

Mathematical modelling of the dynamics of image-informed tumor habitats in a murine model of glioma

Kalina P. Slavkova

The University of Texas at Austin

Sahil H. Patel

Case Western Reserve University

Zachary Cacini

University of Illinois Urbana-Champaign

Anum S. Kazerouni

The University of Washington

Andrea Gardner

The University of Texas at Austin

Thomas E. Yankeelov

The University of Texas at Austin

David A. Hormuth, II (✉ david.hormuth@austin.utexas.edu)

The University of Texas at Austin

Article

Keywords: diffusion-weighted MRI, dynamic contrast-enhanced MRI, quantitative MRI, mathematical model, computational oncology, habitat imaging, ODE model

Posted Date: September 15th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2058436/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Mathematical modelling of the dynamics of image-informed tumor habitats in a murine model of glioma

Kalina P. Slavkova¹, Sahil H. Patel², Zachary Cacini³, Anum S. Kazerouni⁴,
Andrea Gardner⁵, Thomas E. Yankeelov⁵⁻¹⁰, David A. Hormuth, II^{8,9*}

¹ Department of Physics, The University of Texas at Austin, Austin, TX, USA

² Case Western Reserve University, Cleveland, OH, USA

³ University of Illinois, Urbana-Champaign, IL, USA

⁴ Department of Radiology, The University of Washington, Seattle, WA, USA

⁵ Department of Biomedical Engineering, The University of Texas at Austin, Austin, USA

⁶ Department of Diagnostic Medicine, The University of Texas at Austin, Austin, TX, USA

⁷ Department of Oncology, The University of Texas at Austin, Austin, TX, USA

⁸ The Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA

⁹ Livestrong Cancer Institutes, The University of Texas at Austin, Austin, TX, USA

¹⁰ Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

*Please address correspondence to:

David A. Hormuth II, Ph.D.

The University of Texas at Austin

Oden Institute for Computational Engineering and Sciences

201 E 24th Street

Austin, TX 78712, USA

E-mail: david.hormuth@austin.utexas.edu

Keywords: diffusion-weighted MRI, dynamic contrast-enhanced MRI, quantitative MRI, mathematical model, computational oncology, habitat imaging, ODE model

Word count: 4,786

ABSTRACT

Tumors are highly heterogeneous with unique sub-regions termed “habitats”. We evaluate the ability of a mathematical model built on coupled ordinary differential equations (ODEs) to describe and predict tumor habitat dynamics in a murine model of glioma. Female Wistar rats (N=21) were inoculated intracranially with 10^6 C6 glioma cells, a subset of which received 20 (N=5) or 40 Gy (N=8) of radiation. All rats underwent diffusion-weighted (DW) and dynamic contrast-enhanced magnetic (DCE) resonance imaging (MRI) at up to seven time points. All MRI data at each visit were subsequently clustered using *k*-means to identify physiological tumor habitats. A family of four models consisting of three coupled ODEs were developed and calibrated to the habitat time series of eight control rats and eight treated rats and evaluated for predictive capability. The Akaike Information Criterion (AIC) was used for model selection, and the normalized sum-of-square-error (SSE) was used to evaluate goodness-of-fit in model calibration and prediction. Three tumor habitats with significantly different imaging data characteristics ($p < 0.05$) were identified: high-vascularity high-cellularity, low-vascularity high-cellularity, and low-vascularity low-cellularity. Model selection yielded a five-parameter model whose predictions of habitat dynamics yielded SSEs that were similar to the SSEs from the calibrated model. It is thus feasible to mathematically describe habitat dynamics in a preclinical model of glioma using biology-based ODEs, showing promise for forecasting heterogeneous tumor behavior.

1 INTRODUCTION

Gliomas are generally a neoplastic growth of the glial cells which serve supportive functions in the central and peripheral nervous systems¹. High-grade gliomas, such as glioblastoma, are extremely aggressive forms that have a high recurrence rate despite aggressive combinations of surgery, radiotherapy, and chemotherapy, yielding a 5-year-survival rate of only 7.1%². It is well established that the aggressiveness of glioblastomas is largely attributed to intra-tumor heterogeneity³; namely, phenotypic and genotypic differences that lead to variations in proliferation rates and treatment sensitivities across the tumor^{3,4}. Intra-tumor heterogeneity arises from a diverse array of tumor sub-populations as well as stromal cells that infiltrate the tumor and become a part of the tumor microenvironment⁴. As tumor cells divide and evolve, tumors become more heterogeneous, leading to a differential response to therapies and, ultimately, treatment resistance⁵. In addition to heterogeneity arising from genotypic and phenotypic variables, heterogeneity in tumor vasculature also contributes to the mixed tumor landscape and substantially informs treatment outcomes. Previous work has described tumor heterogeneity through the investigation of tumor subregions, termed habitats⁶, with unique physiological features stemming from properties of the tumor vasculature and cellularity^{7,8}. These habitats can be non-invasively identified through multiparametric imaging for three-dimensional (3D) analysis of tumor heterogeneity.

Common methods for evaluating tumor biology involve invasive procedures, such as biopsies or tumor excisions⁹. These techniques are susceptible to sampling error given that they interrogate only a small portion of a heterogeneous tumor, making it unlikely that they provide an accurate description of the whole lesion^{9,10}. Additionally, invasive methods are fundamentally limited in their ability to characterize dynamic events such as vascular perfusion or metabolic activity. Conversely, quantitative magnetic resonance imaging (MRI) allows for noninvasive 3D measurements of physiological tissue characteristics and, therefore, enables measurement at multiple time points¹¹. In particular, diffusion-weighted MRI (DW-MRI)¹² and dynamic contrast-enhanced (DCE-) MRI¹³ are two prominent acquisition techniques that quantitatively assess tissue cellularity and vascularity, respectively, which are tumor features that are known to be dramatically altered in high-grade glioma¹⁴.

Tissue cellularity is quantified *via* DW-MRI, which is sensitive to the movement of water molecules within tissue. DW-MRI can be analyzed to yield an estimate of the apparent diffusion coefficient (ADC , units of mm^2/s) of water in tissue, which has been established as having an inverse relationship with the level of cellularity in tumors¹². Increased cellularity results in an increase in the number of barriers to water movement and therefore a reduction in the ADC . Decreased ADC values have been associated with poorer survival¹⁵, while increased ADC values have been associated with positive response to chemotherapy¹⁵⁻¹⁷.

In DCE-MRI, a subject is injected with a gadolinium-based contrast agent, and images are serially collected before and after contrast delivery. This dynamic data can then be analyzed with (for example) the Kety-Tofts perfusion model¹⁸ to yield quantitative parameters of the volume transfer constant, K^{trans} (units of min^{-1} , a mixed measure of vessel permeability and perfusion), and v_e (a unitless measure extracellular-extravascular volume fraction)¹³. K^{trans} and v_e have been found to characterize vascular properties of tumors¹⁴, and repeatability of these perfusion parameters has been assessed in healthy subjects and patients with high-grade glioma^{15,19}. Additionally, Aydin *et al.*¹⁴ found that K^{trans} and v_e showcase a strong positive correlation with brain tumor grade.

Typical approaches for employing MRI to quantify intra-tumoral heterogeneity include histogram analysis of quantitative parameter maps resulting from DW- and DCE-MRI data^{20,21}. However, this approach eliminates the spatial information available from the imaging data. Other efforts have been made to apply texture analysis and machine learning techniques to identify and enhance heterogeneity without loss of spatial information^{22,23}, but these methods assume that tumors are well-mixed heterogeneous bodies rather than possessing distinct sub-regions⁶, especially in cases of high-grade gliomas with surrounding edema²⁴. Recently, Kazerouni *et al.* demonstrated the ability to identify three distinct tumor habitats using multiparametric preclinical MRI data that were subsequently histologically validated in murine models of breast cancer.⁷ A follow-up study by the same team showed that two distinct tumor phenotypes emerged in which one phenotype exhibited greater sensitivity to treatment²⁵.

Because of the high rate of recurrence of high-grade glioma, treatment strategies are often tailored to each independent case²⁵. In recent years, we²⁶⁻²⁸ and others^{25,29,30} have leveraged

imaging-data to personalize partial differential equation (PDE) models, such as reaction-diffusion models, of glioma progression that predict glioma response to radiotherapy and provide an avenue for non-invasive treatment optimization. These PDE models have several benefits over simpler ordinary differential equation (ODE) models, namely providing spatially resolved predictions. It is non-trivial, however, to build parsimonious multi-species models with limited assumptions on how the tumor habitats interact. ODE models are conducive to data-driven modeling where assumptions can be relaxed, and data alone can be relied upon for parameterizing the modeling framework. While ODE models have been employed by others^{25,31,32}, these models either do not incorporate non-invasive imaging data or do not account for intra-tumor heterogeneity.

Characterization and analysis of tumor heterogeneity via quantitative imaging and mathematical modeling could enable treatment optimization to maximize the effect of therapy across the entire tumor. In this work, we developed a data-driven ODE model to describe the dynamics of tumor habitats, identified via quantitative MRI, with and without radiotherapy in a murine model of glioma. We demonstrate that the identified habitats corroborate previous results in murine models of cancer, in which each cluster exhibits a unique combination of vascularity and cellularity. A family of mathematical models consisting of ODEs is calibrated to the dynamic habitat data and evaluated using a normalized sum-of-square-error as an error metric. After quantitatively selecting the most parsimonious model, we investigated the ability of the chosen model to predict the growth of the tumor habitats across time using two separate approaches. We show that the selected model makes successful predictions of the habitat dynamics that agree with the calibrated model outputs.

2 METHODS

2.1 Animal model and magnetic resonance imaging

All experimental details were previously reported in Hormuth *et al.*^{26,27} For brevity, we include salient details of the methods below (see Supplemental Methods for complete details). All experimental procedures were approved by our Institutional Animal Care and Use Committee and were performed in accordance with relevant guidelines and regulations. This study is reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org/>). Female Wistar rats (N = 21) were inoculated intracranially with C6 glioma cells (1×10^5) through stereotaxic injection. The C6 glioma was chosen as it is widely used in preclinical studies in neuro-oncology and in

mathematical modelling due to its predictable and reliable growth patterns³³. Of the 21 rats, 8 were assigned to the control group, 8 received 40 Gy of radiation, and 5 received 20 Gy of radiation. Rats were imaged with DCE-MRI and DW-MRI protocols beginning 10 days post-inoculation and then every 1-2 days after that for a total of 5-7 MRI data sets per animal.

2.2 Computing multiparametric quantitative data

The DW- and DCE-MRI data from all control and treated rats were analyzed to extract four quantitative parameters that were subsequently clustered to identify tumor habitats (Figure 1A). For each rat, the tumor regions of interest (ROIs) were manually segmented using the DCE-MRI data. Nonlinear least squares methods were then used to fit the standard Kety-Tofts perfusion model^{13,36} to the DCE-MRI data within the tumor ROI, yielding three pharmacokinetic parameters: volume transfer constant (K^{trans} , in units of min^{-1}), extravascular/extracellular volume fraction (v_e), and k_{ep} ($=K^{trans}/v_e$, in units of min^{-1}). Next, the DW-MRI data were analyzed using standard methods^{12,40} to arrive at the ADC for each voxel within the tumor ROI. The result of this analysis of the multiparametric MRI data was a four-dimensional vector, $(K^{trans}, v_e, k_{ep}, ADC)$, at each voxel within the tumor ROI of each rat at each imaging visit. All analyses were performed using MATLAB (Mathworks, Natick, MA).

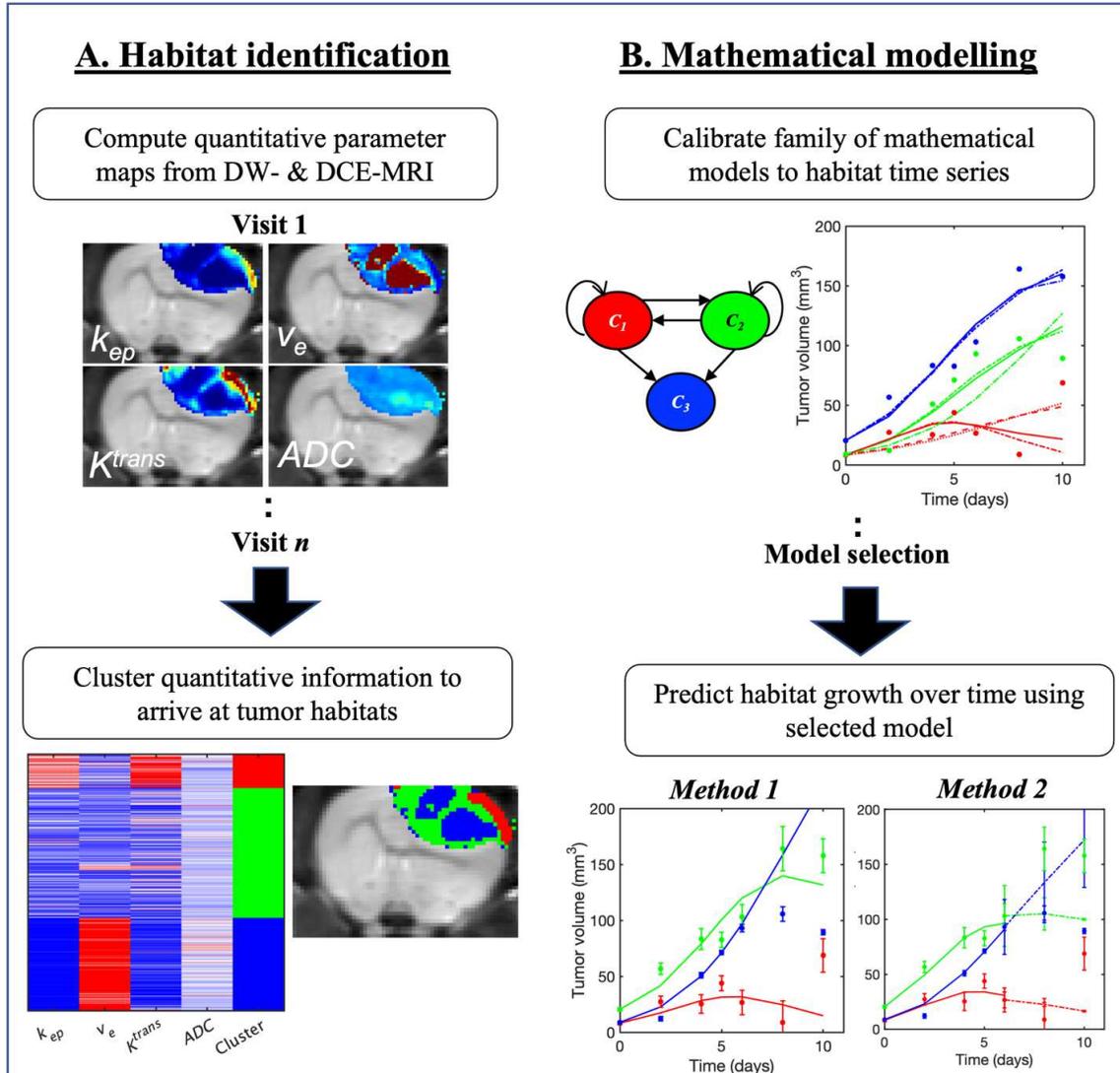


Figure 1. Habitat identification and modelling pipeline. (A) DCE-MRI and DW-MRI data were acquired for each animal over multiple days. The multiparametric MRI data were analyzed to generate four parameter maps (i.e., k_{ep} , v_e , K^{trans} , ADC) across the tumor ROI (top left). The quantitative information at each voxel at all imaging visits for all animals was pooled together and then clustered to identify three tumor habitats with unique physiologies: high-vascularity high-cellularity, low-vascularity high-cellularity, and low-vascularity low-cellularity (bottom left). (B) Once the habitats were identified, each tumor in each rat yielded three-habitat time series that could be analyzed by a family of mathematical models based on three compartments and the transition rates between them (top right). We then tested the ability of the most parsimonious model to predict habitat dynamics through leave-one-out (Method 1) and bootstrapping (Method 2) analyses (bottom right).

2.3 Identification of tumor habitats

All four-dimensional vectors from all voxels from all rat ROIs at all imaging visits were then pooled into one conglomerate matrix, where each column of the matrix represents one

parameter. Each column was normalized to the standard normal distribution in MATLAB to ensure that each parameter had equal weight when clustering⁴⁴. Next, we follow a similar clustering procedure as in previous work⁷ (using k -means⁴⁵ and agglomerative hierarchical clustering⁴⁶) to identify three tumor habitats in terms of the level of vascularity (characterized by K^{trans} and k_{ep}) and cellularity (characterized by ADC and v_e). These habitats are defined as “high-vascularity low-cellularity” (HV-LC), “low-vascularity high-cellularity” (LV-HC), and “low-vascularity low-cellularity” (LV-LC), based on previous work by Kazerouni *et al.*⁷ (See Supplemental Methods for further details). Once the habitats, have been appropriately labeled by their physiology, they are reorganized into a time series of volumes such that each rat now has an associated HV-HC, LV-HC, and LV-LC volume as a function of time. Because no spatial information was included in the clustering, we performed multiregional spatial interaction (MSI) matrix analysis⁴⁷ to evaluate if the identified clusters are spatially contiguous.

2.4 The mathematical model family

We develop a family of models to describe the dynamics of tumor habitats identified (Figure 1B) in the eight control (C) rats and in eight selected rats from the two treated (R) cohorts (See Table S1 for the imaging schedule and radiation dose for these 16 rats). The subset of eight treated rats was selected from the cohort of 13 total treated rats based on uniformity of experimental details, such as the imaging schedule and the execution of the imaging protocol. Given that the data to be modelled consists of three habitats as a function of time, it is natural to consider a three-species model of coupled ordinary differential equations (ODEs) (see Figure 2). This model is defined as follows:

$$V(t) = \frac{C_1(t) + C_2(t) + C_3(t)}{\theta} \quad (1)$$

$$\frac{dC_1(t)}{dt} = k_1 C_1(t)(1 - V(t)) - d_{12} C_1(t) + d_{21} C_2(t) - d_{13} C_1(t) \quad (2)$$

$$\frac{dC_2(t)}{dt} = k_2 C_2(t)(1 - V(t)) + d_{12} C_1(t) - d_{21} C_2(t) - d_{23} C_2(t) \quad (3)$$

$$\frac{dC_3(t)}{dt} = d_{13} C_1(t) + d_{23} C_2(t) \quad (4)$$

Equation (1) defines the total volume as a sum of the habitat volumes in relation to the carrying capacity, θ , where the functions $C_1(t)$, $C_2(t)$, and $C_3(t)$ represent the volumes (mm^3) of the HV-HC, LV-LC, and LV-LC habitats at time t , respectively. The k_n and d_{nm} refer to growth rates

(mm^3/day) of cluster n and decay rates (mm^3/day) from cluster n to cluster m , respectively (Table 1 defines the model parameters). We hypothesize that the HV-HC cluster will grow as the tumor recruits vasculature to its outer rim. As the tumor grows, former HV-HC voxels will become less perfused and transition to the LV-HC compartment (at the rate d_{12}). There is a non-zero probability that K^{trans} may rise in LV-HC voxels as more vasculature is recruited, leading to a transition to HV-HC membership. If no vasculature is recruited to the LV-HC region, the LV-HC voxels will eventually become necrotic and irreversibly transition to LV-LC membership.

Parameters	Description	Lower bound	Upper bound
k_1	Growth rate c_1 (HV-HC)	0	5
d_{12}	$c_1 \rightarrow c_2$ transition rate	0	5
d_{21}	$c_2 \rightarrow c_1$ transition rate	0	5
d_{13}	$c_1 \rightarrow c_3$ transition rate	0	5
k_2	Growth rate c_2 (LV-HC)	0	5
d_{23}	$c_2 \rightarrow c_3$ transition rate	0	5
θ	Carrying capacity	0	2000

Table 1. Parent model parameters and the bounds used in nonlinear fitting for model calibration. All growth rates, k_n , and transition rates, d_{nm} , are in units of mm^3/day , and the carrying capacity, θ , is in units of mm^3 .

Equations (1) – (4) describe the parent model, termed Model A. Based on the results in Section 3 below, we were motivated to generate three additional models to generate a family of models: Model A, Model B ($d_{21} = 0$), Model C ($d_{13} = 0$), and Model D ($d_{21} = d_{13} = 0$). Model B describes the low probability of the LV-HC habitat transitioning to HV-HC, and Model C describes the low probability of the HV-HC habitat directly undergoing necrosis and transitioning to LV-LC membership. Model D combines the assumptions of Models B and C.

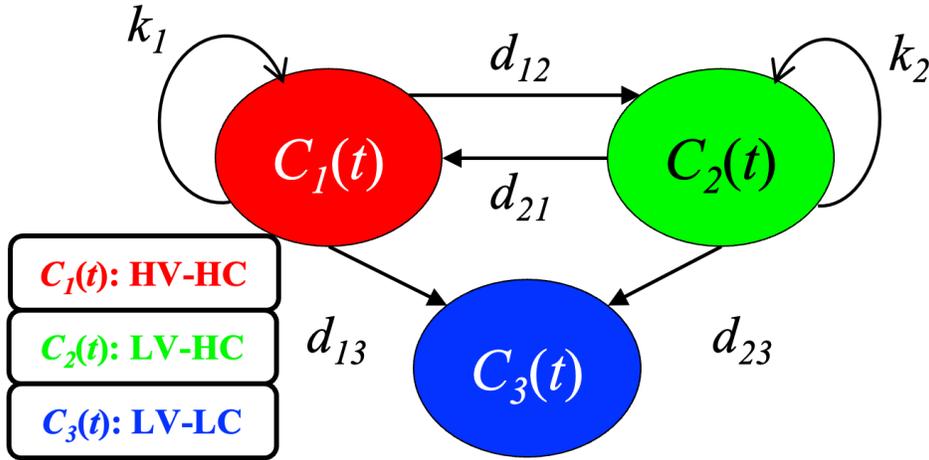


Figure 2. Visualizing a three-compartment mathematical model of tumor habitat growth. A physiological label can be attributed to each cluster in the parent model: $C_1(t)$ corresponds to the HV-HC habitat (red), $C_2(t)$ corresponds to the LV-HC habitat (green), and $C_3(t)$ corresponds to the LV-LC habitat (blue). The parameters k_1 and k_2 are the growth rates of $C_1(t)$ and $C_2(t)$, respectively, while all parameters of the form d_{nm} represent the transition rate from cluster n to cluster m . All model parameters are named and described in Table 1.

2.5 Model calibration to the habitat time series data

All four models were calibrated to the time series of tumor habitat volumes from each rat using the Levenberg-Marquardt method for nonlinear least-squares, implemented in MATLAB. MATLAB's *multistart* function was used to evaluate 50 initial guesses for the model parameters of each model independently. The lower and upper bounds for the parameters are available in Table 1. The initial conditions for $C_1(t)$, $C_2(t)$, and $C_3(t)$ were defined by the volumes of the HV-HC, LV-HC, and LV-LC habitats, respectively, at the first imaging visit on day 10 post-inoculation. The time vectors of the datasets are redefined such that the initial images correspond to day 0.

2.6 Model selection

The Akaike Information Criterion (AIC)⁴⁸, corrected for small sample sizes, was used to select the most parsimonious model from the family of four models. The mathematical expression for the AIC is:

$$AIC = T \ln \left(\frac{RSS}{T} \right) + 2k + \frac{2k(k+1)}{T-k-1}, \quad (5)$$

where T represents the total number of imaging visits, k is the total number of parameters in the model, and the RSS is the root sum-of-squares error. The four models were calibrated using the minimum number of imaging visits available ($T = 5$), and then the AIC was computed for each model calibration for each rat. The model with the smallest AIC value was selected as the most parsimonious model.

2.7 Prediction of tumor habitat dynamics

Two approaches were implemented to investigate the predictive ability of the model selected using the criterion introduced in Section 2.6. The first approach is the leave-one-out (LOO) method in which we leave out the time series of the j^{th} rat that we want to predict. We then compute the normalized weighted average of each model parameter of the remaining seven rats in the population *via* three different weighting schemes: initial tumor volume, initial tumor habitat composition, and initial HV-HC volume (chosen due to having the smallest interquartile range in the treated cohort).

The second approach for evaluating the predictive ability is bootstrapping⁴⁹ (BS). Like the LOO approach, we take out rat j and compute the mean and standard deviation of the model parameters of the remaining rats in the population. Then, we generate a normal distribution from the mean and standard deviation and sample this distribution N_b times between the 25th and 75th quartiles. We run the model forward for each set of sampled parameters and arrive at N_b number of model curves, from which we compute the mean sampled curve and confidence intervals. This is the predicted fit for rat j when no habitat volumes beyond the initial visit are made available. As data becomes available, we update the model fit to by taking a weighted average of the mean model curve and rat j 's individually calibrated fit to the available data. We label this approach as bootstrapping with data updating. (See Supplemental Methods for mathematical details of both approaches)

2.8 Statistical analysis

To determine significant differences in MRI parameter distributions between habitats, we employed the Kolmogorov-Smirnov two-sample test at a 5% significance level as well as a one-way analysis of variance (ANOVA). To evaluate the performance of each model during calibration, we introduce the normalized sum of square error (SSE), which is the sum of the square

of the differences between the data and the model output for each habitat in each rat’s time series, normalized to the maximum volume of each habitat time series. The Wilcoxon rank sum test at a 5% significance level was used to evaluate significant differences in the SSE distributions of each habitat across the family of models to determine if any given model described a particular habitat with lower error. The concordance correlation coefficient (CCC)⁵⁰ was also used to assess the agreement between the model outputs and the measured data, where CCC values above 0.8 are indicative of strong agreement between two distributions. We also use the rank sum test to determine which prediction method (i.e., LOO or bootstrapping) yielded the lowest error in predicting growth of the tumor habitats in a validation dataset.

3 RESULTS

3.1 Classification of image-informed tumor habitats

Figure 1A provides an overview of the clustering pipeline, taking four MRI-informed parameter maps from each rat dataset and outputting three tumor habitats with distinct physiological characteristics, namely HV-HC, LV-HC, and LV-LC. We found that k -means clustering yielded habitats with greater separation between clusters, in terms of higher mean values of K^{trans} for the HC-HV clusters compared to the agglomerative clustering. Moreover, qualitative inspection of the k -means clustering (Figure 3A) and the agglomerative clustering outputs (Figure S2A) shows that k -means collects stray voxels with high values of K^{trans} (colored in red) into the HV-HC habitat, whereas the agglomerative clustering grouped these voxels into the LV-HC cluster, yielding higher mean vascularity (Figure S2B, S2D) in the LV-HC habitat. Thus, we proceed with the habitats identified by the k -means clustering algorithm for the rest of this study.

Physiological characteristics were determined by assessing the mean values of the quantitative parameter distributions (Figure 3B–3E) within each habitat. The mean values of ADC and v_e were used to assess cellularity, while the mean values of K^{trans} and k_{ep} were used to assess vascularity. The first cluster had a mean $[ADC, v_e]$ of $[0.826 \times 10^{-3} (\pm 0.216) \text{ mm}^2/\text{s}, 0.232 (\pm 0.158)]$ and mean $[K^{trans}, k_{ep}]$ of $[0.307 (\pm 0.155) \text{ min}^{-1}, 1.625 (\pm 0.788) \text{ min}^{-1}]$. The second cluster had a mean $[ADC, v_e]$ of $[0.852 \times 10^{-3} (\pm 0.186) \text{ mm}^2/\text{s}, 0.181 (\pm 0.118)]$ and a lower mean $[K^{trans}, k_{ep}]$ of $[0.070 (\pm 0.054) \text{ min}^{-1}, 0.447 (\pm 0.310) \text{ min}^{-1}]$ compared to cluster 1. Finally, the third cluster had a substantially higher mean $[ADC, v_e]$ of $[0.945 \times 10^{-3} (\pm 0.280) \text{ mm}^2/\text{s}, 0.8480 (\pm 0.199)]$ and

significantly lower mean $[K^{trans}, k_{ep}]$ of $[0.0276 (\pm 0.034) \text{ min}^{-1}, 0.029 (\pm 0.097) \text{ min}^{-1}]$ compared to the other clusters. All quantitative parameter distributions were found to be significantly different between the three habitats with $p < 0.0001$ from all statistical tests (Section 2.7). Thus, cluster 1 ($C_1(t)$) was labelled as HV-HC as it had the highest values of vascularity-related parameters (K^{trans} and k_{ep} report on vascularity) and the lowest values of cellularity-related parameters (ADC and v_e are inversely proportional to cellularity). Cluster 2 ($C_2(t)$) was labelled as LV-HC as it was associated with lower K^{trans} and k_{ep} than cluster 1, and cluster 3 ($C_3(t)$) was labelled as LV-LC as it had the highest values of cellularity-related parameters and lowest measures of vascularity.

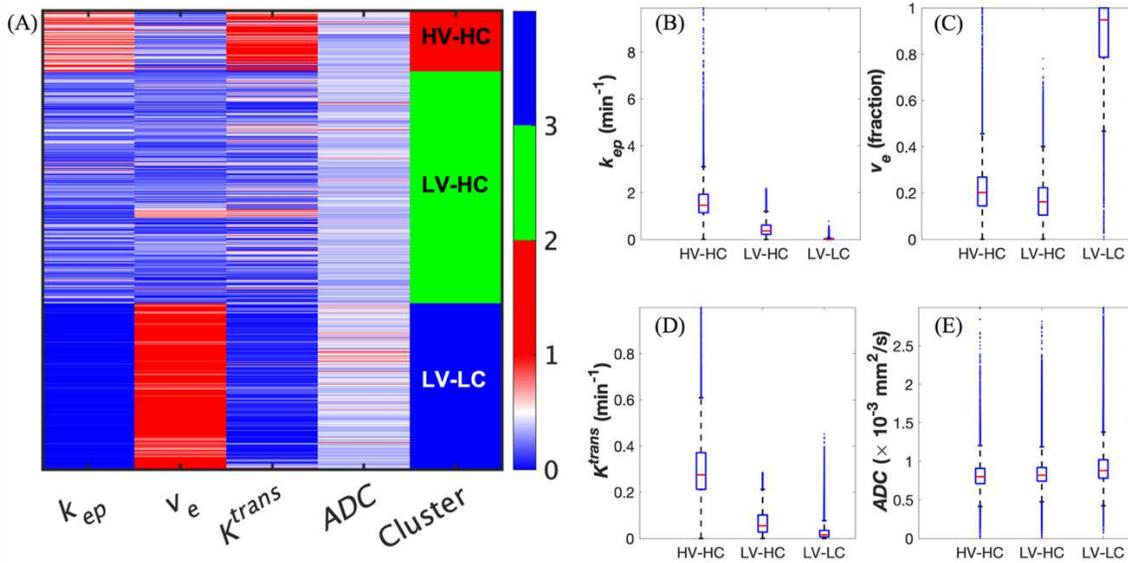


Figure 3: Identification of three tumor habitats with k -means clustering. (A) This is the heat map of normalized quantitative parameters grouped by the three identified clusters: high-vascularity high-cellularity (HV-HC), low-vascularity high-cellularity (LV-HC), and low-vascularity low-cellularity (LV-LC). (B) Boxplots of the distributions of k_{ep} are presented in each identified habit, where the red line in each boxplot represents the median of the distribution. Analogous to (B), panels (C)-(E) show the distributions of v_e , K^{trans} , and ADC , respectively. 168,207 voxels were included in this clustering analysis. We found that all four parameter distributions differ significantly between the three habitats, with $p < 0.05$.

Figure 4 shows the MRI parameter maps (Figure 4A–4D) alongside the corresponding habitat map (Figure 4E) for a representative rat imaged at visit 3. The distributions of k_{ep} , K^{trans} , v_e , and ADC are displayed in Figure 4F-4I, with statistically significant differences between the parameter distributions of the three habitats ($p < 0.001$). The highly perfused HV-HC habitat is at the well-vascularized edge of the tumor, while the LV-LC habitat corresponds to the necrotic

center of the tumor. The LV-HC habitat spatially separates the other two habitats. The habitat maps for two representative rats from the control and treated cohorts, respectively, are displayed in Figure S3 at multiple imaging visits. Taking the original habitat mapping (Figure S4A) and randomly permuting (scrambling) the voxels (Figure S4B), the MSI matrix can be computed for each case (Figure S4C). The MSI values were found to be significantly higher for the original mapping (Figure S4D) than the scrambled case, indicating that the habitats are spatially contiguous.

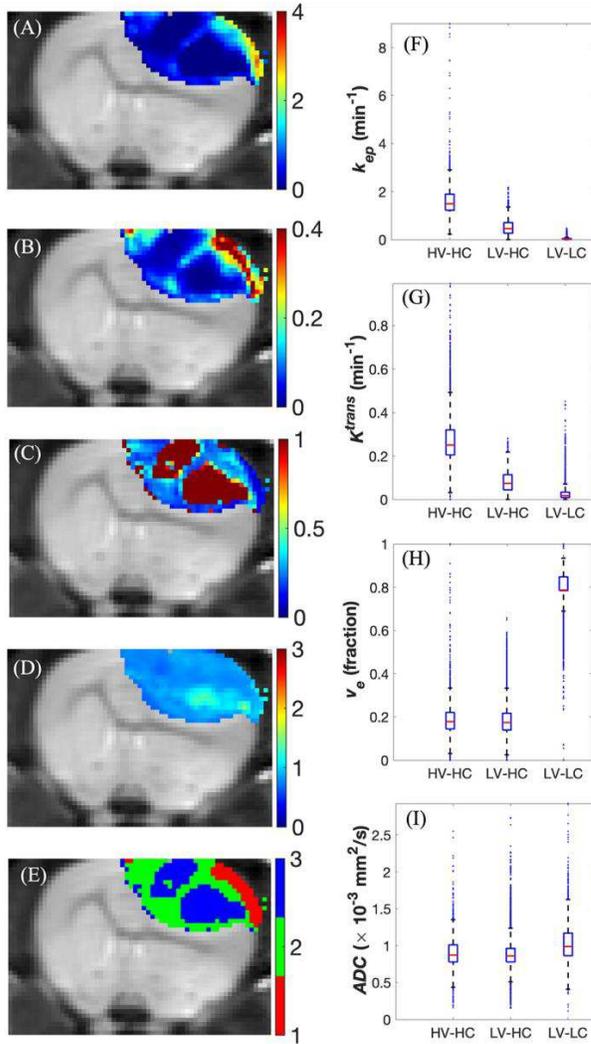


Figure 4: Quantitative parameter maps and tumor habitat map for a representative rat at the third imaging visit. (A)-(D) depict the k_{ep} (min⁻¹), v_e , K^{trans} (min⁻¹), and ADC (10^{-3} s/mm²) maps, respectively. (E) This is the tumor habitat map where HV-HC is in red, LV-HC is in green, and LV-LC is in blue. The LV-LC cluster spatially localizes to the center of the tumor with high values of v_e and high values of the ADC, which align with expected observations in necrotic regions. The HV-HC habitat corresponds to the edge of the tumor that exhibits characteristic high values of K^{trans} and k_{ep} most likely due to breakdown of the blood-brain-barrier. Lastly, the LV-

HC region surrounds the necrotic regions, consisting of voxels with mid-range values of K^{trans} and lower values of v_e compared to the LV-LC habitat. (F)-(I) are the quantitative parameter distributions from each habitat, grouped by k_{ep} , K^{trans} , v_e , and ADC, respectively. All four parameter distributions differ significantly between the three habitats ($p < 0.05$).

3.2 Calibration of the family of mathematical models

The tumor habitat time series from all 16 rat datasets were used to calibrate all four mathematical models (Equations (1)-(4), Figure 2). The calibrated model parameters for Model A (the parent model) are presented in Table S2, and we found that the d_{21} , d_{13} , and θ parameter distributions are significantly different between the control (C) and treated (R) cohorts. Additionally, in Table S2, we observed that the model parameters d_{21} and d_{13} are one to two orders of magnitude smaller compared to the other model parameters, prompting the development of the additional three models that exclude these model parameters.

Panels A-D of Figure 5 show all four models calibrated to the habitat times series of four representative rats from the control group, while panels E-H show these calibrations for four representative rats from the treated group (See Figures S5 and S6 for all 8 control and all 8 treated rat calibrations, respectively). There were no statistical differences in the SSEs for all models when compared within identical habitats. Moreover, the four CCC values (Figure S7), computed from the combined results of each calibration for each model, are all above 0.95, indicating strong agreement between the four models and the data that they were calibrated to.

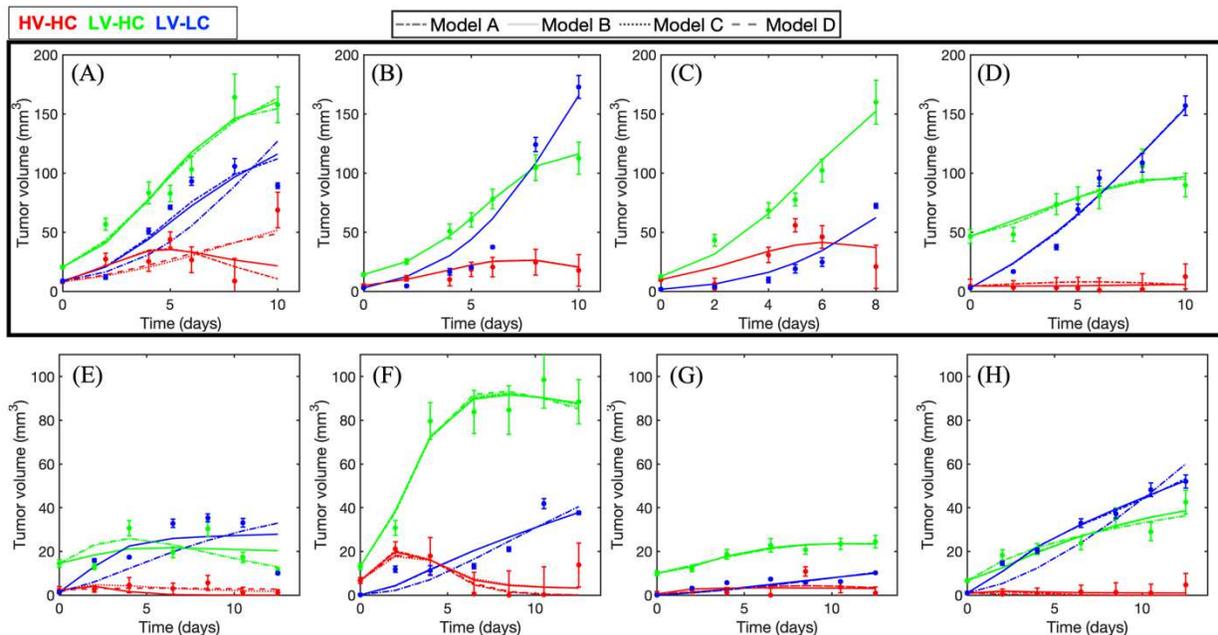


Figure 5. A family of four models calibrated to the habitat data. The first row (black rectangle) shows the results of the calibrations to four example control rats, while the second row shows analogous results for four example rats from the treated cohort (where (E)-(G) received 40 Gy of radiation, and (H) received 20 Gy). Data are represented by solid circles with confidence intervals computed as described in the Methods. Red, green, and blue correspond to the HV-HC, LV-HC, and LV-LC habitats, respectively. Each model is represented by a different line style as depicted in the legend at the top of the figure. There is high similarity between the fits of all four models across all rats, suggesting that each model shows similar performance in describing the dynamics of each habitat time series and allowing for the selection of the most parsimonious model with the least number of free parameters. When comparing tumor sizes at the final imaging visit, the treated rats exhibited significantly smaller tumors than the control rats ($p < 0.05$). This observation is apparent as panels (A)-(D) have double the dynamic range of panels (E).

3.3 Selecting the most parsimonious model

The AIC values were computed using a MATLAB implementation of Equation (5) for each of the four calibrated models across all eight habitat time series, resulting in four AIC values per rat that are collected in Table S3. Model D consistently yielded the lowest AIC for all rat datasets, which is why Model D was selected as the most parsimonious model relative to the other three models. Thus, Model D is used for the remaining analyses in this study.

3.4 Prediction and validation of tumor habitat dynamics

Using each tumor habitat time series as a validation dataset, we applied the LOO and bootstrapping methods to evaluate the ability of Model D to predict the habitat dynamics. To select the appropriate number of samples, N_b , in bootstrapping, we performed the bootstrapping method for different values of N_b , plotted the resultant SSEs against N_b , and averaged the results across all rats in the control (Figure S8A) and treated (Figure S8B) cohorts. We qualitatively inspected the plots in Figure S8 to arrive at $N_b = 200$ as an appropriate sample size at which the SSE values have stabilized and additional samples do not improve the prediction.

Figure 6 shows the results of the predictions for a representative rat. The calibrated model in Figure 6A is juxtaposed with the LOO predictions (Figure 6B-6D) and the bootstrapping predictions (Figure 6E-6H). Using the initial tumor volume (Figure 6B) and tumor composition (Figure 6C) as weighting schemes in the LOO approach resulted in overestimation of the LV-LC and LV-HC habitat volumes at the final two imaging visits beyond the dynamic range of the data. Bootstrapping with data updating (Figure 6F-6H) decreased the error in the predictions at these later imaging visits. While the LOO predictions differ in accuracy depending on the weighting scheme, the bootstrapping predictions with data updating yielded lower SSE values overall (Figure 7). In the control cohort, the SSEs from the bootstrapping predictions were significantly lower ($p < 0.05$) than all SSEs from the LOO predictions for the HV-HC (Figure 7A) and LV-HC (Figure 7B) tumor habitats; this significance was also observed in the LV-LC LOO prediction using the weighting scheme determined by the initial tumor volume (Figure 7C). Figure 7D-7F shows the SSE distributions of the predictions for each habitat in the eight treated rats. The SSE values for bootstrapping with two visits in the data updating step were significantly lower ($p < 0.05$) than the remaining predictions for the HV-HC habitat alone. In all cases (Figure 7A-7F), bootstrapping with two visits yielded statistically similar SSE values compared to the individually calibrated fit.

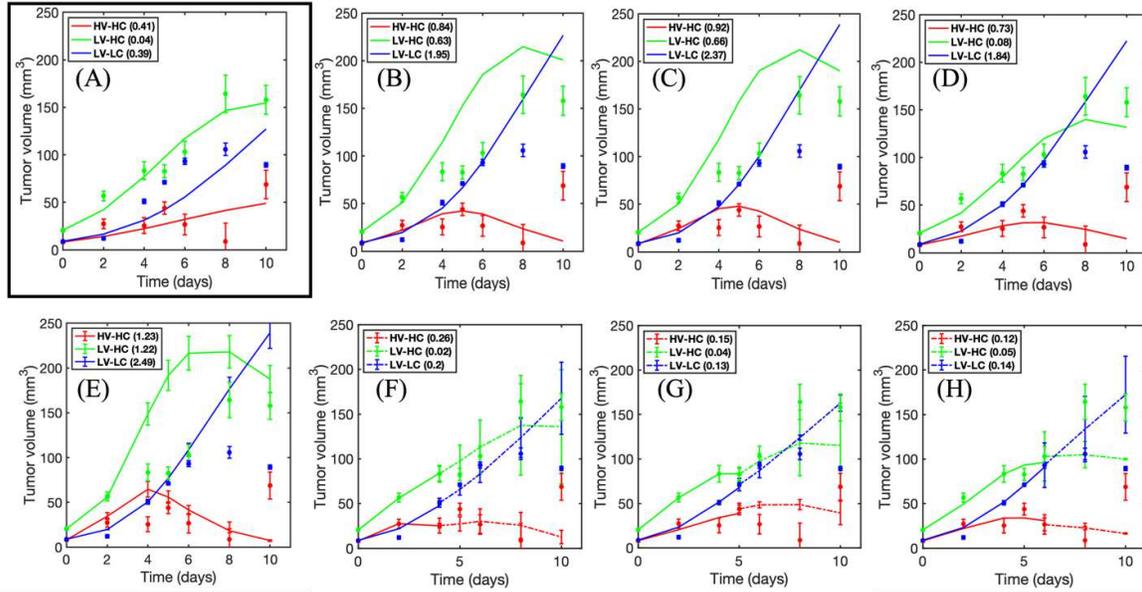


Figure 6. Predictions for a representative rat using the leave-one-out and bootstrapping approaches. (A) Calibrated Model D fit to the habitat time series of a representative rat. The SSEs for each habitat are displayed in the legend. The black box emphasizes that this is the calibrated fit to which the predictions are to be compared. (B)-(C) are the predicted curves using the leave-one-out approach that weighs the prediction from the remaining population of rats by similarity in initial tumor volume, initial tumor composition, and initial HV-HC volume, respectively. (E) The bootstrapped prediction with the initial visit serving as an initial condition is plotted along with corresponding volume measurements. (F)-(H) present bootstrapping with two, three, and four visits, respectively. The solid curves in (F)-(H) are the model calibration using the available data, whereas the dashed curves are the predictions forward. Bootstrapping, when incorporating additional imaging data to update the prediction, yields overall smaller SSEs than the leave-one-out approach.

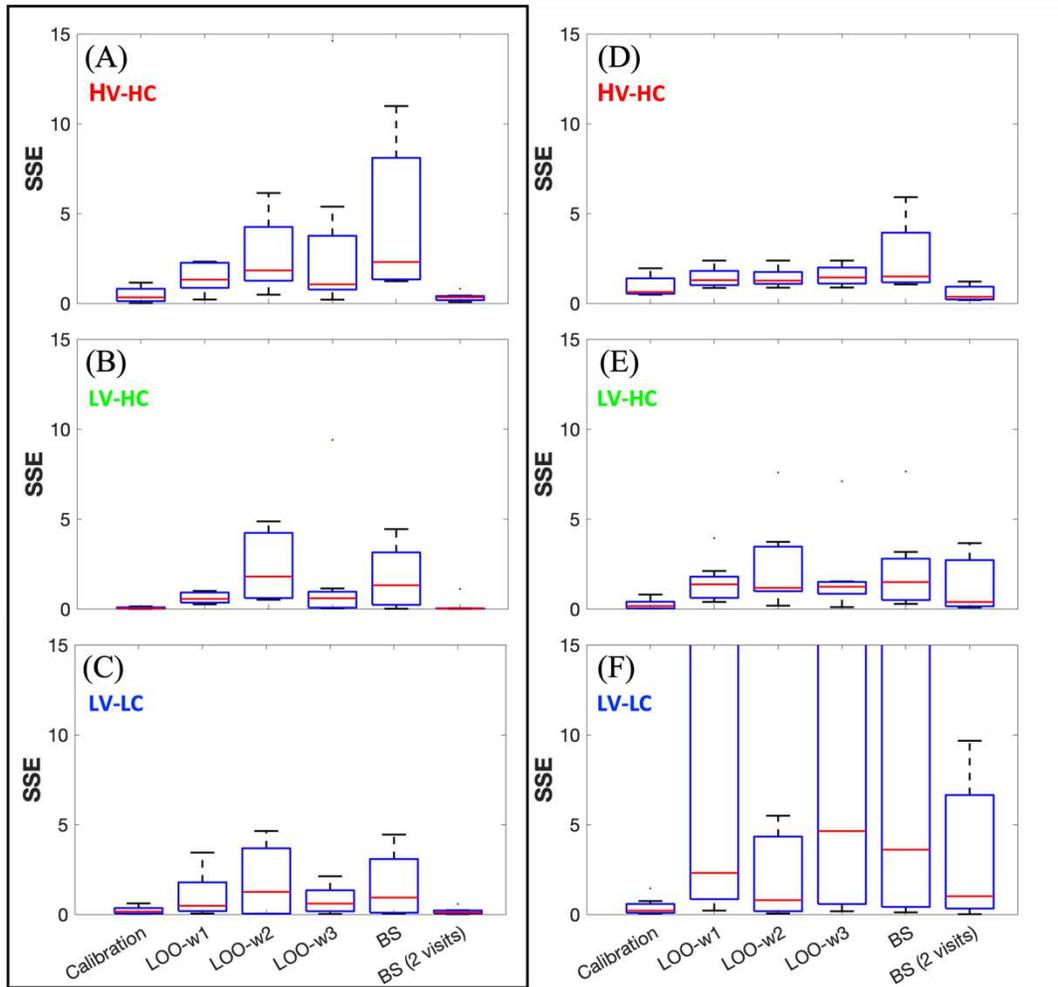


Figure 7. Comparing errors between predictions cross all rat datasets. (A)-(C) display the SSE distributions for six different prediction approaches for the HV-HC, LV-HC, and LV-LC habitats, respectively, for the control rat cohort (black rectangle). From left to right within each panel, the boxplots are model calibration, leave-one-out (LOO) with the initial tumor volume as the weighting scheme (LOO-w1), LOO with the initial tumor habitat composition as the weighting scheme (LOO-w2), LOO with the initial HV-HC volume as the weighting scheme (LOO-w3), bootstrapping (BS) with no data updating, and bootstrapping with two visits (BS (2 visits)). Bootstrapping with two visits yielded significantly lower error than the other four methods ($p < 0.01$) in the HV-HC and LV-HC habitat growth predictions. Furthermore, bootstrapping with two visits showed no significant differences in error between the direct calibrations. (D)-(F) are the analogous results for the treated rat cohort. Similarly, bootstrapping with two visits yielded significantly lower SSEs compared to the remaining prediction methods in the HV-HC habitat growth predictions. SSEs in (F) were significantly higher than the SSEs in the remaining clusters, where the SSE distribution from the LOO-w2 had the lowest median and lowest interquartile range and outlier value. To provide an understanding of the anomalous ranges in panel (F), we report that the 75th percentile of LOO-w3 and BS did not exceed 30, whereas this value for LOO w1 was 53.06.

4 DISCUSSION

To the best of our knowledge, we have presented the first application of an ODE-based model to describe the dynamics of image-informed tumor habitats. Because of the absence of spatial location in the ODE model, our proposed pipeline does not require accurately registered imaging data for habitat identification and subsequent mathematical modelling. This makes our pipeline amendable to fast analysis when image registration tools are not readily available or when the underlying anatomy makes registration challenging. Our method provides a global description of heterogeneous tumor growth – in the presence and absence of radiation therapy – and provides an avenue for predicting changes in heterogeneity and tumor progression.

The three identified habitats were determined based on variations in vascularity and cellularity and labelled accordingly as HV-HC, LV-HC, and LV-LC. Importantly, no spatial information was used during clustering, yet the habitats were found to spatially localize based on their similar properties, which was quantitatively confirmed with MSI analysis (Figure S4). Moreover, the habitats localized in certain regions of the tumor. For example, across all rats, the HV-HC cluster was found at the outer rim of the tumor, while the LV-LC tumor was predominantly located at the center of the tumor (Figure 4); and the LV-HC was found between the LV-LC and HV-HC habitats.

We employed model selection to identify the most parsimonious model for describing the tumor habitat dynamics. Model selection revealed that there are infrequent transitions from the LV-HC habitat to the HV-HC habitat as well as the HV-HC habitat directly to the LV-LC habitat (Table S2). Coupled with the observed localizations of each habitat at certain regions of the tumors, there is strong evidence that the LV-LC habitat correlates with necrosis typically found at the center of growing tumors. The LV-LC habitat exhibited significantly higher *ADC* values compared to the remaining two habitats, and high *ADC* values have been correlated with growth-induced and treatment-induced necrosis.^{51,52} The HV-HC cluster correlates with the highly vascularized rim in which perfusion parameters have been associated with biological markers of vascularization.^{53,54} Lastly, from previous work, the LV-HC region surrounding the necrotic tumor core has been associated with markers of hypoxia from reduced vasculature^{7,55}.

The parameters defining the most parsimonious model as well as the above observations of the spatial localizations of the habitats suggest that, as the tumor grows, regions of HV-HC begin to lose access to vasculature and become regions of low vascularity while still maintaining

cellularity (transitioning to the LV-HC habitat). As time elapses, we hypothesize that the cells within the LV-HC cluster begin to die from lack of oxygen formerly provided by the vasculature, leading to a decrease in cellularity and subsequent necrosis, which completes the transition to the LV-LC habitat. Prior work with more complex formulations of tumor region interactions have demonstrated the effect of hypoxic and necrotic regions in radiation therapy, where necrotic regions in particular have been shown to negatively affect treatment response in *in silico* experiments^{55,56}.

Compared to other heterogeneity analyses and modelling methods, habitat imaging retains spatial information that is crucial for observations across time, like treatment response. Some therapies are known to affect functions of a tumor, such as vascularization, which would manifest as a relative shrinkage of the HV-HC habitat in our proposed analysis. Multi-species PDE model formulations from the literature have successfully linked the dynamics of hypoxic, necrotic, and perfused tumor regions *in silico* and in patient populations^{29,30}, which demonstrates the potential to extend the complexity of our ODE formulation. The challenge, however, is in defining a complete mathematical description of the habitat interactions.

The most parsimonious ODE mathematical model in our formulation was successful in predicting global habitat dynamics, with the most successful method being bootstrapping with data updating. Bootstrapping with data updating yielded significantly lower error than the other prediction methods in the control cohort; however, the LV-LC habitat predictions in the treated cohort exhibited significantly higher error overall, even when using the parent model *in lieu* of the most parsimonious model. While this suggests that there are dynamics due to treatment that are not accurately captured, there is still utility in using the proposed ODE model for prediction. Radiation therapy confers marked decreases in vasculature^{57,58}, manifesting as a decrease in the HV-HC habitat and an increase in the LV-HC habitat over time (Figure S3), which is accurately captured by the model proposed in this study. When no imaging data is available beyond the initial visit, the LOO approach weighted by initial tumor volume or HV-HC volume performs better than bootstrapping alone as it directly considers measured biological information of other tumors in the population. Thus, there is flexibility in choosing a prediction approach based on data availability.

While the proposed pipeline for modelling tumor habitats confers some advances over existing modelling efforts for high grade glioma, there are opportunities for further investigation. First, while the identified habitats agree with previous work^{7,8} in which the habitats were

histologically verified, there is a need to repeat histological analysis in this study of murine glioma. This added analysis would validate the image-based tumor habitats, making it feasible for future work to perform habitat imaging in independent cohorts of animals with greater certainty. Additionally, while the clustering analysis was done on the entire set of available rat data (control and treated), the mathematical modelling analysis was performed on a subset of the treated cohort along with the control cohort as explained in Section 2.4. An important next step is to repeat this analysis using a larger cohort of rats in each treatment group (20 or 40 Gy) to determine whether a modified ODE model is necessary to capture the effects of different therapy dosage regimens on the habitat dynamics.

5 CONCLUSION

This study presented a pipeline for mathematically modelling image-derived tumor habitats in a murine model of glioma using a set of biology-based, coupled ordinary differential equations. We identified three distinct tumor habitats from clustering four sets of quantitative parameters computed from DCE- and DW-MRI data. All models in the family of ODE models exhibited errors that were not significantly different, allowing for the selection of the most parsimonious model that was parameterized by the least number of parameters. Model selection, coupled with observed habitat localizations, revealed important biological insights on glioma behavior that are corroborated by the literature. We found that the selected model could predict tumor dynamics. Further development and validation of this approach could yield accurate predictions of the dynamics of tumor physiology that may be leveraged to optimize therapy for individual tumors.

ACKNOWLEDGEMENTS

We thank the National Institutes of Health for funding through NCI U01CA142565, U01CA174706, U01CA253540, U24CA226110, and R01CA240589, as well as NIH U24EB029240. Z.C. and S.P. were supported by the National Science Foundation 1757885 as Undergraduate Research Scholars, and A.G. was supported by NIH 5R01CA226258. We also extend our gratitude to the American Cancer Society for support through RSG-18-006-01-CCE and the Cancer Prevention and Research Institute of Texas for support through RP220225 and RR160005. T.E.Y. is a CPRIT Scholar in Cancer Research.

DATA AVAILABILITY

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES:

1. Gladson, C. L., Prayson, R. A. & Liu, W. M. The pathobiology of glioma tumors. *Annual Review of Pathology: Mechanisms of Disease* vol. 5 33–50 Preprint at <https://doi.org/10.1146/annurev-pathol-121808-102109> (2010).
2. Ostrom, Q. T., Cote, D. J., Ascha, M., Kruchko, C. & Barnholtz-Sloan, J. S. Adult glioma incidence and survival by race or ethnicity in the United States from 2000 to 2014. *JAMA Oncology* vol. 4 1254–1262 Preprint at <https://doi.org/10.1001/jamaoncol.2018.1789> (2018).
3. Marusyk, A. & Polyak, K. Tumor heterogeneity: Causes and consequences. *Biochimica et Biophysica Acta - Reviews on Cancer* vol. 1805 105–117 Preprint at <https://doi.org/10.1016/j.bbcan.2009.11.002> (2010).
4. Becker, A. P., Sells, B. E., Jaharul Haque, S. & Chakravarti, A. Tumor heterogeneity in glioblastomas: From light microscopy to molecular pathology. *Cancers* vol. 13 1–25 Preprint at <https://doi.org/10.3390/cancers13040761> (2021).
5. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* vol. 12 31–46 Preprint at <https://doi.org/10.1158/2159-8290.CD-21-1059> (2022).
6. Gatenby, R. A., Grove, O. & Gillies, R. J. Quantitative imaging in cancer evolution and ecology. *Radiology* vol. 269 8–15 Preprint at <https://doi.org/10.1148/radiol.13122697> (2013).
7. Syed, A. K., Whisenant, J. G., Barnes, S. L., Sorace, A. G. & Yankeelov, T. E. Multiparametric analysis of longitudinal quantitative mri data to identify distinct tumor habitats in preclinical models of breast cancer. *Cancers (Basel)* **12**, 1–20 (2020).
8. Kazerouni, A. S. *et al.* Quantifying Tumor Heterogeneity via MRI Habitats to Characterize Microenvironmental Alterations in HER2+ Breast Cancer. *Cancers (Basel)* **14**, 1837 (2022).
9. Parker, N. R., Khong, P., Parkinson, J. F., Howell, V. M. & Wheeler, H. R. Molecular heterogeneity in glioblastoma: Potential clinical implications. *Front Oncol* **5**, (2015).
10. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine* **366**, 883–892 (2012).
11. Yankeelov, T. E., Abramson, R. G. & Quarles, C. C. Quantitative multimodality imaging in cancer research and therapy. *Nat Rev Clin Oncol* **11**, 670–680 (2014).
12. Arlinghaus, L. R. & Yankeelov, T. E. Diffusion-Weighted MRI. in *Quantitative MRI in Cancer; Imaging in Medical Diagnosis and Therapy* (eds. Yankeelov, T. E., Pickens, D. R. & Price, R. R.) 91–107 (CRC Press, 2011).
13. Yankeelov, T. & Gore, J. Dynamic Contrast Enhanced Magnetic Resonance Imaging in Oncology: Theory, Data Acquisition, Analysis, and Examples. *Curr Med Imaging Rev* **3**, 91–107 (2007).

14. Aydin, S., Fatihoğlu, E., Koşar, P. N. & Ergün, E. Perfusion and permeability MRI in glioma grading. *Egyptian Journal of Radiology and Nuclear Medicine* **51**, (2020).
15. Henriksen, O. M. *et al.* High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 1: Perfusion and Diffusion Techniques. *Front Oncol* **12**, (2022).
16. Aquino, D., Gioppo, A., Finocchiaro, G., Bruzzone, M. G. & Cuccarini, V. MRI in Glioma Immunotherapy: Evidence, Pitfalls, and Perspectives. *Journal of Immunology Research* vol. 2017 Preprint at <https://doi.org/10.1155/2017/5813951> (2017).
17. Moffat, B. A. *et al.* Functional diffusion map: A noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Targeted Growth, Inc* www.pnas.org/cgi/doi/10.1073/pnas.0501532102 (2005).
18. Tofts, P. S. & Kermode, A. G. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magn Reson Med* **17**, 357–367 (1991).
19. Artzi, M. *et al.* Repeatability of dynamic contrast enhanced v p parameter in healthy subjects and patients with brain tumors. *J Neurooncol* **140**, 727–737 (2018).
20. O'Connor, J. P. B. *et al.* Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome. *Clinical Cancer Research* vol. 21 249–257 Preprint at <https://doi.org/10.1158/1078-0432.CCR-14-0990> (2015).
21. Tofts, P. S. & Parker, G. J. M. DCE-MRI: Acquisition and analysis techniques. *Clinical Perfusion MRI: Techniques and Applications* **9781107013**, 58–74 (2010).
22. Hu, L. S., Hawkins-Daarud, A., Wang, L., Li, J. & Swanson, K. R. Imaging of intratumoral heterogeneity in high-grade glioma. *Cancer Letters* vol. 477 97–106 Preprint at <https://doi.org/10.1016/j.canlet.2020.02.025> (2020).
23. Hu, L. S. *et al.* Multi-parametric MRI and texture analysis to visualize spatial histologic heterogeneity and tumor extent in glioblastoma. *PLoS One* **10**, (2015).
24. O'Connor, J. P. B. Cancer heterogeneity and imaging. *Seminars in Cell and Developmental Biology* vol. 64 48–57 Preprint at <https://doi.org/10.1016/j.semdb.2016.10.001> (2017).
25. Brüningk, S. C. *et al.* Intermittent radiotherapy as alternative treatment for recurrent high grade glioma: a modeling study based on longitudinal tumor measurements. *Sci Rep* **11**, 20219 (2021).
26. Hormuth II, D. A., al Feghali, K. A., Elliott, A. M., Yankeelov, T. E. & Chung, C. Image-based personalization of computational models for predicting response of high-grade glioma to chemoradiation. *Sci Rep* **11**, (2021).
27. Hormuth, D. A. *et al.* A mechanically coupled reaction-diffusion model that incorporates intra-tumoural heterogeneity to predict in vivo glioma growth. *J R Soc Interface* **14**, (2017).
28. Hormuth II, D. A. H. *et al.* Mechanism-Based Modeling of Tumor Growth and Treatment Response Constrained by Multiparametric Imaging Data. (2019) doi:10.1200/CCI.18.
29. Subramanian, S., Gholami, A. & Biros, G. Simulation of glioblastoma growth using a 3D multispecies tumor model with mass effect. *J Math Biol* **79**, 941–967 (2019).
30. Swanson, K. R. *et al.* Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology. *Cancer Res* **71**, 7366–75 (2011).

31. Glazar, D. J. *et al.* Tumor volume dynamics as an early biomarker for patient-specific evolution of resistance and progression in recurrent high-grade glioma. *J Clin Med* **9**, 1–11 (2020).
32. Koziol, J. A., Falls, T. J. & Schnitzer, J. E. Different ODE models of tumor growth can deliver similar results. *BMC Cancer* **20**, (2020).
33. Barth, R. F. & Kaur, B. Rat brain tumor models in experimental neuro-oncology: the C6, 9L, T9, RG2, F98, BT4C, RT-2 and CNS-1 gliomas. *J Neurooncol* **94**, 299–312 (2009).
34. Colvin, D. C. *et al.* Earlier detection of tumor treatment response using magnetic resonance diffusion imaging with oscillating gradients. *Magn Reson Imaging* **29**, 315–23 (2011).
35. Hormuth II, D. A. *et al.* Predicting in vivo glioma growth with the reaction diffusion equation constrained by quantitative magnetic resonance imaging data. *Phys Biol* **12**, 46006 (2015).
36. Tofts, P. S. & Kermode, A. G. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* **17**, 357–367 (1991).
37. Hormuth II, D. A., Skinner, J. T., Does, M. D. & Yankeelov, T. E. A comparison of individual and population-derived vascular input functions for quantitative DCE-MRI in rats. *Magn Reson Imaging* **32**, 397–401 (2014).
38. Donahue, K. M. *et al.* Dynamic Gd-DTPA enhanced MRI measurement of tissue cell volume fraction. *Magn Reson Med* **34**, 423–32 (1995).
39. Li, X. *et al.* Dynamic-contrast-enhanced-MRI with extravasating contrast reagent: rat cerebral glioma blood volume determination. *Journal of Magnetic Resonance* **206**, 190–9 (2010).
40. Barnes, S. L., Whisenant, J. G., Loveless, M. E., Ayers, G. D. & Yankeelov, T. E. Assessing the reproducibility of dynamic contrast enhanced magnetic resonance imaging in a murine model of breast cancer. *Magn Reson Med* **69**, 1721–1734 (2013).
41. Galbraith, S. M. *et al.* Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: comparison of quantitative and semi-quantitative analysis. *NMR Biomed* **15**, 132–142 (2002).
42. Yankeelov, T. E. *et al.* Quantitative pharmacokinetic analysis of DCE-MRI data without an arterial input function: A reference region model. *Magn Reson Imaging* **23**, 519–529 (2005).
43. Barnes, S. L., Whisenant, J. G., Loveless, M. E. & Yankeelov, T. E. Practical Dynamic Contrast Enhanced MRI in Small Animal Models of Cancer: Data Acquisition, Data Analysis, and Interpretation. *Pharmaceutics* **4**, 442–478 (2012).
44. Han, J., Kamber, M. & Pei, J. Data preprocessing. in *Data Mining: Concepts and Techniques* (Elsevier Science & Technology, 2011).
45. Macqueen, J. *SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS.*
46. Nielsen, F. Hierarchical clustering. in *Introduction to HPC with MPI for Data Science; Undergraduate Topics in Computer Science* 195–211 (Springer International Publishing, 2016).

47. Wu, J. *et al.* Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. *Radiology* **288**, 26–35 (2018).
48. Akaike, H. Canonical Correlation Analysis of Time Series and the Use of an Information Criterion. *Mathematics in Science and Engineering* 27–96 (1976) doi:10.1016/S0076-5392(08)60869-3.
49. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap*. (CRC Press, 1993).
50. Lin, L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **45**, 255 (1989).
51. Zhang, M. *et al.* Large-volume low apparent diffusion coefficient lesions predict poor survival in bevacizumab-treated glioblastoma patients. *Neuro Oncol* **18**, 735–743 (2016).
52. Chenevert, T. L., Sundgren, P. C. & Ross, B. D. Diffusion imaging: insight to cell status and cytoarchitecture. *Neuroimaging Clin N Am* **16**, 619–32, viii–ix (2006).
53. Keil, V. C. *et al.* DCE-MRI in Glioma, Infiltration Zone and Healthy Brain to Assess Angiogenesis: A Biopsy Study. *Clin Neuroradiol* **31**, 1049–1058 (2021).
54. Ferrier, M. C. *et al.* Validation of Dynamic Contrast-Enhanced Magnetic Resonance Imaging-Derived Vascular Permeability Measurements Using Quantitative Autoradiography in the RG2 Rat Brain Tumor Model. *Neoplasia* **9**, 546–555 (2007).
55. Lewin, T. D., Maini, P. K., Moros, E. G., Enderling, H. & Byrne, H. M. The Evolution of Tumour Composition During Fractionated Radiotherapy: Implications for Outcome. *Bull Math Biol* **80**, 1207–1235 (2018).
56. Lewin, T. D. *et al.* The importance of dead material within a tumour on the dynamics in response to radiotherapy. *Phys Med Biol* **65**, 015007 (2020).
57. Gupta, K. & Burns, T. C. Radiation-Induced Alterations in the Recurrent Glioblastoma Microenvironment: Therapeutic Implications. *Front Oncol* **8**, (2018).
58. van Vulpen, M., Kal, H., Taphoorn, M. & el Sharouni, S. Changes in blood-brain barrier permeability induced by radiotherapy: Implications for timing of chemotherapy? (Review). *Oncol Rep* (2002) doi:10.3892/or.9.4.683.

AUTHOR CONTRIBUTIONS

KPS performed all analyses, created all figures and tables, and drafted the manuscript. SHP and ZC contributed equally to this work by laying the foundations of the clustering algorithm and preliminary model design and implemented these foundations in MATLAB during their summer as Undergraduate Researchers at The University of Texas at Austin. KPS began with these foundations to advance and refine the analysis pipeline. ASK provided important feedback and expertise on habitat identification, ideas for the figures, and extensively reviewed the manuscript before submission. TEY and DAH advised KPS on the analysis methods – particularly on the mathematical model calibration and prediction methods – and extensively reviewed the manuscript before submission; and, moreover, DAH collected all rat imaging data.

DAH along with AG mentored ZC and SHP during the preliminary stages of this work. All authors reviewed the manuscript before submission.

COMPETING INTERESTS STATEMENTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Habitat identification and modelling pipeline. (A) DCE-MRI and DW-MRI data were acquired for each animal over multiple days. The multiparametric MRI data were analyzed to generate four parameter maps (i.e., k_{ep} , v_e , K^{trans} , ADC) across the tumor ROI (top left). The quantitative information at each voxel at all imaging visits for all animals was pooled together and then clustered to identify three tumor habitats with unique physiologies: high-vascularity high-cellularity, low-vascularity high-cellularity, and low-vascularity low-cellularity (bottom left). (B) Once the habitats were identified, each tumor in each rat yielded three-habitat time series that could be analyzed by a family of mathematical models based on three compartments and the transition rates between them (top right). We then tested the ability of the most parsimonious model to predict habitat dynamics through leave-one-out (Method 1) and bootstrapping (Method 2) analyses (bottom right).

Figure 2. Visualizing a three-compartment mathematical model of tumor habitat growth. A physiological label can be attributed to each cluster in the parent model: $C_1(t)$ corresponds to the HV-HC habitat (red), $C_2(t)$ corresponds to the LV-HC habitat (green), and $C_3(t)$ corresponds to the LV-LC habitat (blue). The parameters k_1 and k_2 are the growth rates of $C_1(t)$ and $C_2(t)$, respectively, while all parameters of the form d_{nm} represent the transition rate from cluster n to cluster m . All model parameters are named and described in Table 1.

Figure 3: Identification of three tumor habitats with k -means clustering. (A) This is the heat map of normalized quantitative parameters grouped by the three identified clusters: high-vascularity high-cellularity (HV-HC), low-vascularity high-cellularity (LV-HC), and low-vascularity low-cellularity (LV-LC). (B) Boxplots of the distributions of k_{ep} are presented in each

identified habit, where the red line in each boxplot represents the median of the distribution. Analogous to (B), panels (C)-(E) show the distributions of v_e , K^{trans} , and ADC, respectively. 168,207 voxels were included in this clustering analysis. We found that all four parameter distributions differ significantly between the three habitats, with $p < 0.05$.

Figure 4: Quantitative parameter maps and tumor habitat map for a representative rat at the third imaging visit. (A)-(D) depict the k_{ep} (min^{-1}), v_e , K^{trans} (min^{-1}), and ADC (10^{-3} s/mm^2) maps, respectively. (E) This is the tumor habitat map where HV-HC is in red, LV-HC is in green, and LV-LC is in blue. The LV-LC cluster spatially localizes to the center of the tumor with high values of v_e and high values of the ADC, which align with expected observations in necrotic regions. The HV-HC habitat corresponds to the edge of the tumor that exhibits characteristic high values of K^{trans} and k_{ep} most likely due to breakdown of the blood-brain-barrier. Lastly, the LV-HC region surrounds the necrotic regions, consisting of voxels with mid-range values of K^{trans} and lower values of v_e compared to the LV-LC habitat. (F)-(I) are the quantitative parameter distributions from each habitat, grouped by k_{ep} , K^{trans} , v_e , and ADC, respectively. All four parameter distributions differ significantly between the three habitats ($p < 0.05$).

Figure 5. A family of four models calibrated to the habitat data. The first row (black rectangle) shows the results of the calibrations to four example control rats, while the second row shows analogous results for four example rats from the treated cohort (where (E)-(G) received 40 Gy of radiation, and (H) received 20 Gy). Data are represented by solid circles with confidence intervals computed as described in the Methods. Red, green, and blue correspond to the HV-HC, LV-HC, and LV-LC habitats, respectively. Each model is represented by a different line style as depicted in the legend at the top of the figure. There is high similarity between the fits of all four models across all rats, suggesting that each model shows similar performance in describing the dynamics of each habitat time series and allowing for the selection of the most parsimonious model with the least number of free parameters. When comparing tumor sizes at the final imaging visit, the treated rats exhibited significantly smaller tumors than the control rats ($p < 0.05$). This observation is apparent as panels (A)-(D) have double the dynamic range of panels (E).

Figure 6. Predictions for a representative rat using the leave-one-out and bootstrapping approaches. (A) Calibrated Model D fit to the habitat time series of a representative rat. The SSEs for each habitat are displayed in the legend. The black box emphasizes that this is the calibrated fit to which the predictions are to be compared. (B)-(C) are the predicted curves using the leave-one-out approach that weighs the prediction from the remaining population of rats by similarity in initial tumor volume, initial tumor composition, and initial HV-HC volume, respectively. (E) The bootstrapped prediction with the initial visit serving as an initial condition is plotted along with corresponding volume measurements. (F)-(H) present bootstrapping with two, three, and four visits, respectively. The solid curves in (F)-(H) are the model calibration using the available data, whereas the dashed curves are the predictions forward. Bootstrapping, when incorporating additional imaging data to update the prediction, yields overall smaller SSEs than the leave-one-out approach.

Figure 7. Comparing errors between predictions cross all rat datasets. (A)-(C) display the SSE distributions for six different prediction approaches for the HV-HC, LV-HC, and LV-LC habitats, respectively, for the control rat cohort (black rectangle). From left to right within each panel, the boxplots are model calibration, leave-one-out (LOO) with the initial tumor volume as the weighting scheme (LOO-w1), LOO with the initial tumor habitat composition as the weighting scheme (LOO-w2), LOO with the initial HV-HC volume as the weighting scheme (LOO-w3), bootstrapping (BS) with no data updating, and bootstrapping with two visits (BS (2 visits)). Bootstrapping with two visits yielded significantly lower error than the other four methods ($p < 0.01$) in the HV-HC and LV-HC habitat growth predictions. Furthermore, bootstrapping with two visits showed no significant differences in error between the direct calibrations. (D)-(F) are the analogous results for the treated rat cohort. Similarly, bootstrapping with two visits yielded significantly lower SSEs compared to the remaining prediction methods in the HV-HC habitat growth predictions. SSEs in (F) were significantly higher than the SSEs in the remaining clusters, where the SSE distribution from the LOO-w2 had the lowest median and lowest interquartile range and outlier value. To provide an understanding of the anomalous ranges in panel (F), we report that the 75th percentile of LOO-w3 and BS did not exceed 30, whereas this value for LOO w1 was 53.06.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SlavkovaAim3Supplementsubmission.docx](#)