for a radiosensitivity parameter ratio of 1), a patient whose tumor is somewhat sensitive to RT (using $\alpha = 0.14$ and $\beta = 0.0467$ for a ratio of $\alpha/\beta = 3$), and a patient whose tumor does not respond meaningfully to RT (using $\alpha = 0.14$ and $\beta = 0.0156$ for a ratio of $\alpha/\beta = 9$). Throughout the investigation, we will refer to these three test scenarios as "high", "medium", and "low" radiosensitivity, respectively. The three sets of CA synthetic

data are shown in Figure 1. In all patients, we consider a typical treatment protocol in which daily doses of 2 Gy are administered Monday through Friday for 6 weeks, as discussed above.
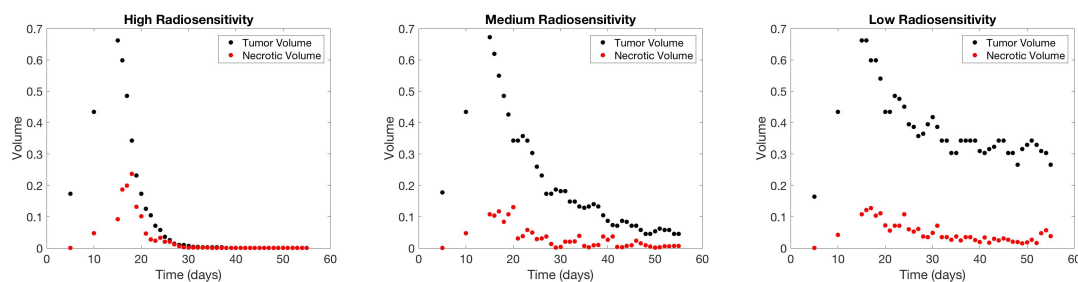

Figure 1. High-fidelity CA data for each of the three radiosensitivity scenarios: high ($\alpha/\beta = 1$, on left), medium ($\alpha/\beta = 3$, in middle), and low ($\alpha/\beta = 9$, on right).

3. Bayesian Information-Theoretic Methodology

Our goal is to calibrate the parameters of a low-fidelity model (Equations (2) or (3)) using as few high-fidelity data collections (or in this case, as few evaluations of the high-fidelity CA model) as possible. In Section 3.1, we outline the mutual information methodology used to choose the most informative high-fidelity data points with respect to low-fidelity parameter uncertainty. Then, in Section 3.2, we present our adaptation of this method to allow for its use on time series data.

3.1. Experimental Design Framework Using Mutual Information

Given a set of initial experimental data (or high-fidelity code evaluations) and a low-fidelity model that we wish to calibrate, we seek to determine the optimal design conditions at which to collect additional data (or evaluate our high-fidelity model) in order to best inform the parameters of the low-fidelity model, $\theta \in \mathbb{R}^p$. We note that throughout this investigation, we will use a high-fidelity code to produce synthetic data to represent experimental data acquisition. We denote the existing high-fidelity data set by $D_{n-1} = \{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_{n-1}\}$, and select the next design condition ξ_n from the set of all possible evaluation strategies or experimental designs, Ξ, in such a way as to maximize reduction in the uncertainty in θ when $\tilde{d}_n = d_h(\xi_n)$—the data point resulting from evaluation at design ξ_n—is appended to the existing data set.

Because we cannot determine the value of $\tilde{d}_n$ without first evaluating the high-fidelity code, we must base our decision of ξ_n on predictions from the low-fidelity model. The low-fidelity model prediction for a specified condition is denoted by $d_n = d_\ell(\theta, \xi_n)$, where $d_\ell(\theta, \xi_n)$ represents the low-fidelity model evaluated at parameter set θ and design ξ_n. Predictions made by this model can be extended into a statistical modeling framework by incorporating a model discrepancy term, $\delta(\xi_n)$, and the possibility of measurement or discretization errors, $\varepsilon_n(\xi_n)$. In this study, we assume no model discrepancy and zero measurement errors, and refer the interested reader to [27] for additional details on the corresponding statistical models for both the low-fidelity and high-fidelity frameworks.

Recall that our objective is to calibrate the low-fidelity model parameters using as little high-fidelity data as possible, so as to minimize unnecessary computational or experimental costs. Thus, we want to choose design conditions for the high-fidelity code so as to contribute the maximum amount of information about our low-fidelity model parameters. We do so using a Bayesian framework. We begin with Bayes' Rule, which gives a method by which to update the knowledge about the parameters based on the addition of a new data point:

$$p(\theta|D_n) = \frac{p(D_n|\theta)p(\theta)}{p(D_n)} = \frac{p(\tilde{d}_n, D_{n-1}|\theta)p(\theta)}{p(\tilde{d}_n, D_{n-1})}.$$

In essence, Bayes' Rule states that the posterior distribution of θ, given data D_n, depends both on the likelihood of obtaining data D_n given the parameter set θ, represented by $p(D_n|\theta)$, and the prior distribution of θ, $p(\theta)$, which incorporates all known information about θ.

As a means of quantifying the information gain obtained from including an additional data point, we utilize the mutual information between the low-fidelity model parameters and the high-fidelity data. We include here the basic layout about the mutual information derivation, and refer the interested reader to [23,24,27] for additional details.

The average level of uncertainty or information inherent in a random variable can be quantified by its Shannon entropy [22]. For a random variable Θ having a corresponding density $p(\theta)$ for $\theta \in \Omega$, the Shannon entropy is defined as

$$H(\Theta) = - \int_\Omega p(\theta) \log(p(\theta)) d\theta$$

for the prior distribution and

$$H(\Theta|D_{n-1}) = - \int_\Omega p(\theta|D_{n-1}) \log(p(\theta|D_{n-1})) d\theta$$

for the posterior distribution, given data D_{n-1}. Since we wish to determine the gain in information that results from incorporating an additional high-fidelity data point $\tilde{d}_n$, we define our utility function to be the following difference:

$$
\begin{aligned}
U(d_n, \xi_n) = & \int_\Omega p(\theta|d_n, D_{n-1}) \log p(\theta|d_n, D_{n-1}) d\theta \\
& - \int_\Omega p(\theta|D_{n-1}) \log p(\theta|D_{n-1}) d\theta.
\end{aligned}
\tag{6}
$$

Note that in Equation (6), we use the low-fidelity model prediction d_n as an estimate in place of the high-fidelity output $\tilde{d}_n$, as we cannot know the value of this high-fidelity output without conducting a potentially expensive experiment.

To compute the average information gain resulting from the contribution of the experimental design ξ_n, we marginalize over the full set of all unknown future observations, $\mathcal{D}$, to obtain

$$\mathbb{E}_{d_n}[U(d_n, \xi_n)] = \int_\mathcal{D} U(d_n, \xi_n) p(d_n|D_{n-1}, \xi_n) dd_n. \tag{7}$$

By substituting Equation (6) into (7) and simplifying, we obtain

$$
\begin{aligned}
\mathbb{E}_{d_n}[U(d_n, \xi_n)] &= \int_\mathcal{D} \int_\Omega p(\theta, d_n|D_{n-1}, \xi_n) \log \frac{p(\theta, d_n|D_{n-1}, \xi_n)}{p(\theta|D_{n-1}) p(d_n|D_{n-1}, \xi_n)} d\theta dd_n \\
&= I(\theta; d_n|D_{n-1}, \xi_n),
\end{aligned}
\tag{8}
$$

which we define to be the mutual information between the low-fidelity model parameters, θ, and the high-fidelity data collected at design ξ_n. Essentially, this is a measure of parameter uncertainty reduction—the larger the mutual information, the more knowledge we expect to gain about the low-fidelity model parameters by collecting data at that experimental design. Thus, we optimize over the set of all available design conditions, choosing the one that maximizes this quantity:

$$\xi_n^* = \underset{\xi_n \in \Xi}{\arg\max} \, I(\theta; d_n|D_{n-1}, \xi_n).$$

Having chosen ξ_n^*, we evaluate the high-fidelity model at this point, $\tilde{d}_n = d_h(\xi_n^*)$, append $\tilde{d}_n$ to the data set D_{n-1}, and re-calibrate the low-fidelity model parameters. For this investigation, we use the Delayed Rejection Adaptive Metropolis (DRAM) algorithm for our model calibration step—see [27,44,45]

for additional details. Once the low-fidelity model parameters have been re-calibrated, this procedure is repeated in full until either (a) the budget of high-fidelity model evaluations has been exhausted or (b) the uncertainty in the low-fidelity model parameters has been reduced below a user-defined threshold.

We note that Equation (8) often cannot be evaluated directly and may be prohibitively expensive to compute via numerical integration. As such, we utilize the *k*th-Nearest-Neighbor (*k*NN) estimate of mutual information described fully in [23,27]. We detail our *k*NN estimate procedure in Appendix B.

3.2. Modified Mutual Information for Time Series Data

When collecting clinical data to determine a tumor's response to treatment, we consider the set of available design conditions to be a series of time steps at which data can be collected sequentially. We note that it is critical to collect measurements at multiple time points to achieve an accurate calibration of the model. Thus, we adapt the calibration framework described in the previous subsection to apply to time-series data. We present this adaptation using the one-compartment model as the low-fidelity model, and note that it can be extended to the two-compartment model by specifying the type of data to be collected (i.e., total tumor volume versus necrotic volume), in addition to the time that it is to be collected.

Let $t_1, t_2, \ldots, t_{n_T}$ denote the set of n_T times at which high-fidelity data can be collected, with $t_i < t_j$ for $i < j$. We use $\tilde{d}(t_i)$ to denote the value of the high-fidelity data, e.g., the tumor volume as evaluated by the CA model, collected at time t_i; the low-fidelity model prediction at t_i is denoted by $d(t_i)$. Suppose that in the previous step of the algorithm, the data point $\tilde{d}(t_r)$ was chosen to append to the data set, which we denote by D_r. When choosing the next data point to add to this data set, for all $i > r$, we let $I(\theta; d(t_i)|D_r)$ denote the mutual information between the low-fidelity model parameters, θ, and the high-fidelity data collected at time t_i, as defined in the general setting in Equation (8).

In some cases, data collected at later time points may provide more overall mutual information, but choosing such a data point then precludes the subsequent collection of data at any previous time. In order to account for this trade-off, we define a score function based on mutual information, which rewards the user for selecting a point with a large mutual information, while penalizing the user for losing the information that could have been collected from the skipped time points. Before presenting this score function, we first define the mutual information relative to the maximum mutual information at a given step in the algorithm. Let I_r^* denote the maximum possible mutual information in the step after data point $\tilde{d}(t_r)$ has been chosen, defined as follows:

$$I_r^* = \max_{i>r} I(\theta; d(t_i)|D_r).$$

Next we define the corresponding relative mutual information provided by each data prediction $d(t_i)$ by

$$R(i, r) = \frac{I(\theta; d(t_i)|D_r)}{I_r^*}.$$

Note that $0 \leq R(i, r) \leq 1$ for all $i > r$. Using this notation, we are now ready to define our score function $S_k(i, r)$, which combines the relative mutual information obtained from choosing data point $d(t_i)$, with a penalty term that captures skipped information. We use the parameter k to vary the weight of the penalty term. We define the score function as follows,

$$S_k(i, r) = R(i, r) - k \left(\frac{\sum_{j=r+1}^{i-1} R(j, r)}{\sum_{l=r+1}^{n_T} R(l, r)} \right), \tag{9}$$

where the sum in the numerator denotes the total amount of relative mutual information that will be lost by choosing data point $\tilde{d}(t_i)$, and the sum in the denominator denotes the total relative mutual information from all data points that may collected after $\tilde{d}(t_r)$.

In Sections 4.2–4.4, we use the score function $S_k(i, r)$ to determine the optimal high-fidelity data points to collect for use in low-fidelity model calibration. At each step of the model calibration algorithm, we choose the data point that maximizes $S_k(i, r)$, that is, $\tilde{d}(t_{i*})$ such that $i^* = \arg\max_{r \leq i \leq n_T} S_k(i, r)$. We test this process multiple times, varying the value of the penalty weight parameter, k, between 0 and 1, in order to investigate the effect of k on the temporal data collection sequence. Note that $k = 0$ is equivalent to using solely the mutual information to determine the next design choice.

4. Simulation Results

In the following section, we outline the results of our algorithm applied to a variety of scenarios. First, we apply our algorithm in a simplified setting where we assume that the clinician has the ability to collect one scan per week, and look for a pattern in the days chosen for scanning. Then, we extend the use of the algorithm to scenarios in which n scans can be collected at any days during the treatment cycle, and investigate how the algorithm performs for both the one- and two-compartment models for three radiosensitivity levels: high, medium, and low. In all of the following scenarios, we fix the pre-treatment parameters (A and B in the one-compartment model, and λ, K, η, and ζ in the two-compartment model) at the values listed in Table A2, assuming that these parameters have been identified prior to the start of the treatment regimen. Additionally, we fix $\alpha = 0.14$ to avoid identifiability issues, as discussed in [15]. Thus, our focus in this investigation is in determining the value of radiosensitivity parameter β as quickly and accurately as possible, so as to increase the predictive power of our models and allow for the alteration or discontinuation of a treatment protocol that is predicted to be ineffective.

We note that throughout this investigation, we have based our nominal parameter values and data collection protocol on experimental procedures using a prostate cancer cell line, with tumor spheroids growing *in vitro* in a lab. Clinically speaking, this is akin to assuming a highly idealized scenario in which clinicians have the ability to take informative measurements on any given day during treatment with limited measurement noise. We recognize that many of our underlying assumptions—in particular, assuming adequate availability of pre-treatment data to allow for accurate calibration of non-treatment parameters prior to beginning RT—will need to be altered in the future to make this directly applicable to a clinical setting.

4.1. Scenario 1: Collecting One Scan Per Week

Frequently in the clinical setting, tumor data collection is constrained to a strict budget due to limited resources. As an example scenario, we consider the case in which one tumor volume scan can be taken per week during the weeks in which treatment is administered. In addition to assuming that data has been collected prior to the start of treatment at days 5 and 10, we automatically provide the scan for day 15 (day one of treatment week one) to obtain an initial fit for parameter β, and then enforce a budget of one scan per week in our mutual information framework with the one-compartment ODE as the low-fidelity model, to determine the optimal day to collect weekly data in weeks 2–6. We complete this procedure for three radiosensitivity levels: high, medium, and low. Table 1 shows the optimal days chosen for each radiosensitivity level, with the day number in the weekly treatment cycle indicated in parentheses (i.e., 1 corresponds to Monday, the first day of treatment each week, and 6 corresponds to Saturday, the day after five sequential days of treatment). Figure 2 provides a visualization of the results summarized in the table.

Table 1. Budget of one scan per week. This table shows the scan choice for high, medium, and low radiosensitivity levels. The day since the start of the simulation is shown, along with the day of each weekly treatment cycle indicated in parentheses. Note that in all cases, we assume that data at days 5, 10, and 15 is already available, so as to avoid the use of improper priors in our parameter calibration. The other scan days were chosen using the mutual information calibration procedure.

	Initial	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
High	5 10	15	22 (1)	29 (1)	36 (1)	43 (1)	50 (1)
Medium	5 10	15	22 (1)	29 (1)	36 (1)	43 (1)	50 (1)
Low	5 10	15	22 (1)	30 (2)	37 (2)	48 (6)	55 (6)
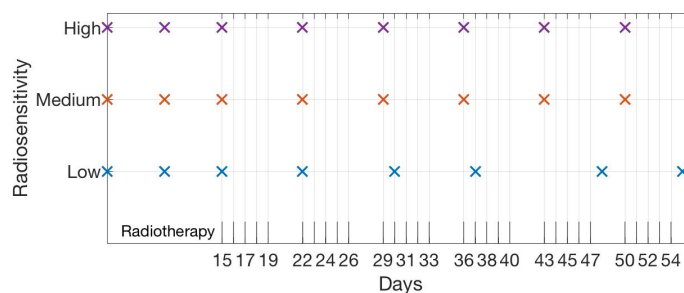
Figure 2. Budget of one scan per week. The scan choices are displayed, with the high radiosensitivity case shown in the upper row in purple, the medium radiosensitivity case shown in the middle in red, and the low radiosensitivity case shown in the lower row in blue.

We observe that in both the high and medium radiosensitivity cases, the first day of each weekly treatment cycle consistently provides the highest level of information for parameter calibration. Figure 3 displays the final model fits for each radiosensitivity case, calibrated using all eight selected scans. In the low radiosensitivity case, the optimal choices are the first day of the treatment cycle in week 2 and the second day of the cycle in weeks 3 and 4—we note that in weeks 3 and 4, days 1 and 2 of the cycle provide nearly identical levels of mutual information. In the final two weeks, day 6 (the Saturday after the treatment cycle ends) provides the highest level of mutual information. Our results suggest that in the latter part of the treatment protocol for this low radiosensitivity case, it becomes most important to assess the full extent of the tumor reduction from each week of radiation doses, encoded in the scan collected on day 6 of each week.

If resources are available for collecting tumor volume data exactly once per week during treatment on any day of the week, then our results suggest that one should begin by collecting this data on the first day of each treatment cycle. If the tumor appears to respond well to the radiotherapy after the first few weeks, then we should continue to collect data on the first treatment day. However, if the tumor is responding slowly after four weeks—for instance, if nearly half of the pre-treatment tumor remains, as in our low radiosensitivity case—then the methodology suggests switching data collection to the end of the week. In particular, it would be optimal in such a scenario to conduct measurements on Saturday, the day following the five-day treatment cycle, for the final two weeks, to provide maximal information for model calibration.

While it may be feasible to collect data on any day of the week in an *in vitro* lab setting, as we look forward towards adapting this framework for a clinical situation, we note that data collection on Saturdays is likely not possible. Thus, it becomes a question of interest to quantify how much information is lost by choosing a "non-optimal" scan in place of the optimal day chosen by the algorithm. In Figure 4, we illustrate the relative mutual information for each day of the week to give an indication of the amount of information loss that would occur if one were to select a different day for data collection. For the high radiosensitivity case, we see that collecting data on the first day of each treatment cycle is the clear optimal choice, as day 1 of each week yields a substantially larger

amount of information than the choice of any other day, particularly in later weeks. However, in the medium and low radiosensitivity cases, the range of information available between days at opposite ends of the week is far less extreme; in these cases, substituting a scan from non-optimal choice will not result in a significant information loss. In particular, with regards to a feasible clinical setting, our previous requirement that data be collected on Saturdays in the low radiosensitivity case can almost certainly be relaxed with no significant repercussions.
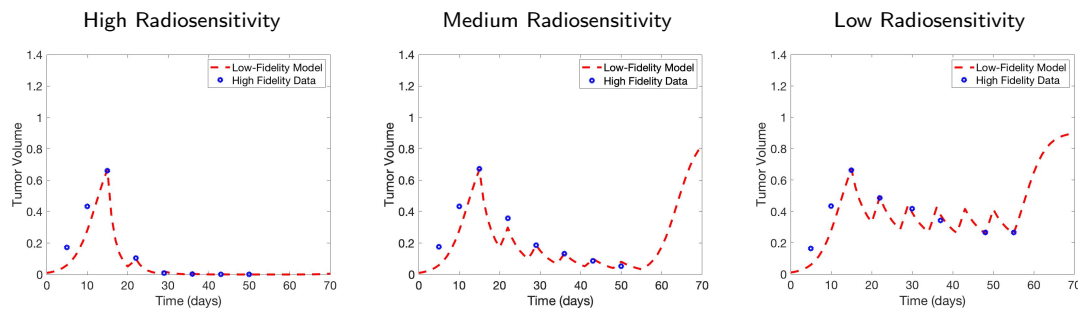
Figure 3. Budget of one scan per week. The fitted models are shown after calibrating with one data point per week, for all six treatment weeks. All three radiosensitivity cases are shown, with high radiosensitivity on the left, medium radiosensitivity in the middle, and low radiosensitivity on the right.
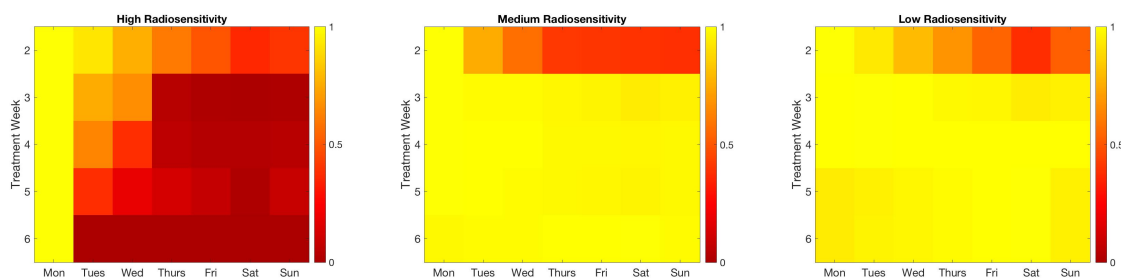
Figure 4. Budget of one scan per week. Relative mutual information is shown for each day of the week. Chosen optimal days have a relative MI of 1. In the high radiosensitivity case, the information gain from choosing day 1 each week is clearly optimal. For the medium and low radiosensitivity cases, it is far less crucial to restrict one's self to the optimal choice in weeks 3–6, as the mutual information is nearly identical for any day of the week.

4.2. Scenario 2: One-Compartment Model with N Number of Scans

In this section, we explore the possibility of designing a more effective scanning schedule to calibrate the one-compartment model, given by Equation (2). For the purpose of the remaining investigations, we assume a very idealized scenario, in which clinicians have the capability to measure the tumor size on any given day. In reality, daily measurements would not be possible, so this methodology must be adapted for use in the clinical setting. Budget constraints can be set based on the clinical availability of the tumor volume measurements.

We now use our proposed score function, (9) in Section 3.2, to determine the scanning schedule. In particular, we study how the suggested scanning schedule changes for different values of the score function parameter k in Equation (9). Moreover, we aim to find an optimal k for each radiosensitivity cases with respect to the accuracy of the model calibration. To quantify the accuracy of the calibration to data, we define

$$\text{error} \doteq \frac{\|\mathbf{d}_h - \mathbf{d}_l\|_2}{\|\mathbf{d}_h\|_2}, \tag{10}$$

where $\mathbf{d}_h = [d_h(t_1), ..., d_h(t_{n_T})]$ is the vector of high-fidelity data at all days during the treatment period, and $\mathbf{d}_l = [V(t_1), ..., V(t_{n_T})]$ is the low-fidelity model approximation evaluated at those same days.

The error includes all the data points up to the final time n_T so that it can be regarded as a measure of predictive power as well.

Figure 5 compares the choice of scan for different values of score function parameter k. We observe that using larger values of k tends to result in the inclusion of more scans from earlier time points, while small values of k often skip those earlier points as a result of the smaller penalty for choosing later points. A sampling of score function values for $k = 0$ and $k = 0.5$, yielding such scan choices, are shown in Figure 6, for the case of high radiosensitivity. The displayed score function values correspond to choosing the third, seventh, and tenth scan. For $k = 0$, which coincides with the standard mutual information of Equation (8), we observe that the score function values vary widely within a single week. In particular, either the first or the last day of the week have large score function values, and thus will be selected. Therefore, by using this score function, the intermediate time points of scans will be skipped. However, by using our score function with $k = 0.5$, more weight is given to earlier time points. This makes it possible to have more frequent earlier scan choices, as shown in Figure 5.
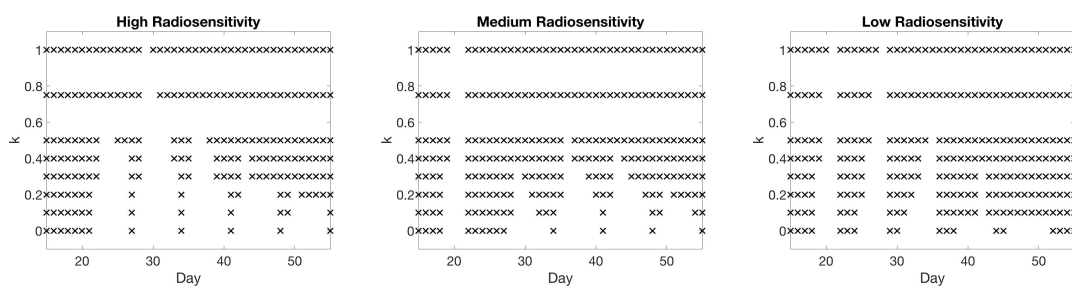

Figure 5. One-compartment model. Choice of scan ($\times$) for different values of score function parameter $k = 0, 0.1, ..., 0.5, 0.75, 1$. We observe that using larger values of parameter k tends to favor the choice of earlier time points, since the choice of point is more heavily penalized for skipping over days. On the other hand, smaller values of k allow the algorithm to skip over earlier time points in favor of gathering points with larger mutual information towards the end of the treatment period.
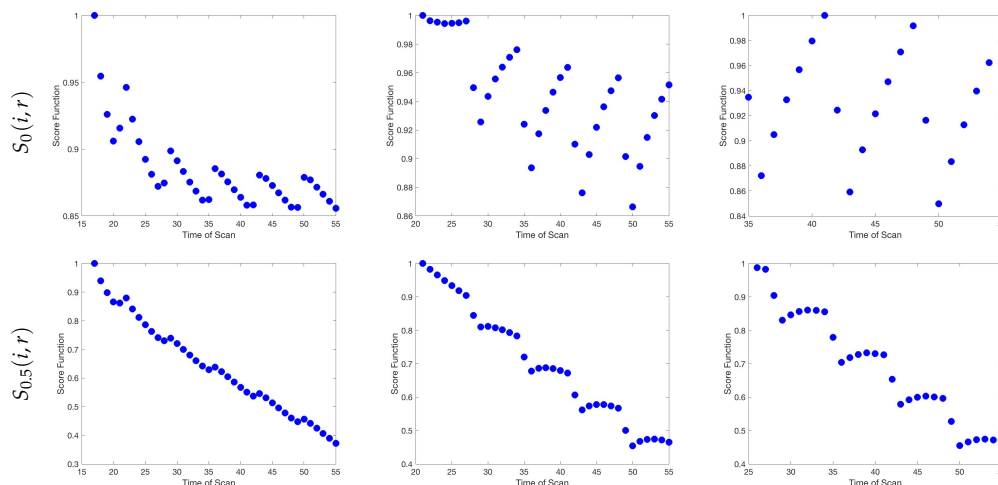

Figure 6. One-compartment model for high radiosensitivity case. Score function $S_k(i, r)$ (9) with respect to $i \in [r + 1, n_T]$ for different values of score function parameter $k = 0$ and $k = 0.5$, when choosing the third (**left**), seventh (**middle**), and tenth (**right**) scan. Larger values of k, for example, $k = 0.5$ give more weight to earlier time points such that the score function values show a decreasing trend, especially shown in the left figure when choosing the third scan. In the case of $k = 0$, the optimal scan choice will skip some of the earlier points.

In Figure 7, we compare the relative error of model fitness to the data as defined in Equation (10) with respect to the number of scans for different values of score function parameter k in three radiosensitivity levels: high, medium, and low. We observe that when using $k = 0$, the error

is reduced most rapidly in the small scan number range. In the case of medium radiosensitivity, any *k* value less than or equal to 0.4 gives similar accuracy when using less than 12 scans. In the low radiosensitivity level, with a fixed budget of 12 scans, $k = 0$ gives the most accurate result. However, larger *k* values give more accurate results as the available scan number become larger. Specifically, *k* values around 0.3 give more accurate outcomes in the medium and low radiosensitivity cases when more than 18 scans are included in the budget. For the high radiosensitivity level, where the tumor responds quickly and drastically to radiation, using larger values of *k* close to $k = 1$ gives the most accurate result with nearly any scan budget. We conclude that using $k = 0$ or a relatively small value of *k* is preferable when the available scan number is small, especially when the tumor is not responsive to treatment, as these small *k* values allow one to skip forward and obtain at least one measurement towards the end of the treatment cycle so as to capture the overall trend in the data. However, when the available scan number increases, $k > 0$ gives more accurate results, especially when the tumor is highly responsive to treatment.
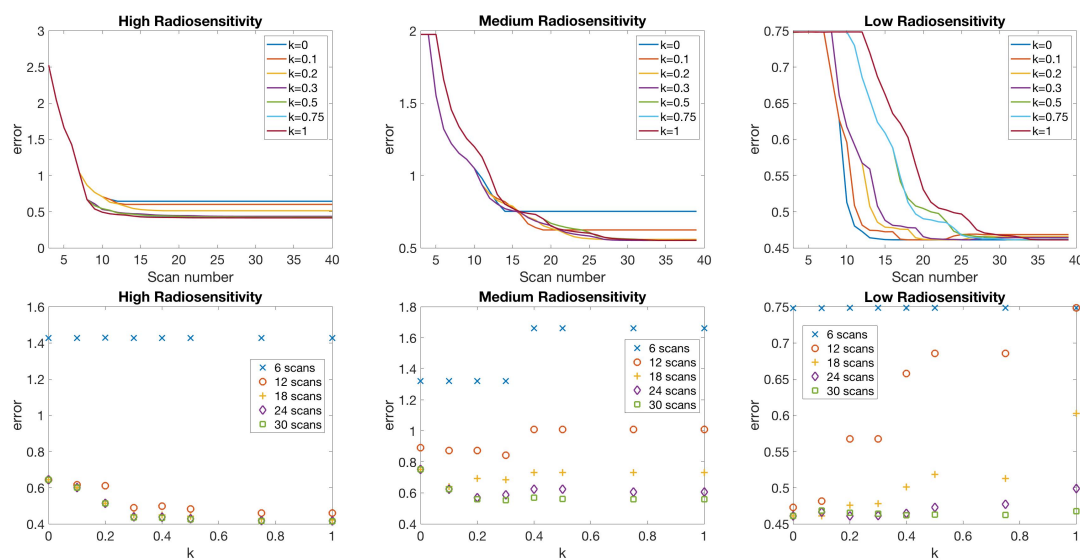

Figure 7. One-compartment model. Relative error of model fitness to the data as defined in Equation (10) with respect to the number of scans for different score function parameter *k* values (**top**), and error with respect to *k* using a fixed number of 6, 12, ..., 30 scans (**bottom**). In general, when the scan number is limited to a small number, (6 and 12 scans), $k = 0$ gives accurate results, particularly when the tumor is less responsive to radiotherapy. However, if a large number of scans is available, $k > 0$ gives more accurate results, especially when the tumor is highly responsive to radiotherapy.

Figure 8 shows the data selected by using a 12 scan budget, with parameter values $k = 0$ and $k = 0.5$, and the fitted model prediction using the one-compartment model (2). When the tumor is highly responsive to radiotherapy, we observe that for larger values of *k*—for example, $k = 0.5$—the fitted model prediction is more accurate. This is unsurprising; since the tumor decay is rapid in these scenarios, it is more informative to learn about the rapidly changing gradient of the data by frequently measuring the tumor data at the earlier time points. On the other hand, when the tumor is less responsive to radiotherapy, learning the overall trend with sparsely placed scans is sufficient, so that using smaller values of *k*, including $k = 0$, gives fairly accurate results. However, the accuracy can be improved with a larger number of scans, using larger *k* values. This is supported by Figure 9, where it can be seen that $k = 0$ results in the fewest total scans, but including more scans (as with the use of larger *k* values) may lead to a more refined final parameter estimate, particularly for the low and medium radiosensitivity cases; the final calibrated *β* parameter values are listed in Table 2 for each of the scenarios illustrated in Figure 9. We see this theme recurring throughout our investigation; small *k* values require fewer scans and are thus less expensive to implement, but in many scenarios—particularly those in which the overall gradient of the data is steeper—choosing a small *k* results in larger prediction errors.
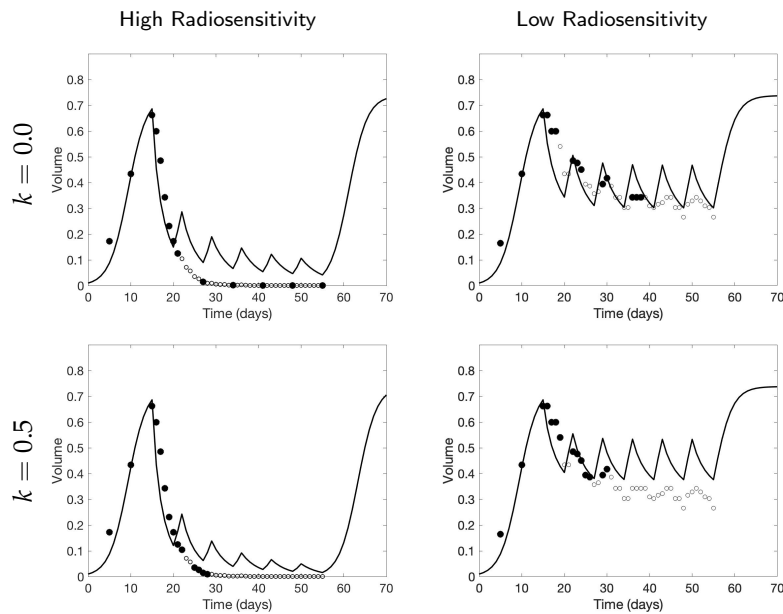
Figure 8. One-compartment model. Choice of high fidelity data (●) among the potential data (○) using a 12 scan budget, and the fitted model prediction (–) of tumor volume using Equation (2) for parameter values $k = 0$ and 0.5. We observe a better fit using $k = 0.5$ in the case of high radiosensitivity and $k = 0$ in the case of low radiosensitivity.

Table 2. One-compartment model. Final calibrated β values for the scenarios shown in Figure 9. We observe that while $k = 0$ uses the fewest number of scans—a fact that makes it the more appealing option in terms of expense—using more scans as in the larger k values for the high and medium radiosensitivity cases results in a significantly different final value, suggesting that more scans may be needed for β to converge to a best estimate.

Radiosensitivity	$k = 0$	$k = 0.2$	$k = 0.5$	$k = 1$
High	0.0838	0.0945	0.1100	0.1194
Medium	0.0526	0.0694	0.0740	0.0745
Low	0.0225	0.0248	0.0239	0.0225

4.3. Scenario 3: Two-Compartment Model with N Number of Scans

In this section, we present results of the selected scanning schedule and accuracy of the calibration in the two-compartment model (3), by selecting either the tumor volume data or the necrotic volume data at each step in the algorithm. Though only one metric can be chosen at a time, we do allow for the next step in the algorithm to pick the second metric at the same time point; i.e., if the algorithm first chooses to measure the tumor volume at day 16, the next point chosen could be the necrotic volume at day 16. The results are similar to the one-compartment model in Section 4.2.

In Figure 10, we observe that the selected scanning schedule becomes more refined in the earlier times for larger values of score function parameter k. We observe that our mutual information criterion often chooses the necrotic volume data over the tumor volume data. This is reasonable, considering that the necrotic volume is more informative for estimating the effect of radiotherapy parameters accurately, since the administration of RT has an immediate effect on the composition of the tumor through the conversion of viable cells to necrotic cells, but only a delayed effect on the total tumor volume.
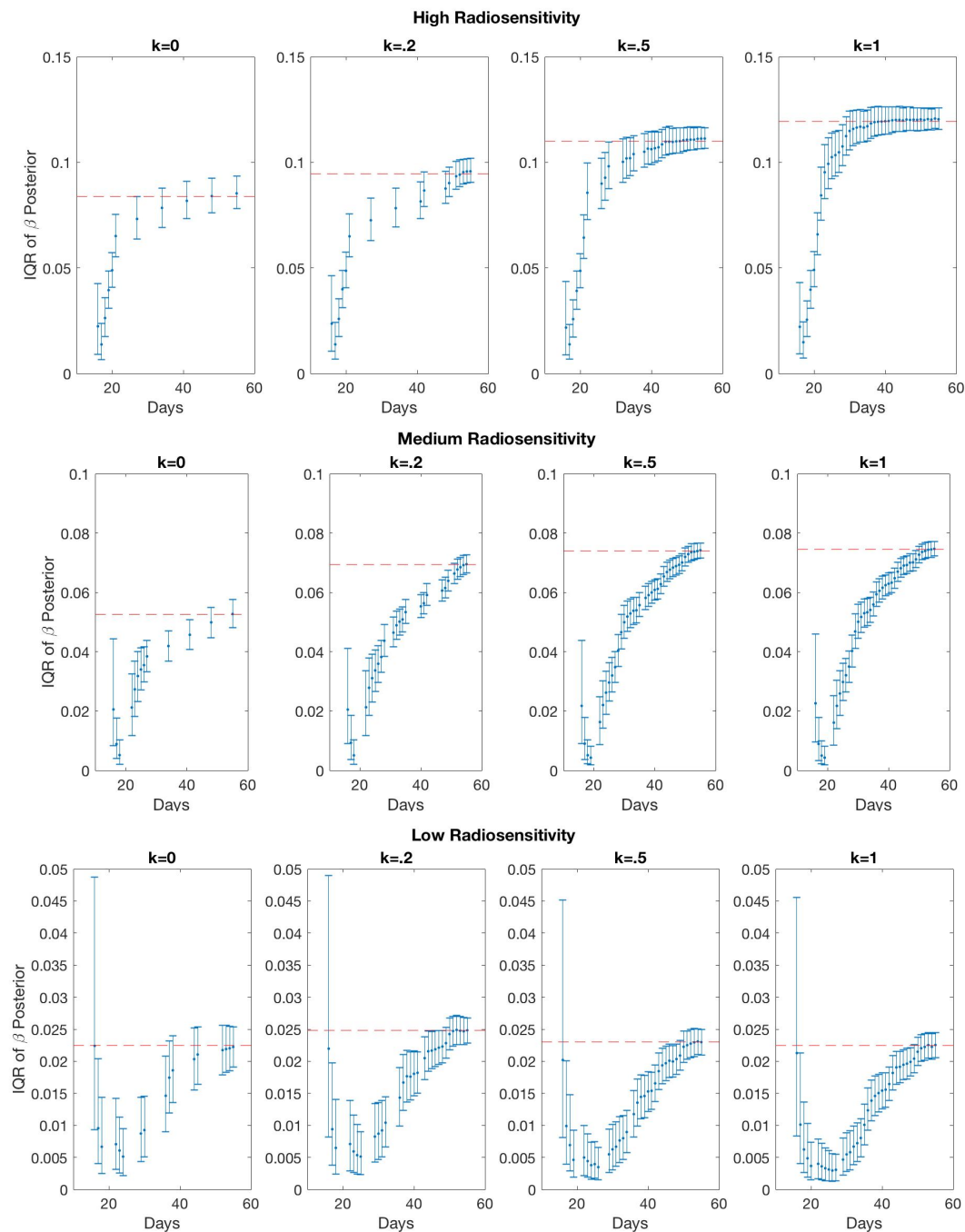
Figure 9. One-compartment model. Interquartile ranges of the posterior distribution for β after the addition of each high-fidelity data point. As more information is gathered, we observe shrinkage in the IQR demonstrating increased precision, and convergence to the final parameter estimate demonstrating increased accuracy. The red dashed lines indicate the final value of the β estimate. For the high and medium radiosensitivity cases, it can be seen that using $k = 0$ does not allow the parameter estimate to fully converge to its actual value, leading to larger errors. In these cases, where the gradient of the data is rapidly changing, we are better off using more scans to achieve our desired accuracy in the final fit.

We remark that choosing necrotic volume data may result in non-monotonic decay in error due to the fact that only providing necrotic volume without tumor volume can deteriorate the accuracy in predicting total tumor volume, as shown in Figure 11. However, the optimal k value with respect to accuracy displays similar results to the one-compartment model. In the case of high radiosensitivity, larger k values, especially $k = 1$, provide the most accurate result. For the medium

to low radiosensitivity cases, $k = 0$ is more accurate when the scan budget is small (for example, less than 25 scans), as it is able to capture some of the final points in this limited scan budget and is thus cognizant of the overall shape of the data across the full treatment period. However, larger values of k (i.e., $k \geq 0.8$) provide a more accurate final fit when large scan budgets are allowed—the inclusion of additional scans in the beginning of the treatment regimen allow for a more refined fit.
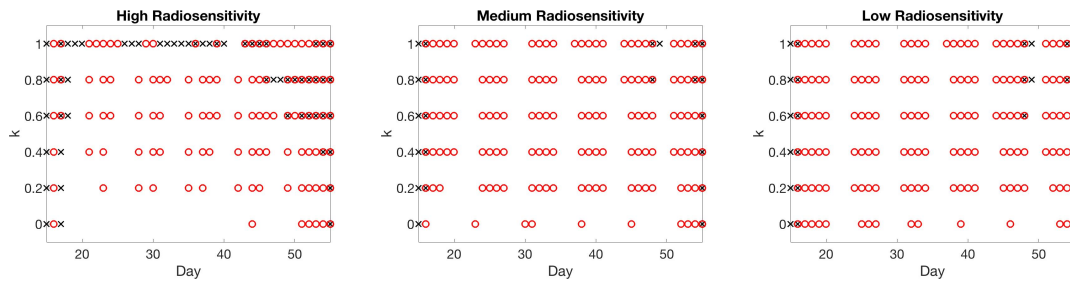
Figure 10. Two-compartment model with one type of point chosen at each step of the algorithm. Choice of scan of either tumor volume ($\times$) or necrotic volume ($\circ$) for different values of score function parameter $k = 0, 0.2, ..., 1$. Larger values of k result in choosing earlier time points in all three radiosensitivity levels.
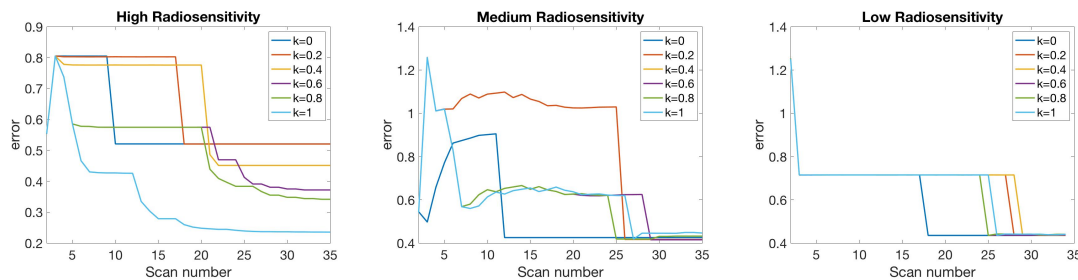

Figure 11. Two-compartment model with one type of point chosen at each step of the algorithm. Relative error of model fitness to the data as defined in Equation (10) with respect to number of scans for each of three radiosensitivity levels. Using $k = 1$ gives the most accurate result in high radiosensitivity level, while $k = 0$ is most beneficial for a low radiosensitivity response.

4.4. Scenario 4: Two-Compartment Model in Practical Setting

While the two-compartment model analysis described in Section 4.3 gives an interesting look at the performance of the algorithm when presented with two alternate metric choices at each time, from a practical standpoint, a clinician who chooses to measure a necrotic proportion at time t would essentially get a tumor volume estimate at time t for "free" from the imaging scan. Thus, in this section we repeat the analysis from Section 4.3, but this time incorporate a tumor volume measurement automatically whenever a necrotic proportion metric is chosen, to better represent the clinical setting.

In Figure 12, we display the design conditions chosen for each of six different k values. In the high radiosensitivity case, there are numerous scenarios in which the MI score function selects tumor volume as the most informative metric, particularly for large values of k when there is a larger penalty for skipping days. In the medium and low radiosensitivity cases, however, the algorithm nearly always chooses necrotic proportion as the more informative metric. Generally, these results agree with those found when tumor volume was not automatically incorporated in Section 4.3.
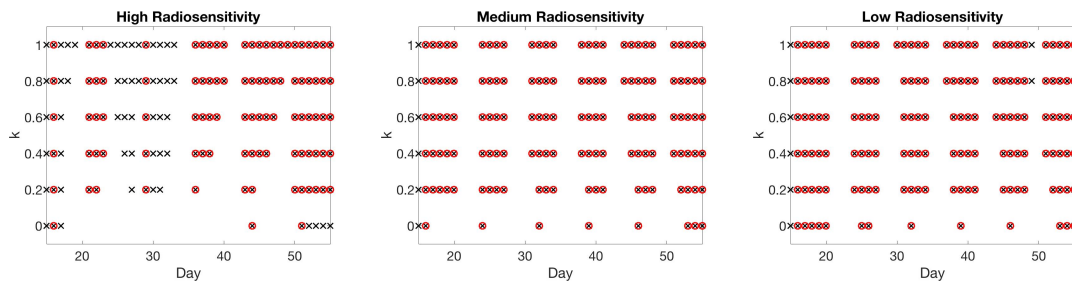
Figure 12. Two-compartment model with tumor volume ($\times$) automatically included when necrotic volume ($\circ$) is chosen. Choice of scan for different values of score function parameter $k = 0, 0.2, ..., 1$. Larger values of k result in choosing earlier time points in all three radiosensitivity levels.

As in the previous two-compartment model analysis, the high radiosensitivity simulation favors larger score function parameter k values in terms of prediction error, while the other two simulations favor smaller k values when scan budget is limited—see Figure 13. Moreover, we emphasize that the prediction error using the two-compartment model is smaller compared to using the one-compartment model when the same number of scans are available, although this is likely due to the additional information about the necrotic fraction. For example, we observe that the two-compartment model using as few as 6 scans in high radiosensitivity with $k = 0.4$—or medium radiosensitivity with $k = 0$—shows far more accurate results compared to using the same number of scans with the one-compartment model; see Figure 14, where we compare the calibrated data fits for the two different models, and Figure 15, where we compare the error with respect to both the number of scans and the days at which scans were collected. Of particular interest in this latter figure is the comparison of error with respect to the time in days, as this gives some indication of a time by which calibration could be considered "good enough"; that is, our parameter values are no longer changing appreciably (as evidenced by a stabilization in the model error), and the procedure can be terminated early, allowing for punctual adaptation of treatment. Further studies must be done with respect to determining the amount of data necessary for ensuring a high degree of confidence in parameter estimates; however, we point the interested reader to [15] where we have already performed preliminary work regarding this very issue.
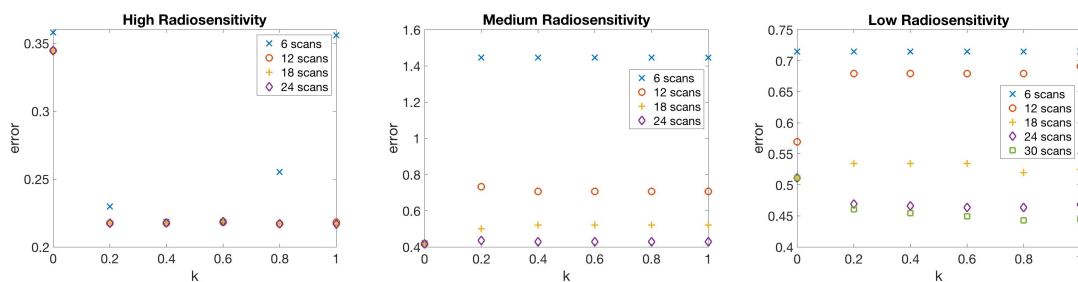

Figure 13. Two-compartment model testing with tumor volume automatically included when necrotic is chosen. Plots of error (10) vs k. In case of high radiosensitivity, using larger values of $k > 0$ is more accurate for all scan budgets. For medium and low radiosensitivity, using $k = 0$ gives the most accurate result when using 6 scans, however, larger values of k yield more accurate predictions when larger scan budgets are available.
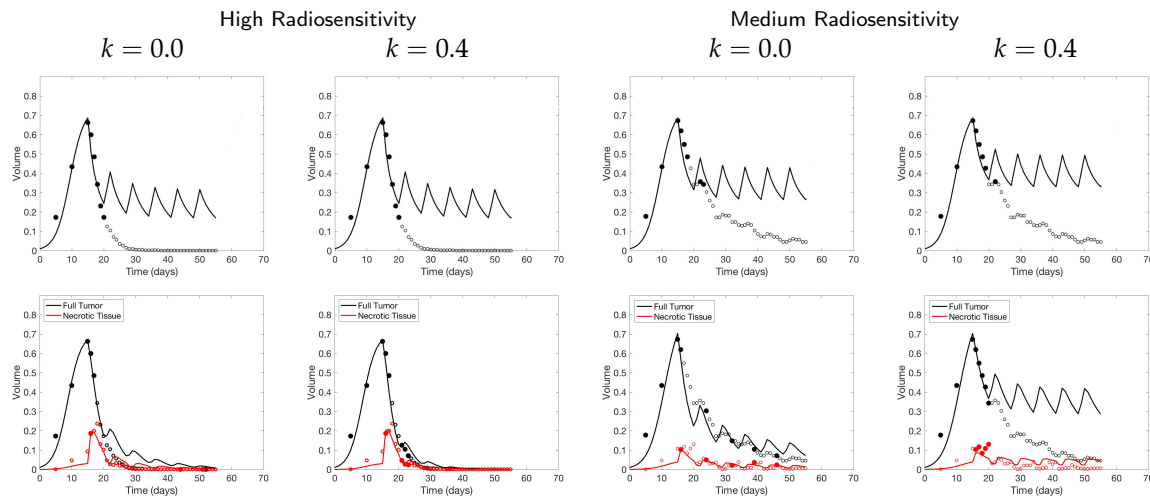
Figure 14. Comparison of data calibration between the one-compartment model (2) (**top**) and two-compartment model (3) (**bottom**) using a budget of 6 scans. The selected scan of tumor volume data (●) and necrotic volume data (●) among the potential data (○, ○) and the fitted model results are plotted. The shown results show how collecting both tumor volume and necrotic fraction and using the two-compartment model calibration can improve the model prediction significantly from the one compartment model. In particular, using $k = 0.4$ in high radiosensitivity and using $k = 0$ in medium radiosensitivity show the best calibrated results.
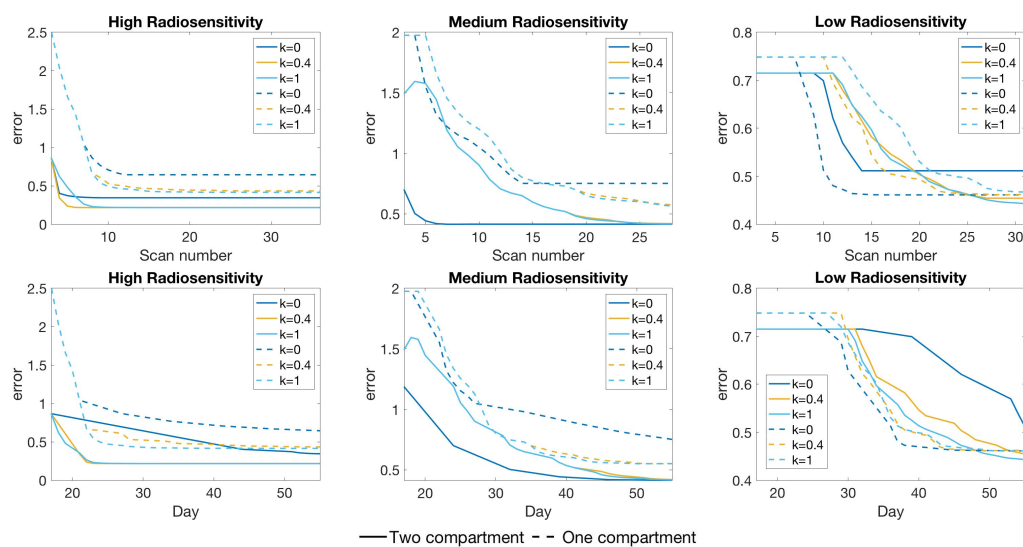

Figure 15. Comparison of accuracy between the one-compartment model calibration (2) and two-compartment model calibration (3). The shown result compares the error (10) with respect to scan number (**top**) and time in days (**bottom**). Using the same number of scans, the two compartment model show more accurate results in high and medium radiosensitivity, especially using our recommended k value. In case of low radiosensitivity, the one-compartment model with $k = 0$ shows better accuracy for scan budgets around 10 to 20.

4.5. Summary of Scan Schedule Recommendations

In what follows, we summarize the major observations from our simulation results in a clinically relevant manner. However, we note that these observations are based on very specific scenarios; results may differ when this methodology is applied to different tumor growth models, treatment regimens, or data collection protocols.

In our first scenario, which investigated this framework under the assumption that a clinician could collect tumor volume data once per week, our results suggest that the scan should be taken

on the first day of treatment for the first three weeks. At this point, if roughly half or more of the pre-treatment tumor volume still remains, it would be beneficial to switch to day 6 (Saturday) for weekly data collection, in order to maximize the mutual information provided by each scan.

However, if clinicians are not restricted to taking one scan per week, as in Scenarios 2–4, we use our score function, $S_k(i, r)$—defined in Equation (9)—to enforce a penalty for choosing later scans. Our results suggest that in the the high radiosensitivity case, large values of penalty parameter $k \approx 1$ provide the optimal scan schedule for any scan budget number, allowing for the inclusion of numerous scans early in the treatment schedule. In the case of low radiosensitivity, small values of $k = 0$ give the optimal scan schedule, especially when the scan budget is small—the use of a small k allows the algorithm to skip ahead and obtain data toward the end of the treatment protocol so that an overall data trend can be captured. Such a score function recommends data collection largely on the first and last days of the week, skipping most intermediate time points. In the case of medium radiosensitivity, $k = 0$ is optimal when the scan budget is less than 15, with larger k values favored to improve accuracy as more scans are available.

In general, the optimal choice of k is highly dependent upon the shape of the data; data with a rapidly changing gradient (i.e., high radiosensitivity) will require different handling of the scanning schedule than data that is nearly constant, such as our low radiosensitivity scenario. This suggests that we might benefit from updating k as we learn about a patient's sensitivity to radiotherapy, thus gaining information about the trend of their data. Though this will require an abundance of future work to investigate in depth, we give the following examples of the style of suggestion that could be offered to clinicians, once this framework is fully adapted to a clinical setting:

If only total tumor volume is measured, then for a small scan budget, our results recommend to start with a score function with small k, within $0 \leq k \leq 0.3$. Then, if the patient is highly responsive to radiotherapy, increase $k \geq 0.5$, or if the patient is less responsive to radiotherapy, reduce to $k = 0$. When the budget of total scans is high (for example, more than 15), we suggest using large k in the range of $k > 0.3$, and then to increase k further if the patient is highly responsive to radiotherapy.

If both the tumor volume and necrotic volume can be measured, then for a small scan budget, our results suggest to start with a score function with parameter $k \approx 0.2$. Similarly to the one-compartment case, we might further recommend increasing k for highly responsive patients, or reducing k to $k = 0$ for less responsive patients. For a scan budget at or above 15 scans, we suggest using a non-zero k, for instance, $k \geq 0.2$, in all scenarios.

5. Conclusions

In summary, we have illustrated a Bayesian information-theoretic framework for determining the optimal sampling of time-series data in order to accurately calibrate low-fidelity models for use in clinical decision-making. We applied this methodology to one- and two-compartment ODE models of tumor response to radiotherapy, calibrated using synthetic data generated from a cellular automaton model, simulating tumors with varying radiosensitivity levels and parameterized using experimental data from the prostate cancer cell line, PC3.

We first enforced a budget of exactly one tumor volume scan per week, and used the mutual information between the low-fidelity model parameters and high-fidelity data to determine the optimal day within each week for data collection. Our results suggested that clinicians with a weekly scan budget should start by measuring the tumor volume on the first day of the weekly treatment regime. If the tumor does not respond well after three weeks of treatment, then scans should be collected on the sixth day of the treatment schedule for the final three weeks to garner additional information about the magnitude of tumor reduction over the course of one dosage cycle; otherwise, scans should continue to be collected on day 1.

In order to relax the scan schedule restrictions to allow for n scans collected at any time throughout the course of treatment, we developed and applied an algorithm relying on a score function,

which rewards the user for choosing a data point that yields a large mutual information between the low-fidelity model parameters and high-fidelity data, and enforces a penalty with weight k for skipping forward to later time points. We tested this algorithm on both the one- and two-compartment models; for the latter, we first incorporated a single metric at each step, and then repeated the analysis in a more practical setting by providing the total tumor volume whenever the necrotic volume metric was chosen. We assessed the predictive power of the model resulting from the calibration algorithm using the error between all high-fidelity data points and the low-fidelity model approximation.

In all cases, we observed the algorithm favoring a larger penalty parameter, k, for tumors with a high level of radiosensitivity and any scan budget size; for cases such as this, the accuracy of the model benefits from the inclusion of data at numerous early time points. For tumors with low and medium levels of radiosensitivity and a small scan budget, the calibration was most accurate when using small k values, allowing for samples interspersed throughout the full treatment schedule and observance of the overall data trend. However, in these lower radiosensitivity cases, larger k values became more favorable as the scan budget increased, allowing for the inclusion of more data. from early in the treatment regimen. When calibrating the two-compartment model, the algorithm largely preferred necrotic data to tumor volume, with some exceptions in the high radiosensitivity case. Overall, we observed a smaller prediction error when using the two-compartment model than when using the one-compartment model, due to the additional information provided by the necrotic fraction.

Although this work used only fixed values of the penalization parameter k, as future work we propose the development of an adaptive framework to alter k throughout the course of the treatment regimen as we gain information about a patient's sensitivity to treatment. Additionally, we plan to further investigate robustness of the optimal measurement protocol on different potential patients, based not only on radiosensitivity, but also in terms of other characteristics—including tumor aggressiveness and micro-environment—and study the impact of measurement errors on optimal measurement protocol. The results in this work are highly dependent upon the models, treatment regimen, and data collection protocols chosen. In particular, more work must be done to investigate which metrics are most informative and practical for use in a clinical setting and to incorporate metrics from imaging techniques developed for early assessment of tumor response [31–34]. Having illustrated the efficacy of the methodology from a mathematical standpoint in an highly-idealized setting, we next plan to work with experimentalists to test the robustness of our framework in a messy real-world setting, by incorporating the effects of issues such as measurement noise, tumor response lag times, and clinical budget constraints. Other points of future interest include the expansion of this work to models that can incorporate two less closely-related metrics—such as tumor volume and immune cell count—as well as models containing other treatment types, such as chemotherapy, immunotherapy, or some combination of these modalities with radiotherapy. We plan to use such extensions of our scanning framework to help inform decision-making about adjusting therapy types and schedules during tumor treatment at the clinical level.

Abbreviations

The following abbreviations are used in this manuscript:

ODE	Ordinary Differential Equation
L-Q	Linear-Quadratic
RT	Radiotherapy

CA   Cellular Automaton
*k*NN   *k*th-Nearest-Neighbor
DRAM   Delayed Rejection Adaptive Metropolis

Appendix A. Parameter Values

Table A1. A summary of the parameters used in the CA Model and their default values. Parameter values are estimated using experimental data from the prostate cancer cell line, PC3, in [14].

Parameter	Description	Value	Units
l	Cell size	0.0018	cm
L	Domain length	0.36	cm
$\bar{\tau}_{cycle}$	Mean (standard deviation) cell cycle time	18.3 (1.4)	h
c_∞	Background O_2 concentration	2.8×10^{-7}	mol cm^{-3}
D	O_2 diffusion constant	1.8×10^{-5}	cm^{2}s^{-1}
c_Q	O_2 concentration threshold for proliferating cells	1.82×10^{-7}	mol cm^{-3}
c_N	O_2 concentration threshold for quiescent cells	1.68×10^{-7}	mol cm^{-3}
κ_P	O_2 consumption rate of proliferating cells	1.0×10^{-8}	mol cm^{-3}s^{-1}
κ_Q	O_2 consumption rate of quiescent cells	5.0×10^{-9}	mol cm^{-3}s^{-1}
p_{NR}	Rate of lysis of necrotic cells	0.015	hr^{-1}

Table A2. A summary of the fixed pre-treatment parameter values used for low-fidelity model simulations. Throughout the investigation, we assume that pre-treatment parameters have been identified prior to the administration of RT, and leave the work of ensuring pre-treatment parameter identifiability during calibration to a future investigation.

Parameter	One-Compartment Model			Two-Compartment Model		
	High	Medium	Low	High	Medium	Low
A	0.4584	0.4584	0.4584			
B	0.6213	0.6213	0.6213			
λ				0.5193	0.4845	0.6236
k				0.7201	0.9529	0.9031
η				0.0366	0.0828	0.1762
ζ				0.7270	1.3476	1.6638

Appendix B. *K*Nn Estimate of Mutual Information

Our definition of mutual information in Equation (8) requires an integral that often cannot be evaluated directly and may be prohibitively expensive, even using many numerical methods. As such, we utilize the *k*NN (*k*th-Nearest-Neighbor) estimate of mutual information to make our procedure computationally feasible. This estimate (and a proof of convergence) was fully developed in [23]—below, we summarize the major details.

To begin, at each calibration step, we estimate the posterior distribution $p(\theta|D_{n-1})$ using the Delayed Rejection Adaptive Metropolis (DRAM) algorithm for model calibration [44]. This variation on a traditional Metropolis-Hastings algorithm includes an additional two steps; an adaptation step that allows for periodic updating of the covariance matrix as information is learned about parameter dependencies, and a delayed rejection step, whereby the algorithm avoids stagnating for long periods of time by replacing rejected parameter set candidates with an alternate candidate chosen from a narrowed proposal distribution. We refer the interested reader to [44,45] for additional details about the DRAM algorithm, and to [27] for more information on how this calibration procedure fits into our mutual information framework.

To estimate the mutual information between the low-fidelity model parameters and a specified design condition ξ_n, we draw M samples from the prior distribution $p(\theta|D_{n-1})$ computed via our Metropolis algorithm, and append low-fidelity model estimates to each parameter set drawn—this results in a chain $X = (\theta, d_n(\xi_n))$ of size $M \times (p+1)$. We satisfy the kNN algorithm assumption of independent parameter samples by systematically thinning our dependent Metropolis chains, selecting every 10th sample in the posterior distribution to include in our set of M samples.

Following the procedure laid out in [23], for each chain element X_i, we compute the distance $\epsilon(i)/2 = ||X_i - X_{k(i)}||_\infty$, where $X_{k(i)}$ represents the kth-nearest neighbor to X_i in the chain $\{X_i\}_{i=1}^M$ (determined using the supremum norm). We determine the number of points in each marginal subspace n_θ and n_d that lie within $\epsilon(i)/2$ of the projected point; see Figure A1 for an example of this computation. The mutual information can approximated by

$$I(\theta; d_n | D_{n-1}, \xi_n) \approx \psi(k) - \frac{1}{M} \left[\sum_{i=1}^M \psi(n_\theta(i) + 1) + \sum_{i=1}^M \psi(n_d(i) + 1) \right] + \psi(M),$$

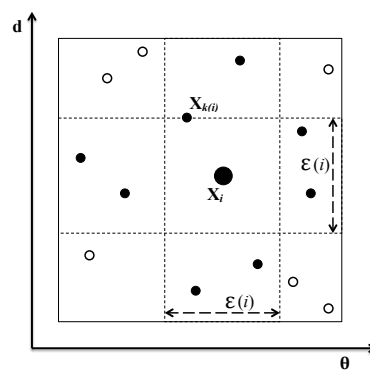where $\psi(\cdot)$ is the digamma function.

Figure A1. Calculation of $\epsilon(i)$, $n_\theta(i)$, and $n_d(i)$ for the case $k = 1$ from [23]. Here we have $n_\theta(i) = 3$ and $n_d(i) = 4$. Note that the kth-nearest neighbor is not included in the determination of n_θ and n_d.

For this study, we estimate mutual information using $k = 6$, as suggested in [24], though we note that one could calculate the mutual information for a vector of different k values and choose the value that yields the maximum mutual information.

References

1.   Altrock, P.M.; Liu, L.L.; Michor, F. The mathematics of cancer: Integrating quantitative models. *Nat. Rev. Cancer* **2015**, *15*, 730–745. [PubMed]
2.   Lavi, O.; Gottesman, M.M.; Levy, D. The dynamics of drug resistance: A mathematical perspective. *Drug Resist. Updat.* **2012**, *15*, 90–97. [CrossRef] [PubMed]
3.   Swierniak, A.; Kimmel, M.; Smieja, J. Mathematical modeling as a tool for planning anticancer therapy. *Eur. J. Pharmacol.* **2009**, *625*, 108–121. [CrossRef] [PubMed]
4.   Byrne, H.M. Dissecting cancer through mathematics: From the cell to the animal model. *Nat. Rev. Cancer* **2010**, *10*, 221–230. [PubMed]
5.   Rockne, R.C.; Hawkins-Daarud, A.; Swanson, K.R.; Sluka, J.P.; Glazier, J.A.; Macklin, P.; Hormuth, D.A.; Jarrett, A.M.; Lima, E.A.; Tinsley Oden, J.; et al. The 2019 mathematical oncology roadmap. *Phys. Biol.* **2019**, *16*, 1–33. [CrossRef]
6.   Chambers, D.A.; Amir, E.; Saleh, R.R.; Rodin, D.; Keating, N.L.; Osterman, T.J.; Chen, J.L. The Impact of Big Data Research on Practice, Policy, and Cancer Care. *Am. Soc. Clin. Oncol. Educ. Book* **2019**, *39*, e167–e175. [CrossRef]
7.   Bibault, J.E.; Giraud, P.; Burgun, A. Big Data and machine learning in radiation oncology: State of the art and future prospects. *Cancer Lett.* **2016**, *382*, 110–117. [CrossRef]

8. Murphy, H.; Jaafari, H.; Dobrovolny, H.M. Differences in predictions of ODE models of tumor growth: A cautionary example. *BMC Cancer* **2016**, *16*, 1–10. [CrossRef]

9. Collis, J.; Connor, A.J.; Paczkowski, M.; Kannan, P.; Pitt-Francis, J.; Byrne, H.M.; Hubbard, M.E. Bayesian Calibration, Validation and Uncertainty Quantification for Predictive Modelling of Tumour Growth: A Tutorial. *Bull. Math. Biol.* **2017**, *79*, 939–974. [CrossRef]

10. Koziol, J.A.; Falls, T.J.; Schnitzer, J.E. Different ODE models of tumor growth can deliver similar results. *BMC Cancer* **2020**, *20*, 1–10. [CrossRef]

11. Liu, Y.; Purvis, J.; Shih, A.; Weinstein, J.; Agrawal, N.; Radhakrishnan, R. A multiscale computational approach to dissect early events in the erb family receptor mediated activation, differential signaling, and relevance to oncogenic transformations. *Ann. Biomed. Eng.* **2007**, *35*, 1012–1025. [CrossRef] [PubMed]

12. Ramis-Conde, I.; Chaplain, M.A.; Anderson, A.R.; Drasdo, D. Multi-scale modelling of cancer cell intravasation: The role of cadherins in metastasis. *Phys. Biol.* **2009**, *6*. [CrossRef] [PubMed]

13. Hawkins-Daarud, A.; Prudhomme, S.; van der Zee, K.G.; Oden, J.T. Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth. *J. Math. Biol.* **2013**, *67*, 1457–1485. [CrossRef] [PubMed]

14. Kannan, P.; Paczkowski, M.; Miar, A.; Owen, J.; Kretzschmar, W.; Lucotti, S.; Kaeppler, J.; Chen, J.; Markelc, B.; Kunz-Schughart, L.; et al. Radiation resistant cancer cells enhance the survival and resistance of sensitive cells in prostate spheroids. *bioRxiv* **2019**. [CrossRef]

15. Cho, H.; Lewis, A.; Storey, K.; Jennings, R.; Shtylla, B.; Reynolds, A.; Byrne, H. A framework for performing data-driven modeling of tumor growth with radiotherapy treatment. In *Springer Special Issue: Using Mathematics to Understand Biological Complexity, Women in Mathematical Biology*; Springer: Berlin/Heidelberg, Germany, 2020; accepted.

16. Thames, H.D.; Rodney Withers, H.; Peters, L.J.; Fletcher, G.H. Changes in early and late radiation responses with altered dose fractionation: Implications for dose-survival relationships. *Int. J. Radiat. Oncol. Biol. Phys.* **1982**, *8*, 219–226. [CrossRef]

17. Fowler, J.F. The linear-quadratic formula and progress in fractionated radiotherapy. *Br. J. Radiol.* **1989**, *62*, 679–694. [CrossRef]

18. Rockne, R.; Rockhill, J.K.; Mrugala, M.; Spence, A.M.; Kalet, I.; Hendrickson, K.; Lai, A.; Cloughesy, T.; Alvord, E.C.; Swanson, K.R. Predicting the efficacy of radiotherapy in individual glioblastoma patients in vivo: A mathematical modeling approach. *Phys. Med. Biol.* **2010**, *55*, 3271–3285. [PubMed]

19. Corwin, D.; Holdsworth, C.; Rockne, R.C.; Trister, A.D.; Mrugala, M.M.; Rockhill, J.K.; Stewart, R.D.; Phillips, M.; Swanson, K.R. Toward patient-specific, biologically optimized radiation therapy plans for the treatment of glioblastoma. *PLoS ONE* **2013**, *8*. [CrossRef]

20. Sunassee, E.D.; Tan, D.; Ji, N.; Brady, R.; Moros, E.G.; Caudell, J.J.; Yartsev, S.; Enderling, H. Proliferation saturation index in an adaptive Bayesian approach to predict patient-specific radiotherapy responses. *Int. J. Radiat. Biol.* **2019**, *95*, 1421–1426. [CrossRef]

21. Enderling, H.; Alfonso, J.C.L.; Moros, E.; Caudell, J.J.; Harrison, L.B. Integrating Mathematical Modeling into the Roadmap for Personalized Adaptive Radiation Therapy. *Trends Cancer* **2019**, *5*, 467–474. [CrossRef]

22. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]

23. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138. [CrossRef] [PubMed]

24. Terejanu, G.; Upadhyay, R.R.; Miki, K. Bayesian experimental design for the active nitridation of graphite by atomic nitrogen. *Exp. Therm. Fluid Sci.* **2012**, *36*, 178–193. [CrossRef]

25. Bryant, C.; Terejanu, G. An information-theoretic approach to optimally calibrate approximate models. In Proceedings of the 50th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, Nashville, TN, USA, 9–12 January 2012. [CrossRef]

26. Liepe, J.; Filippi, S.; Komorowski, M.; Stumpf, M.P. Maximizing the Information Content of Experiments in Systems Biology. *PLoS Comput. Biol.* **2013**, *9*, 1–13. [CrossRef] [PubMed]

27. Lewis, A.; Smith, R.; Williams, B.; Figueroa, V. An information theoretic approach to use high-fidelity codes to calibrate low-fidelity codes. *J. Comput. Phys.* **2016**, *324*, 24–43. [CrossRef]

28. Lorenzo, G.; Pérez-García, V.M.; Mariño, A.; Pérez-Romasanta, L.A.; Reali, A.; Gomez, H. Mechanistic modelling of prostate-specific antigen dynamics shows potential for personalized prediction of radiation therapy outcome. *J. R. Soc. Interface* **2019**, *16*, 1–13. [CrossRef]

29. Walker, A.J.; Ruzevick, J.; Malayeri, A.A.; Rigamonti, D.; Lim, M.; Redmond, K.J.; Kleinberg, L. Postradiation imaging changes in the CNS: How can we differentiate between treatment effect and disease progression? *Future Oncol.* **2014**, *10*, 1277–1297. [CrossRef]

30. Ghaye, B.; Wanet, M.; El Hajjam, M. Imaging after radiation therapy of thoracic tumors. *Diagn. Interv. Imaging* **2016**, *97*, 1037–1052. [CrossRef]

31. Rashidian, M.; Keliher, E.J.; Bilate, A.M.; Duarte, J.N.; Wojtkiewicz, G.R.; Jacobsen, J.T.; Cragnolini, J.; Swee, L.K.; Victora, G.D.; Weissleder, R.; et al. Noninvasive imaging of immune responses. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6146–6151. [CrossRef]

32. Shuhendler, A.J.; Ye, D.; Brewer, K.D.; Bazalova-Carter, M.; Lee, K.H.; Kempen, P.; Dane Wittrup, K.; Graves, E.E.; Rutt, B.; Rao, J. Molecular magnetic resonance imaging of tumor response to therapy. *Sci. Rep.* **2015**, *5*, 1–14. [CrossRef]

33. Kasoji, S.K.; Rivera, J.N.; Gessner, R.C.; Chang, S.X.; Dayton, P.A. Early assessment of tumor response to radiation therapy using high-resolution quantitative microvascular ultrasound imaging. *Theranostics* **2018**, *8*, 156–168. [CrossRef] [PubMed]

34. Zhou, Z.; Deng, H.; Yang, W.; Wang, Z.; Lin, L.; Munasinghe, J.; Jacobson, O.; Liu, Y.; Tang, L.; Ni, Q.; et al. Early stratification of radiotherapy response by activatable inflammation magnetic resonance imaging. *Nat. Commun.* **2020**, *11*, 1–12. [CrossRef] [PubMed]

35. Hall, E.J. *Radiobiology for the Radiologist*; J.B. Lippincott: Philadelphia, PA, USA, 1994.

36. Enderling, H.; Chaplain, M.A.; Hahnfeldt, P. Quantitative Modeling of Tumor Dynamics and Radiotherapy. *Acta Biotheor.* **2010**, *58*, 341–353. [CrossRef] [PubMed]

37. Lewin, T.D. Modelling the Impact of Heterogeneity in Tumor Composition on the Response to Fractionated Radiotherapy. Ph.D. Thesis, University of Oxford, Oxford, UK, 2018.

38. Lewin, T.D.; Byrne, H.M.; Maini, P.K.; Caudell, J.J.; Moros, E.G.; Enderling, H. The importance of dead material within a tumour on the dynamics in response to radiotherapy. *Phys. Med. Biol.* **2020**, *65*, 015007. [CrossRef]

39. Lewin, T.D.; Maini, P.K.; Moros, E.G.; Enderling, H.; Byrne, H.M. A three phase model to investigate the effects of dead material on the growth of avascular tumours. *Math. Model. Nat. Phenom.* **2020**, *15*, 1–29. [CrossRef]

40. Lea, D.E.; Catcheside, D.G. The mechanism of the induction by radiation of chromosome aberrations in Tradescantia. *J. Genet.* **1942**, *44*, 216–245. [CrossRef]

41. Pérez-García, V.M.; Bogdanska, M.; Martínez-González, A.; Belmonte-Beitia, J.; Schucht, P.; Pérez-Romasanta, L.A. Delay effects in the response of low-grade gliomas to radiotherapy: A mathematical model and its therapeutical implications. *Math. Med. Biol.* **2015**, *32*, 307–329. [CrossRef]

42. Prokopiou, S.; Moros, E.G.; Poleszczuk, J.; Caudell, J.; Torres-Roca, J.F.; Latifi, K.; Lee, J.K.; Myerson, R.; Harrison, L.B.; Enderling, H. A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation. *Radiat. Oncol.* **2015**, *10*, 1–8. [CrossRef]

43. Poleszczuk, J.; Walker, R.; Moros, E.G.; Latifi, K.; Caudell, J.J.; Enderling, H. Predicting Patient-Specific Radiotherapy Protocols Based on Mathematical Model Choice for Proliferation Saturation Index. *Bull. Math. Biol.* **2018**, *80*, 1195–1206. [CrossRef]

44. Haario, H.; Laine, M.; Mira, A.; Saksman, E. DRAM: Efficient adaptive MCMC. *Stat. Comput.* **2006**, *16*, 339–354. [CrossRef]

45. Smith, R. *Uncertainty Quantification: Theory, Implementation, and Applications*; SIAM: Philadelphia, PA, USA, 2014.