Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 17(4): 3660–3709.
DOI: 10.3934/mbe.2020207
Received: 03 February 2020
Accepted: 08 May 2020
Published: 19 May 2020

Review

A tutorial review of mathematical techniques for quantifying tumor heterogeneity

Rebecca Everett[1], Kevin B. Flores[2,*], Nick Henscheid[3], John Lagergren[2], Kamila Larripa[4], Ding Li[5], John T. Nardini[2,6], Phuong T. T. Nguyen[7], E. Bruce Pitman[8] and Erica M. Rutter[9]

[1] Department of Mathematics and Statistics, Haverford College, Haverford, PA, USA
[2] Department of Mathematics, North Carolina State University, Raleigh, NC, USA
[3] Department of Medical Imaging, University of Arizona, Tucson, AZ, USA
[4] Department of Mathematics, Humboldt State University, Arcata, CA, USA
[5] Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA
[6] Statistical and Applied Mathematical Sciences Institute, Durham, NC, USA
[7] Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA
[8] Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY, USA
[9] Department of Applied Mathematics, University of California, Merced, Merced, CA, USA

* **Correspondence:** Email: kbflores@ncsu.edu; Tel: +19195130821; Fax: +19195137336.

Abstract: Intra-tumor and inter-patient heterogeneity are two challenges in developing mathematical models for precision medicine diagnostics. Here we review several techniques that can be used to aid the mathematical modeller in inferring and quantifying both sources of heterogeneity from patient data. These techniques include virtual populations, nonlinear mixed effects modeling, non-parametric estimation, Bayesian techniques, and machine learning. We create simulated virtual populations in this study and then apply the four remaining methods to these datasets to highlight the strengths and weaknesses of each technique. We provide all code used in this review at https://github.com/jtnardin/Tumor-Heterogeneity/ so that this study may serve as a tutorial for the mathematical modelling community. This review article was a product of a Tumor Heterogeneity Working Group as part of the 2018–2019 Program on Statistical, Mathematical, and Computational Methods for Precision Medicine which took place at the Statistical and Applied Mathematical Sciences Institute.

Keywords: cancer heterogeneity; mathematical oncology; tumor growth; glioblastoma multiforme; virtual populations; nonlinear mixed effects; spatiotemporal data; Bayesian estimation; generative adversarial networks; non-parametric estimation; variational autoencoders; machine learning

Contents

Abbreviations: BX/STR: biopsy/subtotal resection; CAT: computed axial tomography; CE: Comparative E ectiveness; CT: computed tomography; DC: dendritic cell; ECM: extracellular matrix; EHR: electronic health records; ^{18}F-FLT: Fluorothymidine F-18; FET-PET: fluoroethyl-l-tyrosine positron emission tomography; GBM: glioblastoma multiforme; GTR: gross total resection; HE: Hematoxylin and eosin stain; LGG: low-grade glioma; MRI: Magnetic Resonance Imaging; MTD: mean tumor diameter; OV: oncolytic virus; PCV: procarbazine and vincistrine; PET: Positron Emission Tomography; PSA: prostate-specific antigen; RCT: Randomized Controlled Trial; TMZ: temozolomide; AE: Auto-encoder; AIC: Akaike Information Criteria; ANOVA: Analysis of Variance; ASAP: Adjoint Sensitivity Analysis Procedure; BRATS: Brain Tumor Segmentation Challenge; CE: Comparative Effectiveness; DG: Days Gained; DTI: Diffusion Tensor Imaging; DWMRI: Diffusion Weighted Magnetic Resonance Imaging; Fisher-KPP: Fisher-Kolmogorov-Petrovsky-Piskunov; FSAP: Forward Sensitivity Analysis Procedure; GAN: Generative Adversarial Networks; HMC: Hamiltonian Monte Carlo; KL divergence: Kullback-Leibler; LETKF: Local Ensemble Transform Kalman Filter; MAPEL: Mechanistic Axes Population Ensemble Linkage; MCMC: Markov Chain Monte Carlo; MTL: Multi-task learning; NLME: Nonlinear Mixed Effects; ODE: Ordinary Differential Equation; PCA: Principal Components Analysis; PDE: Partial Differential Equation; PDF: Probability Density Function; PINN: physics-informed neural networks; PMF: Prohorov Metric Framework; QoI: Quantity of Interest; RDE: Reaction-Diffusion-Equation; SSAE: Stacked Sparse Autoencoder; T1Gd: T1-weighted Gadolinium contrast enhanced; T2-FLAIR: T2-weighted FLuid-Attenuated-Inversion-Recovery; VAE: variational autoencoder; VEPART: Virtual Expansion of Populations for Analyzing Robustness of Therapies

1. Introduction

Heterogeneity in cancer is exhibited at both the intra-tumor and inter-patient levels [1]. Intra-tumor heterogeneity in the form of genotypic and phenotypic variability between tumor cells is present in most solid tumors [2] and has also been identified in hematopoietic cancers such as leukemia [3, 4]. Intratumor heterogeneity is driven by the introduction of genetic [5] and epigenetic alterations characterized by genomic instability and evolutionary selection of advantageous cell phenotypes from the cancer cell population [5, 6]. Recent studies have revealed that genetic heterogeneity does not necessarily correlate with phenotypic heterogeneity, since not all genetic alterations result in a phenotypic alteration, and only a few genetic alterations provide a fitness advantage [7].

Intratumor heterogeneity is also affected by the tumor microenvironment whereby gradients of positive and negative growth factors, such as oxygen and lactic acid, have direct effects on cell proliferation and survival. Heterogeneous tumor microenvironments also contribute to intercellular variability as tumor cells adaptively switch between different phenotypes, e.g., changing metabolic [8] or migratory states, in response to the availability of nutrients. Furthermore, distinct areas of a tumor may experience different immunologic responses which cause further heterogeneity within the tumor, and these immunological responses can be affected by local nutrient concentrations. For example, most solid tumors experience regions of hypoxic stress [9] which in turn shapes cell phenotype and can induce resistance to an immune response. The presence of this intratumor heterogeneity leads to variability in tumor response to drug treatment [10, 11] and makes predictions of tumor growth and spread particularly challenging [12]. In the context of tumor cell invasion in glioblastoma multiforme

(GBM), phenotypic heterogeneity is often modeled by using several equations to represent phenotypically distinct subpopulations of the tumor. For example, this approach has been used to model the 'go or grow' mechanism in which one subpopulation is more mobile and the other is more proliferative [13, 14], or to model phenotypic switching between hypoxic and normoxic subpopulations [15, 16].

The presence of intratumor heterogeneity is increasingly recognized as one of the primary drivers of heterogeneity in patient outcomes. The uniqueness and ongoing evolution of each patient's tumor microenvironment, genetic profile, and cell phenotype composition contributes to the difficulty in predicting the effect of cancer therapies. For example, tumor cells can change their microenvironment to induce immunusuppression [17–19] or to protect themselves from drug therapy [20]. A heterogeneous tumor microenvironment may also contribute to variability in drug penetration, creating inter-cellular variability from which therapy resistant cell populations or stem cell phenotypes can be promoted [21, 22]. A recent meeting that included cancer biologists, clinicians, and industry representatives highlighted the central problem with cancer therapy clinical trials under the current paradigm [7]: "...trials for which a single therapy agent is tested in a cohort with a matched biomarker do not provide information about the impact of heterogeneity or the longitudinal evolution of clonal or subclonal cells."

Mathematical modeling has successfully described the heterogeneity in the growth and progression of tumors from individual patient data [23]. For example, in [24], a mathematical model was used with MRI data from glioma patients to estimate patient-specific rates of cell proliferation and diffusion, revealing that patients could be stratified into those with diffuse tumors (tumors with a small rate of proliferation as compared to the rate of diffusion) versus nodular tumors (tumors with a high rate of proliferation as compared to the rate of diffusion). Patients with diffuse tumors (identified through the use of a mathematical model) showed no significant survival benefit with gross total resection (GTR) versus biopsy/subtotal resection (BX/STR). Nodular tumor patients on the other hand do exhibit a survival benefit with GTR. This is just one example of the insight that mathematical models provide for precision medicine, but mathematical models can also be used to predict a subject's response to drug treatment [25–27], infer differences in environmental factors between metastatic and benign tumors [28], or describe how metastatic cells spread [29]. We note that a subject here may refer to a tissue culture, mouse, or human, all of which are vital parts of the drug development process. There is a growing trend in the mathematical modeling community to focus on personalized mathematical models that investigate not only the mean patient response, but also how individual patients' responses will vary due to the above-mentioned sources of heterogeneity [25, 30]. Although large omics datasets exist due to high-throughput platforms, a current challenge is to integrate these data into a clinical setting [31, 32]. There has been some progress in connecting clinical and statistical investigators [33], but treatment outcomes will be improved if clinicians can continue to be informed by refined models that account for tumor heterogeneity and predict treatment outcomes.

The goal of this review article is to summarize mathematical, statistical, and machine learning methods that can be utilized for estimating intra- and inter-tumor heterogeneity from longitudinal data. This review article was a produced by members of the Tumor Heterogeneity Working Group as part of the 2018–2019 Program on Statistical, Mathematical, and Computational Methods for Precision Medicine which took place at the Statistical and Applied Mathematical Sciences Institute

(SAMSI). The aim of the SAMSI program was to facilitate interdisciplinary exchange between mathematical, statistical, computational, and health sciences researchers in developing data-driven methodology for precision medicine. We focus on the methods for estimating phenotypic heterogeneity from patient data with the use of a mathematical model rather than deducing genetic heterogeneity by clustering genomics data. Methods of modeling genetic, phenotypic, cell signaling, and spatial heterogeneity in systems biology have recently been reviewed in [34]. Here, we discuss the applicability of the methods for quantifying phenotypic heterogeneity to cancer prognostication and treatment. In section 2, we describe a canonical spatiotemporal Reaction-Diffusion-Equation (RDE) model of tumor growth, the Fisher-Kolmogorov-Petrovsky-Piskunov (Fisher-KPP) model. In section 3, we review the concepts and techniques for generating virtual populations, and their application to the study of tumor heterogeneity. We also generate a virtual population based on the Fisher-KPP model. This virtual population will be used in the subsequent sections for the illustration of the different methods for quantifying tumor heterogeneity. In sections 4 and 5, we discuss Nonlinear Mixed Effects Modeling (NLME) and non-parametric estimation for quantifying tumor heterogeneity. We illustrate an application of these two methods to a homogeneous virtual population. In section 6, we summarize Bayesian estimation methods. In section 7, we highlight the use of recent machine learning techniques for the study of tumor heterogeneity. In particular, we illustrate the application of generative adversarial networks (GANs) and variational auto-encoders (VAEs) for generating virtual populations. While we do not utilize clinical data in this paper, the generated data are intended to closely match the attributes of clinical data. By generating our own data, we are able to provide an evaluation of the reviewed methods by comparing estimated parameter and data distributions to their ground truth values. Most of the sections in this paper are accompanied by open-source code that can be accessed in a Github repository at https://github.com/jtnardin/Tumor-Heterogeneity/. This code provides data-driven tools for cancer modeling researchers that can be used to develop methods for connecting tumor heterogeneity to variability in treatment outcomes.

2. The Fisher-KPP model for tumor growth

Reaction-diffusion equation models were introduced into the cancer modeling literature in the mid 1990's by several groups [35–37] and have been employed with preclinical and clinical data [38–42]. To generate synthetic data sets for demonstrating the performance of various methods for quantifying tumor heterogeneity in a tutorial-like setting, we choose to focus on the Fisher-KPP Reaction-diffusion equation model of tumor growth in the absence of treatment for this review. While this model does not directly account for tumor necrosis, vascularity, immune response or treatment response, it is frequently used to model tumor growth from longitudinal imaging data, so we focus on it for the rest of this review [23, 24]. If one instead has access to a time-varying scalar data (such as tumor volume over time), then an ordinary differential equation, such as the logistic equation, would be appropriate [30, 43, 44]. We will state a general version of the PDE model first, then in later sections we discuss simplified parameterizations that are useful for demonstrating key statistical concepts. Several of our assumptions are motivated by clarity instead of accuracy to any clinical or preclinical context.

The Fisher-KPP model for tumor growth in the absence of treatment takes the form of a nonlinear reaction-diffusion type Partial Differential Equation (PDE) for a density of tumor cells $\boldsymbol{n} = n(\boldsymbol{x}, t)$

(units number per unit volume):

$$\begin{cases} \frac{\partial \boldsymbol{n}}{\partial t} = \nabla \cdot (\boldsymbol{D}\nabla \boldsymbol{n}) + \boldsymbol{\rho}\boldsymbol{n}\left(1 - \frac{\boldsymbol{n}}{\boldsymbol{\kappa}}\right) & (\boldsymbol{x}, t) \in V \times [0, T] \\ n(\boldsymbol{x}, 0) = n_0(\boldsymbol{x}) & \\ \mathcal{B}\boldsymbol{n} = 0 & \text{boundary cond.} \end{cases} \tag{2.1}$$

The spatial domain $V \subset \mathbb{R}^d$ corresponds to the biological domain the tumor is growing in, which we take to be the planar set $V = [0, 1]^2$ for convenience. The units of $(\boldsymbol{x}, t)$ are centimeters and days, respectively, and the units of all other quantities follow by dimensional analysis.

The boundary condition $\mathcal{B}\boldsymbol{n}$ can be taken to be Dirichlet, no-flux or free; we assume homogeneous Dirichlet, that is $n(\boldsymbol{x}, t) = 0$ on the domain boundary, for convenience. We also assume that the initial condition $\boldsymbol{n}_0$ is fixed and known, that all tumors grow from an initial population of $N_0 = 5$ cells localized around a central point $(x_0, y_0) = (0.5, 0.5)$ according to an isotropic Gaussian distribution with standard deviation $\sigma = 0.01$cm (i.e., 100 μm):

$$n_0(\boldsymbol{x}) = \frac{N_0}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}\left((x - x_0)^2 + (y - y_0)^2\right)\right). \tag{2.2}$$

In its full generality, the model parameters $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$ are permitted to be spatiotemporally inhomogeneous fields, i.e., $\boldsymbol{D} = D(\boldsymbol{x}, t), \boldsymbol{\rho} = \rho(\boldsymbol{x}, t), \boldsymbol{\kappa} = \kappa(\boldsymbol{x}, t)$. Such generality allows the user to model intrapatient and intratumoral heterogeneity, but owing to statistical or computational complexity, it is frequently convenient to restrict one or all of these fields to be constant or piecewise constant in time and/or space. Such simplifications are assumed in sections 4–6 below, while sections 3 and 7 maintain the spatially varying parameters. Extending the methods discussed in sections 4–6 to the fully spatial case is a subject of future work. In any case, we assume that $\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa}$ are described by a p-dimensional parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$. For example, in the fully homogeneous case the three coefficients in (2.1) are assumed constant, i.e., $D(\boldsymbol{x}, t) \equiv D_0, \rho(\boldsymbol{x}, t) \equiv \rho_0$ and $\kappa(\boldsymbol{x}, t) \equiv \kappa_0$, and hence we can take $\boldsymbol{\beta} = (D_0, \rho_0, \kappa_0) \in \mathbb{R}^p$ with $p = 3$. For the spatially inhomogenous case, $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$ are functions of a spatial variable $\boldsymbol{x}$, so we must select a parameterization that is able to generate such spatial variability using a finite-dimensional $\boldsymbol{\beta}$. This is achieved by way of a *field synthesis map*, which is constructed as follows. First, an appropriate set of functions $\mathbb{X}$ is selected for the PDE parameter fields, say $(D(\boldsymbol{x}), \rho(\boldsymbol{x}), \kappa(\boldsymbol{x})) \in \mathbb{X} = \mathbb{X}_{\boldsymbol{D}} \times \mathbb{X}_{\boldsymbol{\rho}} \times \mathbb{X}_{\boldsymbol{\kappa}}$, where $\mathbb{X}$ is a set of coefficient functions selected to guarantee classical well-posedness of (2.1). Then, a linear or nonlinear synthesis function $\boldsymbol{\Phi}$ is constructed that maps parameter vectors $\boldsymbol{\beta} \in \mathbb{R}^p$ to their field counterparts in $\mathbb{X}$, that is $\boldsymbol{\Phi} : \mathbb{R}^p \to \mathbb{X}$. For instance, selecting a set of basis functions $\phi_1(\boldsymbol{x}), \ldots, \phi_{p_D}(\boldsymbol{x})$, we can synthesize a spatially inhomogenous diffusion parameter by the linear mapping:

$$D(\boldsymbol{x}; \boldsymbol{\beta}) = (\boldsymbol{\Phi}_D \boldsymbol{\beta})(\boldsymbol{x}) = \sum_{m=1}^{p_D} \beta_m \phi_m(\boldsymbol{x}), \tag{2.3}$$

where p_D denotes the number of basis functions. Note that one must be careful to select $\boldsymbol{\phi}_m$ and $\boldsymbol{\beta}$ such that the resulting diffusion coefficient is non-negative. Selecting similar basis sets for the coefficients $\boldsymbol{\rho}$ and $\boldsymbol{\kappa}$ results in a complete parameter-to-coefficient linear synthesis map $\boldsymbol{\Phi}$. More generally, a nonlinear synthesis map such as the model (3.3) below can be employed. In machine learning, a field

synthesis map is analogous to a generative model, such as those described in section 7, that produces samples from a distribution by feeding a latent noise vector through a neural network.

To model both population heterogeneity and patient-specific uncertainty we treat the parameter vector $\boldsymbol{\beta}$ as random, with distribution $\mathbb{P}_{\beta}$. For the homogeneous parameter case, $\boldsymbol{\beta}$ is a three-dimensional random vector. In the more general spatially inhomogeneous case, $\boldsymbol{\beta}$ is a p-dimensional random vector, where $p \geq 3$ is the total number of parameters employed to describe the spatial variability. The synthesis map $\boldsymbol{\Phi}$ then allows us to treat the PDE coefficients $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$ as spatial random fields, since they are a function of both a random component ($\boldsymbol{\beta}$) and a spatial index $\boldsymbol{x} \in V$. More explicitly, denote a realization of the random vector $\boldsymbol{\beta}$ as $\boldsymbol{\beta}_{\omega}$. Then, we can express $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$ as a vector-valued random field as follows:

$$(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa}) : V \times \Omega \to \mathbb{R}^3, \quad (\boldsymbol{x}, \omega) \mapsto (D(\boldsymbol{x}, \beta_{\omega}), \rho(\boldsymbol{x}, \beta_{\omega}), \kappa(\boldsymbol{x}, \beta_{\omega})). \tag{2.4}$$

The induced probability distribution of the coefficient fields is, by construction of the synthesis map, guaranteed to be supported in a set $\mathbb{X}$ of functions such that (2.1) is well-posed for all (or almost all) $\boldsymbol{\beta}$. For example, we can require that for almost all $\boldsymbol{\beta}$, the synthesized carrying capacity $\boldsymbol{\kappa}(\beta)$ satisfies $\kappa(\boldsymbol{x}; \boldsymbol{\beta}) \geq \kappa_0 > 0$ for all $\boldsymbol{x} \in V$. Note that the ubiquitous Gaussian random field model does not satisfy such a condition; we discuss a particular non-Gaussian random field model below that does provide such a guarantee, and refer readers to [45–47] for a more complete background on the theory of random and stochastic differential equations. Note that we have been careful to ensure that (2.1) can be treated path-wise, i.e., without resorting to the theory of stochastic calculus. See [48] for a discussion of the more general case where equations similar to (2.1) are treated in a fully statistical manner using functional calculus.

We remark that, to accurately infer a general random field $\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa}$ as a random function of spatial coordinate $\boldsymbol{x}$ from clinical data requires a large dataset, curated from a large patient population. To our knowledge, such a dataset is not currently available. However we can simulate these random fields and ensure these fields are consistent with the available data, by using statistical methods to create virtual patients and virtual populations.

In section 3 below, we will introduce the concept of a *virtual patient*. In the context of the model (2.1), each realization of $\boldsymbol{\beta} \sim \mathbb{P}_{\beta}$ corresponds to a virtual patient. Note that the full statistical description of $\mathbb{P}_{\beta}$ may require additional population parameters to be fully specified (e.g., means and (co)variances), and we denote these parameters by $\boldsymbol{\theta}$ throughout this study and denote the probability distribution with parameter $\boldsymbol{\theta}$ as $\mathbb{P}_{\beta|\theta}$, or alternately with a Probability Density Function (PDF) $p(\boldsymbol{\beta}|\boldsymbol{\theta})$ (abusing notation by using the variable $\boldsymbol{\beta}$ to denote both the random variable and the argument of its PDF). As the probability distribution $\mathbb{P}_{\beta}$ is a precise mathematical description of a heterogeneous population, as modeled via the patient-specific parameter $\boldsymbol{\beta}$ and the PDE (2.1), *quantifying inter-patient heterogeneity in the modeling context of this paper is analogous to performing statistical inference on* $\mathbb{P}_{\beta}$, for instance estimating the mean $\bar{\boldsymbol{\beta}}$ and covariance matrix $\boldsymbol{K}_{\beta}$.

Given a realized parameter $\boldsymbol{\beta}$, it is usually the primary goal to predict the future state or some vector Quantity-of-Interest (QoI) $\boldsymbol{y} \in \mathbb{R}^q$. Such a prediction takes the form of a mathematical model

$$\boldsymbol{y}^{(model)} = \boldsymbol{\mathcal{M}}(\boldsymbol{\beta}). \tag{2.5}$$

The dimension q of $\boldsymbol{y}$ is typically small, as this vector tends to only contain a limited number of human-interpretable, and preferably observable, quantities of interest (QoI). While in the formulation (2.5) we

have $\mathcal{M} : \mathbb{R}^p \to \mathbb{R}^q$, it is worth noting that it may be necessary to first predict a high- or infinite-dimensional quantity to compute the final QoI $\boldsymbol{y}$. For example, assuming that (2.1) is well-posed for all $\boldsymbol{\beta}$, we can first construct a deterministic *parameter-to-solution map* for the PDE, written as:

$$n^{(model)}(\boldsymbol{x}, t; \boldsymbol{\beta}) = \mathcal{M}^{(PDE)}(\boldsymbol{\beta}). \tag{2.6}$$

From the solution (2.6), we can further predict (say) the total tumor size (in number of cells) at some fixed time T via

$$N^{(model)}(\boldsymbol{\beta}; T) = \int_V n^{(model)}(\boldsymbol{x}, T; \boldsymbol{\beta})\, d\boldsymbol{x} = \int_V \mathcal{M}^{(PDE)}(\boldsymbol{\beta})(\boldsymbol{x}, T)\, d\boldsymbol{x} = \mathcal{M}(\boldsymbol{\beta}). \tag{2.7}$$

In (2.7), the final QoI model can be written as a mapping $\mathcal{M} : \mathbb{R}^p \to \mathbb{R}$, despite the fact that (2.7) first computes a spatiotemporal function via (2.6). Note that typically, (2.1) does not have an explicit closed-form solution, so it may be necessary to numerically approximate the parameter-to-solution map (2.6), say at M spatial locations over N timepoints. In this case, we can consider the numerical solution at the $q = NM$ points as a large vector QoI. Thus, when convenient, $\boldsymbol{y}^{(model)}$ may also represent a high-dimensional quantity such as a tumor cell density defined on a computational grid. From such a computational solution, a final low-dimensional QoI such as (2.7) can be estimated.

We employ the commonly used modeling assumption that the true value of $\boldsymbol{y}$, denoted $\boldsymbol{y}^{(true)}$, is related to the model prediction via

$$\boldsymbol{y}^{(true)} = \boldsymbol{y}^{(model)} + \boldsymbol{\epsilon}^{(model)} = \mathcal{M}(\boldsymbol{\beta}) + \boldsymbol{\epsilon}^{(model)}, \tag{2.8}$$

where $\boldsymbol{\epsilon}^{(model)}$ is a model error, which may consist of both systematic error (including numerical discretization errors) and calibration noise, and which depends on $\boldsymbol{\beta}$. A measured data point $\boldsymbol{y}^{(meas)}$ will be related to $\boldsymbol{y}^{(model)}$ via

$$\boldsymbol{y}^{(meas)} = \boldsymbol{y}^{(true)} + \boldsymbol{\epsilon}^{(meas)} = \boldsymbol{y}^{(model)} + \boldsymbol{\epsilon}^{(model)} + \boldsymbol{\epsilon}^{(meas)}, \tag{2.9}$$

where $\boldsymbol{\epsilon}^{(meas)}$ is a measurement noise, which typically has mean zero indicating an unbiased measurement method. We discuss measurement error further in section 6.

Note that as written, $\mathcal{M}(\boldsymbol{\beta})$ is potentially over-parameterized, particularly if the dimension of the QoI is small and a large number of parameters have been employed to account for spatial variability. This leaves open the opportunity of performing model order reduction, which seeks to eliminate some parameters by selecting $\boldsymbol{\beta}' \in \mathbb{R}^{p'}$ with $p' < p$. We briefly discuss sensitivity-based model reduction strategies in section 6.

3. Virtual populations

A virtual individual is a digital representation of a biological counterpart (e.g., a cell, mouse, or patient), which is comprised of a set of parameters describing the biological characteristics of the individual. Some parameters can be directly measured from the real individual, for example, the results of a blood panel, gene sequencing data, imaging data, or other time series measurements. Other parameters may not be directly measurable, but can be estimated by fitting a mathematical model to available clinical or experimental data. For example, the logistic ordinary differential

equation model can be used to estimate a patient's tumoral carrying capacity from tumor volume data over time [30, 43]. A virtual population, defined as an ensemble of many virtual individuals, provides a natural *in silico* platform to investigate tumor heterogeneity.

Mathematically, a virtual individual is characterized by a vector parameter $\boldsymbol{\beta}$, which can in principle be high- or infinite-dimensional to model functional data such as time series, scalar and vector field quantities. Spatiotemporal parameters can effectively quantify many forms of intra-patient and intra-tumoral heterogeneity, and their (direct or indirect) measurement is usually possible in the laboratory or clinic via biosensors and imaging techniques [49]. In this paper, we will assume for simplicity that $\boldsymbol{\beta} \in \mathbb{R}^p$ is finite-dimensional, so all spatiotemporal quantities are assumed to be approximated by an appropriate linear or nonlinear synthesis map. For example, a spatially varying diffusion coefficient $D(\boldsymbol{x})$ could be approximated by its values on a discrete grid, or alternately by its projection onto a finite-dimensional subspace.

Inter-patient (i.e., population) heterogeneity is then described by treating $\boldsymbol{\beta}$ as a random vector with probability distribution $\mathbb{P}_{\boldsymbol{\beta}}$. A realization of $\boldsymbol{\beta}$ is denoted as $\boldsymbol{\beta}_k$, corresponding to 'virtual individual k', so that a collection of virtual individuals $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ corresponds to a virtual population. In the setting of precision medicine, it is frequently useful to incorporate patient-specific uncertainty by considering samples from the conditional distribution $\mathbb{P}_{\boldsymbol{\beta}|\boldsymbol{g}_k}$, where $\boldsymbol{g}_k$ is some patient-specific data such as quantitative imaging and/or blood test results. We will provide a more detailed review on patient-specific modeling in section 6. Note that unless otherwise necessary, we will use the same notation for both a population and patient-specific random vector, namely the random vector $\boldsymbol{\beta}$ has realizations $\boldsymbol{\beta}_k$ and distribution $\mathbb{P}_{\boldsymbol{\beta}}$, but it is important to recognize that that this vector is random for different reasons in the population and patient-specific cases. In a Bayesian mindset, a sample from a population is analogous to a sample from a prior distribution, while a patient-specific virtual patient is a sample from a posterior distribution. Note also that the statistical description of $\mathbb{P}_{\boldsymbol{\beta}}$ can either be parametric, i.e., $\mathbb{P}_{\boldsymbol{\beta}} = \mathbb{P}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$, with $\boldsymbol{\theta}$ being a finite-dimensional population parameter, or non-parametric where no such parametric assumption is made. Sections 3,4,6 and 7 focus on parametric descriptions, while section 5 discusses a patient-specific non-parametric estimation technique.

3.1. The virtual population framework

When the probability distribution $\mathbb{P}_{\boldsymbol{\beta}}$ is known either explicitly or implicitly, a virtual population can be generated directly by drawing samples $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ from $\mathbb{P}_{\boldsymbol{\beta}}$, as will be demonstrated in section 3.3. Depending on the complexity of $\mathbb{P}_{\boldsymbol{\beta}}$, this may be computationally straightforward, e.g., for low-dimensional uniform or Gaussian random $\boldsymbol{\beta}$, or more demanding, for example when $\boldsymbol{\beta}$ parameterizes functional parameters such as diffusion coefficients or a drug concentration $c(\boldsymbol{x}, t)$ which may require solving an auxiliary random PDE in order to sample, or when $\boldsymbol{\beta}$ is conditional on patient-specific data, i.e., sampling a posterior distribution $\mathbb{P}_{\boldsymbol{\beta}|\boldsymbol{g}_k}$.

In practice, however, the probability distribution $\mathbb{P}_{\boldsymbol{\beta}}$ is often partially or completely unknown. In such situations, one can estimate the underlying population probability distribution, $\mathbb{P}_{\boldsymbol{\beta}}$, from empirically sampled data that may be directly or indirectly related to $\boldsymbol{\beta}$ [50, 51]. One possible approach is to define pre-specified parameter ranges that are known to generate biologically feasible output from the mathematical model. For example, supposing that $\boldsymbol{\beta} = (D_0, \rho_0, \kappa_0)$ is a spatially homogeneous parameter in (2.1), it may be possible to determine a set of intervals for each

component of $\boldsymbol{\beta}$ that will lead to biologically plausible output from the QoI model, for instance non-negative finite tumor sizes. Generating virtual patients by initially sampling uniformly from this predefined plausible parameter range allows us to sufficiently explore the broad range of responses that are possible. However, these preliminary model predictions might not reflect the distribution of some observed population-level data, which means the generated virtual populations do not have the correct population characteristics.

Several methods have been developed to address this issue [50–53]. For example, in the method presented by Allen et al. [50], a "plausible patient population" is generated by first matching predefined input (parameter) ranges and output (quantity-of-interest) ranges. Following our notation, this method supposes initially that $\mathbb{P}_{\boldsymbol{\beta}}$ is a uniform distribution on some set $L^{(input)} \subset \mathbb{R}^p$. Then, defining a plausible range of outputs $\boldsymbol{y}^{(model)} \in L^{(output)} \subset \mathbb{R}^q$, we say that a parameter $\boldsymbol{\beta}_k \in \mathbb{P}_{\boldsymbol{\beta}}$ is *plausible* if $\boldsymbol{\mathcal{M}}(\boldsymbol{\beta}_k) \in L^{(output)}$. The method proposed in [50] suggests generating a virtual population by first sampling plausible patients by applying a simulated annealing algorithm to an optimization objective designed to enforce the plausibility constraint. Then, a given plausible patient $\boldsymbol{\beta}_k$ is included in the final virtual population with some probability, calculated based on an observed QoI distribution. The inclusion probability is based on: i) probability of inclusion for the whole plausible population, and ii) the ratio of the output probability density of the observed population to an estimated probability density for the plausible population. In this sampling step, the probability of inclusion for the whole plausible population is optimized until the generated virtual population (which is a subset of plausible patients) matches the distribution of the observed population data by some measure of statistical similarity such as the Kolmogorov-Smirnov test statistic.

It is worth noting that the ordinary accept-reject sampling approach presented in [50] suffers from poor sampling efficiency since it searches a large space of parameters and generates a significant number of plausible but unlikely parameter realizations that are ultimately rejected. Rieger et al. [51] improved upon this methodology by introducing more efficient sampling methods which include nested simulated annealing, a genetic algorithm, and Metropolis-Hastings-based importance sampling, and compared these different methods in their work. Another approach called the MAPEL (Mechanistic Axes Population Ensemble Linkage) algorithm, presented in [53], also follows the two-step process of generating a cohort of plausible patients followed by matching statistics to an observed population. Instead of rejecting plausible patients, however, the MAPEL method computes a prevalence weight for each plausible patient, with the prevalence weight computed to match the model output statistics to response bins measured in a clinical trial, according to a composite goodness-of-fit objective that acts at the level of histogram bins. While in theory each virtual patient could have a distinct prevalence weight, the MAPEL method elects to divide the parameter axes into bins, assigning a uniform prevalence weight to each bin (effectively modeling the parameter distribution $\mathbb{P}_{\boldsymbol{\beta}}$ as a multinomial distribution). A Nelder-Mead optimization method is then employed to update the prevalence weights.

Other methods for generating virtual patient populations for use in *in silico* clinical trials may also be applicable for the study of tumor heterogeneity. We refer readers to Chapter 5 of [54] for a more complete overview of recent successes in computer-aided clinical trials, and to [55] for a more theoretical discussion of virtual clinical trials.

3.2. Previous literature of virtual populations for tumor heterogeneity

Virtual populations have been employed in various contexts that are relevant to the study of tumor heterogeneity in both the drug development and clinical contexts. Specifically, virtual cell populations, mouse populations and human patient populations can be employed to understand the sources of heterogeneity at different stages of drug development.

The main motivation for generating *virtual cell populations* is to characterize population-level responses to a perturbation such as anti-cancer treatments. Due to the variations in the concentration of intracellular signaling species at the time of the perturbation, different cells can have different response to the same perturbation. Albeck et al. [56] study variability in cell apoptosis by generating a virtual cell population from a mathematical model of intracellular signaling networks. They validated this model with experimental data, and suggest that this virtual population technique can be applied to the development of pro-apoptotic cancer therapies. Note that an initial distribution of cell states is needed in order to simulate the cell population's response. For example, in [57] a gamma distribution was assumed for the concentrations of intracellular signaling species [58].

Once a potential anti-cancer treatment has been discovered, *virtual mouse populations* may aid the analysis and therapeutic design for preclincial studies of a potential drug treatment. In [25], a technique called the Virtual Expansion of Populations for Analyzing Robustness of Therapies (VEPART) was introduced and used to study the robustness of a tumor population in response to different treatment protocols. Given time course data on tumor volume and a mathematical model that has been initially fit to aggregate data, a non-parametric bootstrap technique is used to amplify the sample population to a large cohort of statistically plausible patients. This method is illustrated on mice with melanoma that receive oncolytic virus (OV) or dendritic cell (DC) vaccines. The virtual population was then used to design the optimal ordering of three doses of the OV and three doses of the DC vaccine (e.g., OV-OV-OV-DC-DC-DC designates three consecutive OV doses followed by three administrations of the DC vaccine over 6 days). Each treatment was ranked by how often it was found to be the optimal treatment for the entire virtual population. The current standard of care was ranked as the best treatment in some cases, but only lead to tumor eradication in 43% or less of the virtual population. By changing the doses of OV and DC, however, the authors showed that a different treatment schedule can lead to a robust treatment that eradicates tumors in 84% of the virtual population.

Moving up to a clinically relevant scale, *virtual patient populations* have been employed in a wide array of medical contexts, ranging from predicting outcomes for blunt force trauma patients [59] to studying insulin sensitivity in type 2 diabetes [52], and have been suggested as a method for evaluating patient-specific chemotherapy regimens [55]. For drug development, it is important to predict not only the mean response to an intervention but also the overall distribution of the population responses. In particular, virtual populations have been employed to help identify the key mechanistic parameters that give rise to variation in the response [53]. Virtual patient populations have also been used for evaluating medical radiation safety and performance of imaging systems [60, 61] and to evaluate various radiotherapy protocols [62]. In the context of medical device assessment, a virtual patient is usually called a *computational human phantom*, and several phantom models employ randomization, morphing and posing to generate a virtual patient population that mimics a real population. While computational phantoms have historically emphasized anatomical structures, others are working to also incorporate functional (i.e., physiological) mechanisms, and thus may eventually provide a compelling platform for treatment evaluation in cancer therapy.

See [61] for a review of the current state-of-the-art in computational human phantoms.

In addition to the development and optimization of therapeutic agents, virtual populations have been used to guide the design of cancer biomarkers. Depending on the cancer, more than a decade may be needed to finish the Randomized Controlled Trials (RCTs) to validate the effectiveness of diagnostic markers. In contrast, emerging biomarkers may diffuse into clinical practice long before survival data becomes available. In the interim, Comparative Effectiveness (CE) studies typically collect data on how testing for a biomarker impacts an intermediate endpoint for treatment recommendations. For example, [63] shows an example of using a virtual population to help evaluate the effectiveness of biomarkers to target cancer treatment. It presents a statistical simulation tool that facilitates the modeling of the mortality impact of interventions. Given CE studies of typical intermediate endpoints, i.e., the common pre-mortality outcomes used to demonstrate the effectiveness of interventions, bootstrapping was used to generate a virtual population to mimic the population in the CE study. The treatment and mortality of the virtual population was then modeled to understand the long-term potential of emerging diagnostic biomarkers. In [64], a virtual population is generated to simulate the population response to a clinically tested therapy. This provides a mechanistic explanation into the heterogeneous response to the treatment and also helps identify several potential prognostic biomarkers.

Virtual population techniques have also been used to understand the impact of healthcare system administration on real patients. In [65], for example, an agent-based model representing colon and colorectal cancer patients was developed which includes a biological cancer evolution model and patients that interact with the healthcare system, including physicians, nurses and equipment. Simulations show promising interpolation results with respect to chemotherapy dosage and radiotherapy dosage. However, the model's ability to interpolate different administration protocols is still limited, and therefore calibration is required for each protocol.

3.3. Case study: Virtual population data generation from the Fisher-KPP model

In this section, we present the details of generating simulated tumor growth data with the Fisher-KPP model. We will generate two such virtual populations arising from either: i) a spatially-homogeneous model with inter-patient heterogeneity or ii) a spatially-heterogeneous model with intra-patient heterogeneity. These two datasets provide an ideal platform for demonstrating the performance of various methods for quantifying tumor heterogeneity presented in the subsequent sections. In Table 1, we detail which methods will be used to interpret these datasets through case studies found throughout this article. Each virtual population is produced by simulating a randomized Fisher-KPP model, with differences between the virtual populations arising from differences between the random field synthesis method employed for the model parameters $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$, as well as different underlying probability distributions $\mathbb{P}_{\boldsymbol{\beta}}$.

We will select a statistical description for $\boldsymbol{\beta}$ and corresponding random field synthesis method to demonstrate two tumor types: Spatially homogeneous and spatially heterogeneous tumors. For *homogeneous* tumors, $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$ are the same for all spatial points, i.e., $\boldsymbol{D} \equiv D, \boldsymbol{\rho} \equiv \rho$ and $\boldsymbol{\kappa} \equiv \kappa$, and hence $\boldsymbol{\beta} = (D, \rho, \kappa) \in \mathbb{R}^3$. We will only randomize the growth parameter ρ, fixing $D = 10^{-6}, \kappa = 10.001$. Note that the nominal value of $\kappa = 10.001$ selected here was chosen arbitrarily since it does not influence any subsequent analysis performed on the homogeneous dataset; a more biologically realistic value would be straightforward to incorporate. We sample ρ from the probability

Table 1. Description of which datasets will be investigated in through case studies in later sections of this article. "Homogeneous" refers to the dataset arising from the spatially-homogenous Fisher-KPP model with inter-patient heterogeneity. "Heterogenous" refers to the dataset arising from the dataset arising from the spatially heterogeneous Fisher-KPP model with intra-patient heterogeneity.

Method	Sections	Dataset
NLME	4.3	Homogeneous
Non-parametric Estimation	5.2	Homogeneous
Bayesian Methods	6.5	Homogeneous
Virtual Populations	3.3	Heterogeneous
Machine Learning	7.2,7.3	Heterogeneous

distribution given by

$$\rho = \rho_0 + I_k \rho_M \tag{3.1}$$

where

$$\rho_0 \sim \mathcal{N}(\mu_0, \sigma_0^2), \;\; \rho_M \sim \mathcal{N}(\mu_M, \sigma_M^2) \tag{3.2}$$

and $I_k = 0$ for patients with benign tumors and $I_k = 1$ for patients with malignant tumors. We obtain a set of 25 tumor density data $\boldsymbol{y}_k = y_{i,j,k}$, with $y_{i,j,k} \approx n_k(\boldsymbol{x}_i, t_j)$ denoting the approximated cell density at spatiotemporal point $(\boldsymbol{x}_i, t_j)$ and $k = 1, 2, \ldots, 25$ representing the virtual patient. We produced 15 benign tumors and 10 malignant tumors. The parameter ρ thus samples from a bimodal distribution comprised of benign (low ρ values) and malignant (high ρ values) tumors. The parameter μ_0 corresponds to the mean benign tumor rate of growth, and μ_M, corresponds to the mean increase in the rate of growth for malignant tumors. The population parameters σ_0^2, σ_M^2 are the variances associated with these values. We set $\mu_0 = 1 \times 10^{-3}, \sigma_0 = 2 \times 10^{-4}, \mu_M = 0.03$, and $\sigma_M = 1 \times 10^{-2}$ for this data set. Once we have sampled ρ_k from Eq (3.1), we generate $\boldsymbol{\beta}_k = (D, \rho_k, \kappa)$ and simulate the Fisher-KPP solution as $\boldsymbol{y}_k = \boldsymbol{\mathcal{M}}(\boldsymbol{\beta}_k)$ using a numerical approximation to Eq (2.6). The data set generated in this homogeneous setting will be analyzed in sections 4.3 and 5.2.

For spatially *heterogeneous* tumors, $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$ are assumed to be realizations of spatial random fields as described below. The inhomogeneous data will be employed for Bayesian-specific Bayesian inference in section 6.1 and for methods of machine learning in sections 7.2 and 7.3. We assume that $\boldsymbol{D} = D(\boldsymbol{x}), \boldsymbol{\rho} = \rho(\boldsymbol{x})$ and $\boldsymbol{\kappa} = \kappa(\boldsymbol{x})$, and model each as independent realizations of a 'lumpy-type' random field model [66, 67]. For instance, $\boldsymbol{D}$ is generated by

$$D(\boldsymbol{x}, \omega) = \sum_{l=1}^{L(\omega)} \ell\left(\boldsymbol{x} - \boldsymbol{c}_l(\omega); \boldsymbol{\gamma}\right), \tag{3.3}$$

and the formula for generating $\boldsymbol{\kappa}$ and $\boldsymbol{\rho}$ can be written analogously. In (3.3), the 'lump' function $\ell(\boldsymbol{x}; \boldsymbol{\gamma})$ is always bounded, integrable and strictly positive throughout the support set V. The number of 'lumps' $L(\omega)$ is Poisson distributed with mean $\bar{L}$, and the random 'lump centers' $\boldsymbol{c}_l$ are independent and identically distributed (i.i.d.) with distribution $\boldsymbol{c}_l \sim \mathbb{P}_{\boldsymbol{c}}$. We will assume for simplicity that the lumps have fixed shape $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, that the centers are uniformly distributed in V. We also assume the

lump function is an isotropic Gaussian:

$$\ell(\boldsymbol{x};\boldsymbol{\gamma}) = A\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{x}\|^2\right), \quad \boldsymbol{\gamma} = (A,\sigma^2). \tag{3.4}$$

In this case, the random field model is fully specified by the population parameter $(\bar{L}, A, \sigma^2)$. This parameter can be estimated for a given population by applying statistical estimation techniques; see e.g., [68] and section 6 below. Generalizations of these assumptions lead to more intricate field statistics, but at the expense of more parameters to specify and hence significantly more clinical data from which to infer reasonable distributions (see [69], for example). Note that we have also assumed that the three fields are statistically independent; this assumption is likely not supported by clinical data, but could be generalized, for instance by introducing a nontrivial statistical coupling between the lump centers in (3.3). Inferring the actual structure of this coupling, for instance the mutual covariances between the field parameters, would require extensive data and is a topic of future work. See also [70] in this issue for further discussion.

The model (3.3) is flexible and rapid to simulate, and realizations can be guaranteed to satisfy conditions required for well-posedness of (2.1) such as strict positivity. The random field $\boldsymbol{\beta}$ is fully specified by the 9-dimensional population parameter $\boldsymbol{\theta} = (\bar{L}_\rho, \bar{L}_D, \bar{L}_\kappa, A_\rho, A_D, A_\kappa, \sigma^2_\rho, \sigma^2_D, \sigma^2_\kappa)$; the choice of $\boldsymbol{\theta}$ we employed for the example virtual population is shown in Table 2, and example realizations are shown in Figure 1. Once a sample of the parameter vector $\boldsymbol{\beta}$ has been drawn, the PDE model (2.1) is simulated forward in time for each $\boldsymbol{\beta}_k$ using a finite difference method-of-lines approach [71], employing Matlab's `ode45` integrator for the time stepping. The result is a vector $\mathbf{y}_k^{(model)} = y_{i,j,k}^{(model)} = \boldsymbol{\mathcal{M}}(\boldsymbol{\beta}_k)$, where $y_{i,j,k}^{(model)}$ is the approximate cell density $y_{i,j,k} \approx n_k^{(model)}(\boldsymbol{x}_i, t_j)$ for 'patient' k at grid position $\boldsymbol{x}_i$ and time t_j (see [70] for further discussion of this numerical technique).

Table 2. Population parameters $\boldsymbol{\theta}$ employed to generate the test tumor population. These parameters were hand-tuned to obtain reasonable model output.

Parameter	$\bar{L}$	A	σ^2
Growth rate $\rho(\boldsymbol{x})$	200	1e-2	2e-3
Diffusion coefficient $D(\boldsymbol{x})$	20	8e-8	4e-2
Carrying capacity $\kappa(\boldsymbol{x})$	20	5e7	5e-2

Using the population parameters $\boldsymbol{\theta}$ provided in Table 1 for the lumpy model (3.3), we generate a set of $K = 128$ sampled realizations of $\boldsymbol{\beta}$ and $\boldsymbol{y}_k^{(model)} = \boldsymbol{\mathcal{M}}(\boldsymbol{\beta}_k)$. In the top row of Figure 1, four realizations of the tumor cell density at a fixed time are shown, demonstrating a wide variety of tumor morphology during the initial growth phase. In the bottom row of Figure 1, we compare these virtual tumor cell densities to a collection of four Apparent Diffusion Coefficient (ADC) maps obtained from a publicly available clinical imaging dataset [72]. A 2D slice of the 3D ADC map for each tumor was extracted using a segmentation mask provided in the dataset. While ADC maps have been related to tumor cellularity [41], we provide these images for qualitative comparison only, and no parameter estimation step was attempted to match the virtual tumors to the real ADC maps. Such parameter estimation is discussed in [70]. In Figure 2, a single realization is shown for four time points, together with the corresponding random field realizations and the total cell number $N(t)$. In Figure 3, population

heterogeneity in the quantity-of-interest $N(t)$ is demonstrated by plotting each of the 128 sample $\boldsymbol{y}_k = N_k(t)$ computations together with the sample average $\bar{N}(t)$ and the resulting histogram of $p_{N(x,t)}$ for several time points. In Figure 4, we display the sample average tumor cell density for four times, which demonstrates a distinctly uniform 'spheroid' growth.

For the illustration of methods for quantifying heterogeneity in subsequent sections, we take the resulting simulated tumor cell density population $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K$, and any quantity-of-interest functionals computed from these solutions, to be the ground-truth for subsequent experiments. Thus, we treat a given realization of the coefficient field generated using (3.3) to be the true latent parameter, and the corresponding QoI functionals applied to the model solution (2.6) as $\boldsymbol{y}^{true}$. Subsequent models may make further simplifying assumptions, such as spatial homogeneity, which will introduce modeling errors $\boldsymbol{\epsilon}^{model}$ relative to our simulated ground-truth. While we have elected to use virtual tumors as the ground-truth in this review, this is only for clarity and simplicity in describing subsequent statistical methodologies. Applying these methodologies to real clinical or preclinical data is feasible, but requires more careful data processing and is the topic of future work.

4. Nonlinear mixed effects modeling

Nonlinear mixed effects (NLME) modeling, also known as hierarchical nonlinear modeling, is a statistical framework for simultaneously characterizing both population-level behavior and individual variations. The response behavior can be described using a parameterized mathematical model, from which we can deduce both the average population parameter values as well as how individual parameter values vary from these mean values. The ultimate aim is to characterize the typical behavior of the population, and the extent to which the behavior varies across the individuals, as well as whether or not the variation is associated with individual attributes [73].

As an example, assume we aim to understand the intra-subject processes underlying tumor spread and growth from a large dataset of measured patient tumor volumes over space and time. An example of such dataset could be the virtual population data described in section 3.3. Applying the NLME framework for the fully inhomogeneous random differential equation would be a challenging task due to the large number of parameters required to account for spatial variability. Thus, in this section we will consider the homogeneous model for which $(\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa}) = (D, \rho, \kappa) \in \mathbb{R}^3$. By fitting this deterministic PDE model to patient data, we can understand the average rates of spread, growth, and maximal tumor volume for the population as well as the level of uncertainty in these parameter estimates. Such an understanding may serve as a first step in developing treatment and dosing strategies as well as guidelines for clinical studies [24]. When certain attributes of patients in the population are known (e.g., age, benign/malignant diagnosis, weight), the NLME framework allows one to further understand how these attributes influence the average behavior or uncertainty in the underlying processes.

We first describe the generic NLME framework in section 4.1 and then review previous applications of this framework to investigate tumor heterogeneity in section 4.2. Then, in section 4.3, we illustrate how to apply this framework to infer the underlying patient proliferation rate distribution from the homogeneous virtual population data described in section 3.3.

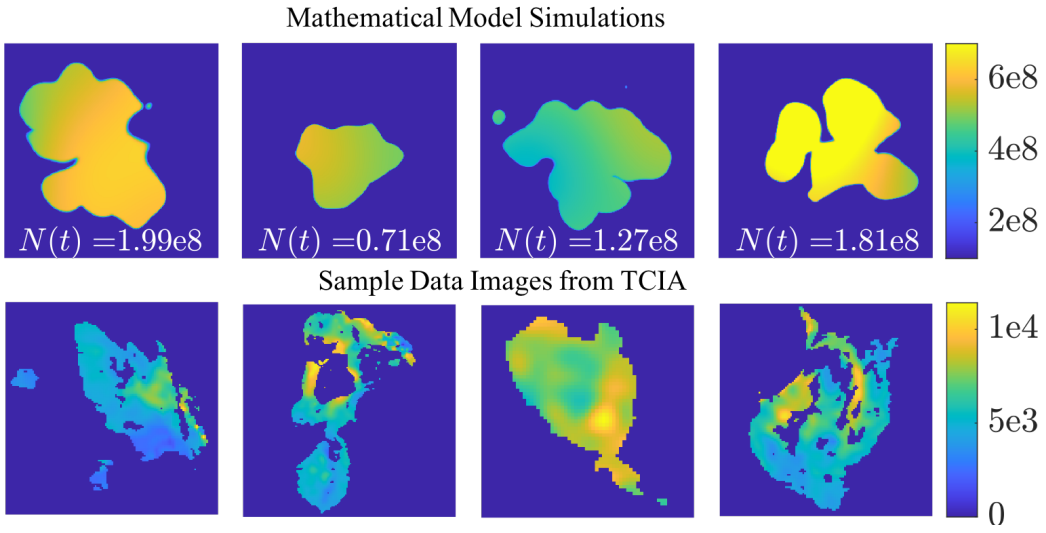

Figure 1. In the top row, we show a virtual population of tumor cell densities $n(\boldsymbol{x}, t)$, at t = 335.2 days. These were generated using Eq (2.1) with lumpy random field coefficients using parameters given in Table 2. We give a common color scale to illustrate heterogeneity within the population. In the bottom row, we show MRI-based Apparent Diffusion Coefficient (ADC) maps for a collection of four human brain tumors gathered from the Brain-Tumor-Progression dataset on The Cancer Imaging Archive [72, 74]. We provide these images for qualitative comparison only; the virtual tumors and real ADC maps are not intended to match.

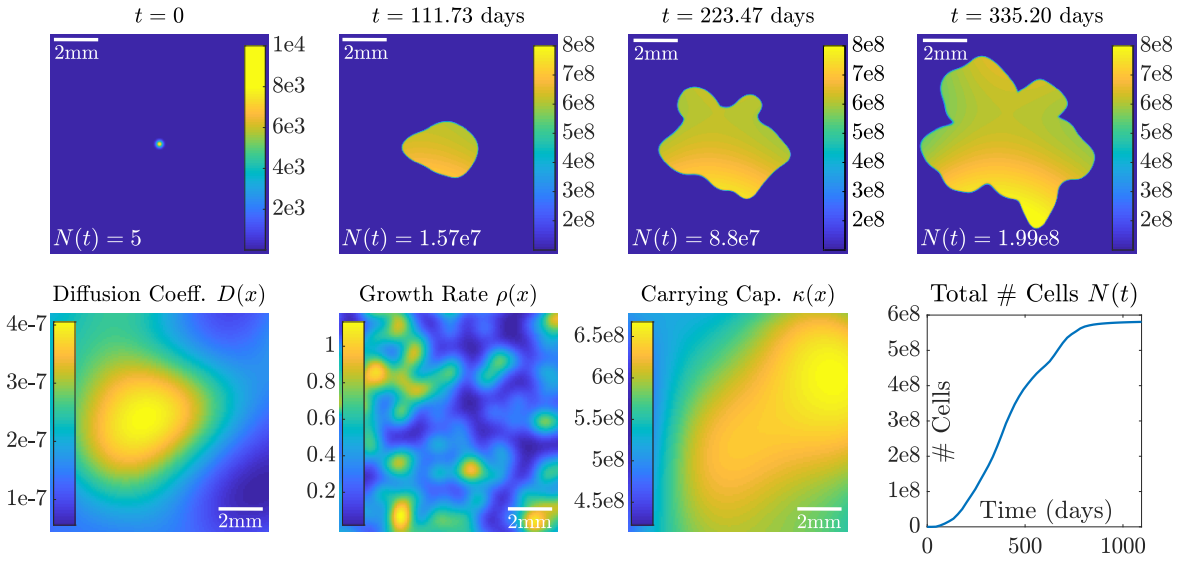

Figure 2. Example time course of a single simulated random heterogeneous tumor, with the random coefficient fields and the resulting $N(t)$ displayed below.

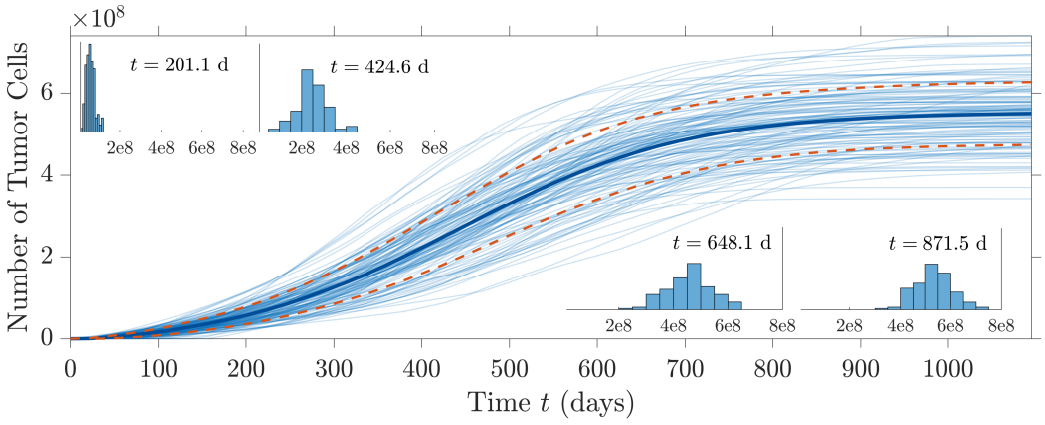

Figure 3. Plot of the predicted total tumor burden $N(t)$ (in number of cells) for each of the $K = 128$ simulated heterogeneous tumors. The estimated mean $\bar{N}(t)$ is shown in bold, while an estimated $\bar{N}(t) \pm \sigma(t)$ confidence band is shown with the dashed lines. Histogram estimates of the time-dependent probability density $p_{N(x,t)}$ are shown for four times.

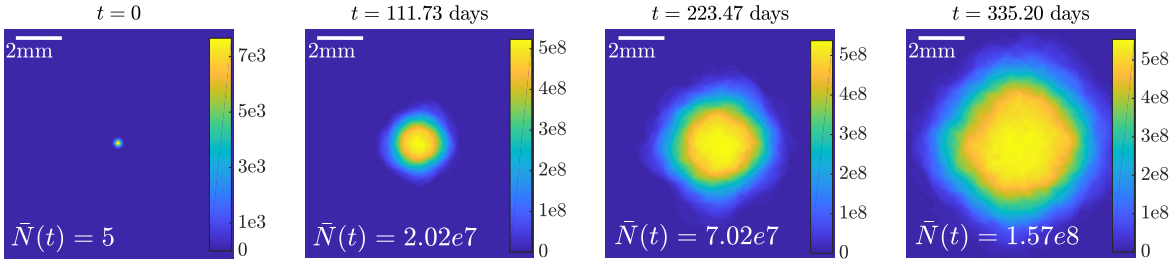

Figure 4. Plot of the simulated sample average tumor cell density $\bar{n}_{ij} = \frac{1}{K}\sum_{k=1}^{K} n_{ij,k}$ for four time points and $K = 128$ samples. See Figure 3 for a plot of the average $\bar{N}(t)$.

4.1. The general NLME framework

We will briefly summarize the main procedure of NLME modeling and refer the reader to [73] for an exhaustive review. The general NLME framework consists of two stages: an individual-level model and a population level model. Let $y_{i,j,k}$ represent the quantity of interest at time t_j and spatial location $\boldsymbol{x}_i$ from the k^{th} patient. The *individual-level model* is then given by

$$y_{i,j,k} = f(\boldsymbol{x}_i, t_j; \boldsymbol{\beta}_k(\boldsymbol{u}_k)) + \epsilon_{i,j,k}, \tag{4.1}$$

where f represents a function governing within-individual behavior, $\boldsymbol{\beta}_k$ represents the $p \times 1$ vector of model parameters specific for patient k (which may depend on the patient's covariates $\boldsymbol{u}_k$, which may include age, weight, diagnosis, etc.), and $\epsilon_{i,j,k}$ denotes realizations from both $\boldsymbol{\epsilon}^{(meas)}$ and $\boldsymbol{\epsilon}^{(model)}$ from Eq (2.9) representing the combination of many potential errors, including random intra-individual uncertainties, natural inter-individual variations, and measurement error). We assume for simplicity that $\mathbb{E}(\epsilon_{i,j,k}|\boldsymbol{\beta}_k) = 0$. For the problem of tumor spread and growth, $y_{i,j,k}$ represents the tumor density at spatio-temporal point $(\boldsymbol{x}_i, t_j)$ for patient k, and f is the solution to the homogenous Fisher-KPP PDE, given by Eq (2.1) with $\boldsymbol{\beta} = (D, \rho, \kappa) \in \mathbb{R}^3$. The vector $\boldsymbol{\beta}_k$ contains the individual patient parameter

values $\boldsymbol{\beta}_k = (D_k, \rho_k, \kappa_k)$.

The *population-level model* describes how variations in the model parameters between individuals are due to both individual attributes and random variations. The general formulation for the population model is given by

$$\boldsymbol{\beta}_k = \boldsymbol{d}(\boldsymbol{u}_k, \boldsymbol{\beta}_k^{(fixed)}, \boldsymbol{\beta}_k^{(rand)}), \tag{4.2}$$

where the function $\boldsymbol{d}$ has a p-dimensional output, $\boldsymbol{u}_k$ denotes the covariate vector for patient* k, $\boldsymbol{\beta}_k^{(fixed)}$ represents *fixed effects* (i.e., nonrandom population parameters), and $\boldsymbol{\beta}_k^{(rand)}$ represents *random effects* (i.e., random parameters). A simple and common example for function $\boldsymbol{d}$ is given by

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^{(fixed)} + \boldsymbol{\beta}_k^{(rand)}, \tag{4.3}$$

in which the parameters representing each patient are assumed centered about the fixed effect $\boldsymbol{\beta}_k^{(fixed)}$ that is common for the whole population with the same covariates, $\boldsymbol{u}_k$, with random variation $\boldsymbol{\beta}_k^{(rand)}$. There are cases where parameter values depend on patients' covariate values. For example, for the problem of tumor spread and growth described earlier, the growth rate ρ for the malignant patients should be higher than that for benign patients. In this case, we can write the growth rate as:

$$\rho_k = \rho_0 + I_k \rho_m, \tag{4.4}$$

where ρ_0 denotes the growth rate for benign tumors and $\boldsymbol{\rho}_m$ denotes the increase in the growth rate for malignant tumors, and I_k is a discrete variable designated as 0 for benign patients and 1 for malignant patients. Consequently, the parameter vector $\boldsymbol{\beta}_k$ has the form:

$$\boldsymbol{\beta}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & I_k & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} D \\ \rho_0 \\ \rho_m \\ \kappa \end{bmatrix} + \boldsymbol{\beta}_k^{(rand)}. \tag{4.5}$$

There are cases where the parameter values for one patient type have less variation than that for other patient type. This leads to the general linear population model provided by

$$\boldsymbol{\beta}_k = \boldsymbol{A}_k \boldsymbol{\beta}_k^{(fixed)} + \boldsymbol{B}_k \boldsymbol{\beta}_k^{(rand)}. \tag{4.6}$$

Here $\boldsymbol{A}_k$ is a design matrix determining how patients' covariates influence model parameters and $\boldsymbol{B}_k$ is a design matrix describing the variation associated with each of these estimates for different groups.

The function f may not capture all within-individual processes, or may exhibit local fluctuations. As such, f may be viewed as an average over all possible realizations for patient k and $\boldsymbol{\beta}_k$ parameterizes these different possible realizations. For example, we observe in Figure 4, that averaged RDE data appears similar to a homogoenous simulation of the Fisher-KPP simulation.

*Note that in the typical NLME framework, $\boldsymbol{x}_i, \boldsymbol{t}_j$, and $\boldsymbol{u}_k$ would be concatenated together into one variable, e.g., $\boldsymbol{X}_{i,j,k} = (\boldsymbol{x}_{i,k}, t_{j,k}, \boldsymbol{u}_k)$, but we have chosen to write them as separate variables in this study to be consistent with typical mathematical modeling literature. If one only has access to time-varying values, then we would have $\boldsymbol{X}_{j,k} = (t_{j,k}, \boldsymbol{u}_k)$.

4.2. Previous literature of NLME for tumor heterogeneity

In tumor biology, it is fundamental to understand both the population-wide behavior driving growth and progression as well as the inherent differences between individuals due to heterogeneity in immune responses and spatial microenvironment variables. This has prompted the common use of nonlinear mixed effects modeling in studies in tumor growth. Previous studies of tumor heterogeneity that use NLME modeling include [75–80]. Here, we review in detail a few studies that use NLME modeling to predict tumor responses to drug treatment [81, 82] and the differences in metastatic spreading between patients [28, 29]. For a broader review on the use of NLME for anticancer drug treatment, the reader is referred to [83].

Ferenci et al. [81] use empirical models to describe the effect of an angiogenesis inhibitor, *bevacizumab*, on final tumor size in mice. Tumor size was measured over time with digital calipers and small animal MR images during the initial growth phase of a xenograft tumor after mice were implanted with colon adenocarcinoma. Patients were divided into two groups: One receiving a large dose on the first and seventeenth day of the experiments, and the other group receiving a smaller daily dose. The considered models include the exponential, logistic, and Gompertz equations. Accordingly, $\boldsymbol{\beta}_k$ contains the parameters for growth rate, carrying capacity, and time scaling for the different models. If $I_k^{(\text{group})}$ denotes the patient groups (two large doses or daily small doses) and $I_k^{(\text{meas})}$ denotes the type of measurement (digital caliper or MRI), then the population model is given by

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}^{(\text{mean})} + \boldsymbol{\beta}_k^{(\text{rand})} + \boldsymbol{\beta}^{(\text{group})} I_k^{(\text{group})} + \boldsymbol{\beta}^{(\text{meas})} I_k^{(\text{meas})} \tag{4.7}$$

where $\boldsymbol{\beta}^{(\text{group})}$ denotes the changes to the mean population response, $\boldsymbol{\beta}^{(\text{mean})}$, for different groups and $\boldsymbol{\beta}^{(\text{meas})}$ denotes the changes for the different measurement sources. Ultimately the authors determined that the exponential model could describe patient data robustly, whereas the sigmoidal models had large amounts of uncertainty in parameter estimates. The exponential model exhibited decreased growth rates for the daily dosing regime as compared to the mean population growth parameter, suggesting that smaller daily doses are more effective in preventing tumor growth than a small number of larger doses.

In [82], a compartmental ODE model was used to study the reduction in size of low-grade glioma (LGG) in response to three types of therapy: Procarbazine and vincistrine (PCV) chemotherapy, temozolomide (TMZ) chemotherapy, and radiotherapy. The dependent variables in the model consist of proliferative cells, quiescent cells, damaged quiescent cells, and concentration of the treatment. The model is fit to data measuring mean tumor diameter (MTD) from MRI scans for 21 PCV patients, 24 TMZ patients, and 25 radiotherapy patients. Each parameter in the compartmental model assumes log-normal random effects. The model for TMZ patients was validated on 96 other TMZ patients, and can accurately predict 5%, 50%, and 95% percentiles of the observed MTD time course using 200 virtual patients from the model. The model was further able to fit individual MTD trajectories based on treatment type and pretreatment MTD. The accuracy of this model in predicting both population-level distributions and individual trajectories suggests it is a viable tool for future use in predicting treatment efficacy for individual patients.

In [29], a transport model was used to describe metastasitic spreading during orthotopic breast tumor xenograft experiments in immunodeficient mice. Tumor size was measured over time with 3D bioluminescence imaging. An ODE model is used to describe tumor growth at the primary tumor site in combination with a size-structured PDE transport model to describe how metastatic cells spread to

other areas of the body. The empirical 10–90% intervals of metastatic mass data for all patients varied over two orders of magnitude at all time points, demonstrating the large inter-individual variation in the data. The authors used log-normally distributed parameter values to match the model to these empirical distribution values, and concluded that the total metastatic burden was a sufficient metric from which to infer the rates of tumor growth and spread. Their use of NLME in this study demonstrated that individual differences in the primary tumor size lead to increased rates of spreading in patients.

Morrell et al. [28] investigate longitudinal levels of prostate-specific antigen (PSA), which is a biomarker for prostate cancer. From a longitudinal study [84], the PSA levels before, during, and after prostate cancer diagnosis were available from 18 men: 11 who had local or regional tumors and 7 who had advanced or metastatic tumors. The authors used a NLME framework to understand if the PSA levels behave different for the two types of patient classifications. To do so, a piecewise linear-exponential model was fit to PSA levels over time, with random effects incorporated into the transition time between the linear and exponential phases as well as the growth rates of the exponential phase. This analysis demonstrated that a significant difference between metastatic and benign diagnoses is the lag between this transition time and time of diagnosis, suggesting that early detection of the tumor is the key to preventing harmful spreading.

4.3. Case study: Applying NLME to infer inter-patient heterogeneity

In this section, we illustrate the use of NLME on the spatially homogeneous virtual dataset generated from Eq (2.1) in section 3.3 with $\boldsymbol{D} = D \in \mathbb{R}, \boldsymbol{\rho} = \rho \in \mathbb{R}, \boldsymbol{\kappa} = \kappa \in \mathbb{R}$. To recap, there are 25 tumor profiles, $\{\boldsymbol{y}_k\}_{k=1,\dots,25}$, where $\boldsymbol{y}_k = n(\boldsymbol{x}, \boldsymbol{t}; \boldsymbol{\beta}_k)$ from Eq (2.6) denotes each tumor's spatiotemporal volume over time. Note that for simplicity in this tutorial we are assuming that $\boldsymbol{y}^{(meas)} = \boldsymbol{y}^{(true)}$. We assume some of the patients have been diagnosed as malignant and the others have been diagnosed as benign: Each patient is given a covariate variable, I_k, which is 0 if they are benign and 1 if they are malignant. For simplicity in this section, we assume D and κ are known: we only need to infer ρ. We let this growth rate take the population model given by

$$\rho = \rho_0 + I_k \rho_M \tag{4.8}$$

where

$$\rho_0 \sim \mathcal{N}(\mu_0, \sigma_0^2), \;\; \rho_M \sim \mathcal{N}(\mu_M, \sigma_M^2). \tag{4.9}$$

The parameter ρ_0, thus corresponds to the benign tumor rate of growth, ρ_M, corresponds to the increase in the malignant rate of growth, and the σ_0^2, σ_M^2 parameters are the variation associated with these values. Note that several of the NLME studies discussed in section 4.2 assumed log-normal distributions for some parameters. Such parameters in these situations are log-transformed during computation so that the parameters being inferred are normally distributed. We thus have the population-level model given by

$$\boldsymbol{\beta_k} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & I_k & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} D \\ \rho_0 \\ \rho_M \\ \kappa \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & I_k \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_0 \\ \sigma_M \end{bmatrix} \tag{4.10}$$

and we aim to infer $\boldsymbol{\theta} = (\rho_0, \rho_M, \sigma_0, \sigma_M)$ from $\{\boldsymbol{y}_k\}_{k=1,\dots,25}$. Recall that this synthetic data set was generated by setting $\mu_0 = 1 \times 10^{-3}, \sigma_0 = 2 \times 10^{-4}, \mu_M = 0.03,$ and $\sigma_M = 1 \times 10^{-2}$.

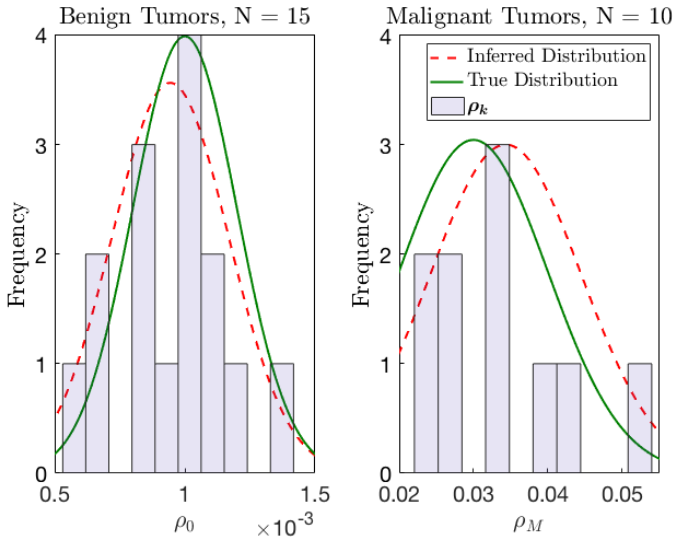

Figure 5. Using the NLME framework to infer the distribution underlying ρ for the homogeneous data set. The solid green curves represent the true underlying distribution for ρ and The red dashed lines represent our inferred distribution for ρ using NLME. The purple histograms correspond to the number of realizations of ρ for the $K = 25$ patients. The left panel corresponds to benign patients and the right panel corresponds to malignant patients.

We performed estimation of this distribution for ρ using **MATLAB**'s **nlmefit** command. This function estimates the underlying distribution by maximizing an approximation to the marginal likelihood assuming that the random effects are normally distributed and observation errors are independent of the random effects and *i.i.d.* normally distributed [85]. We computed this with an absolute tolerance of 1.0.

Our NLME implementation accurately inferred each of the four distribution parameters as: $\hat{\mu}_0 = 9.440 \times 10^{-4}, \hat{\sigma}_0 = 2.239 \times 10^{-4}, \hat{\mu}_M = 3.44 \times 10^{-2}$, and $\hat{\sigma}_M = 1.01 \times 10^{-2}$. The distribution (Eq (4.9)) generated by these estimated values, is depicted against the true underlying data distribution in Figure 5. The inferred distribution closely matches the true underlying distribution; using a two-sample Kolmogorov-Smirnov test between the true and inferred distributions (using 15 benign and 10 malignant samples from both distributions), we fail to reject the null hypothesis (that these two samples are from different distributions) at the 5% significance level.

5. Non-parametric estimation

As mentioned in section 3, the parameters in random differential equations are random variables, i.e., they are associated with a probability distribution. In this section, we describe the techniques for estimating these probability distribution using spatio-temporal aggregate data. The focus is on non-parameteric frameworks that make no assumptions about the form of the probability distribution. This framework has previously been used to estimate the distribution of growth-rates of shrimp populations [86, 87] and to identify the inter-individual heterogeneity in the relationship between

alcohol transdermal sensor measurements and blood alcohol concentration [88]. In the context of cancer modeling, this framework has been used to characterize the heterogeneity within a tumor [89]. In Rutter et al. [89], a reaction-diffusion equation was proposed in which the growth rate, $\boldsymbol{\rho}$, and the diffusion rate $\boldsymbol{D}$ are random variables. Synthetic data were generated with a variety of underlying distributions for the parameter (such as bigaussian to represent a situation in which a portion of the tumor cells grew faster and the remainder grew slower, as well as point distributions to represent the traditional reaction diffusion equation with constant parameters). The authors then used the prohorov metric framework (PMF) to estimate the probability distribution of the random variables. Probability distributions of the parameter were succesfully recovered even with the introduction of 5% noise [89]. Below, we illustrate the application of the PMF for the inverse problem associated with the Fisher-KPP model using the same synthetic data as in section 4.

5.1. The prohorov metric framework

We consider the case of the Fisher-KPP model where only the growth rate $\boldsymbol{\rho}$ is a random variable defined on a probability space, denoted by Ω, and the diffusion rate D and carrying capacity κ are constant. First, we rewrite the random differential equation for the Fisher-KPP model in a more explicit form as follows:

$$\frac{\partial n(x,t,D,\boldsymbol{\rho})}{\partial t} = \nabla \cdot (D\nabla n(x,t,D,\boldsymbol{\rho})) + \boldsymbol{\rho} n(x,t,D,\boldsymbol{\rho})\left(1 - \frac{n(x,t,D,\boldsymbol{\rho})}{\kappa}\right), \tag{5.1}$$

and calculate the average tumor volume, $\bar{n}(x,t)$, aggregated over the individuals by taking the expectation over the probability distributions:

$$\bar{n}(x,t) = \mathbb{E}\left[n(x,t,\cdot,\cdot), P_{\boldsymbol{\rho}}\right] = \int_{\Omega} n(x,t,D,\boldsymbol{\rho}) dP_{\boldsymbol{\rho}}. \tag{5.2}$$

One of the many advantages to this approach is that it is flexible enough to model the regular reaction-diffusion equation (where $\boldsymbol{\rho}$ has an underlying point distribution).

We describe the method here for completeness, but refer the reader to [90, 91] (and the references therein) for proofs regarding the consistency and convergence of the inverse problem. The inverse problem is formulated to find an estimated probability distribution $\hat{\mathbb{P}}_{\rho}$ that minimizes the least squares problem:

$$\hat{\mathbb{P}}_{\boldsymbol{\rho}} = \underset{P_{\boldsymbol{\rho}}^M(\Omega)}{\operatorname{argmin}} \sum_{i,j} \left(y_{ij} - \int_{\Omega} n(x_i, t_j; D, \boldsymbol{\rho}) dP_{\boldsymbol{\rho}}^M \right)^2 \tag{5.3}$$

where y_{ij} represents the data at spatial location i and time j, where $i = 1, .., N_x$ and $j = 1, ..., N_t$, and n represents the numerical approximations to the solution for the partial differential equation, and M represents the number of elements (or nodes) used in the finite-dimensional approximation (denoted $P_{\boldsymbol{\rho}}^M$) of the underlying distribution $\hat{\mathbb{P}}_{\rho}$.

When the underlying distributions are discrete, we use discrete approximations for which delta functions are equispaced over the desired parameter space. For example, if $\boldsymbol{\rho}$ lies in a finite interval, a parameter mesh grid of equispaced M nodes between the end points can be used: $\boldsymbol{\rho}^M = \{\rho_k, k = 1, ..., M\}$. The inverse problem is simplified to:

$$\hat{\mathbb{P}}_{\boldsymbol{\rho}} = \underset{\mathbb{R}}{\operatorname{argmin}} \sum_{j,i} \left[y_{ij} - \left(\sum_{k} n(x_i, t_j; D, \rho_k) w_k \right) \right]^2 \tag{5.4}$$

where w_k are the weights of the respective points. Thus, we aim to determine the weights w_k that minimize the distance between the computed solution and the data. Since the w_k describe a probability distribution, we require that $w_k \geq 0$ and $\sum_{k=1}^{M} w_k = 1$.

When the underlying distributions are continuous, we can use spline approximations composed of hat functions to parameterize a continuous distribution. Similar to the discrete case, we generate a mesh of M nodes for the random variable: $\boldsymbol{\rho}^M = \{\rho_l, l = 1, ..., M\}$. However, in this case the nodes ρ_l are used for constructing hat functions as follows

$$s_l(\rho) = \begin{cases} \frac{\rho - \rho_{l-1}}{\rho_l - \rho_{l-1}} & \text{if } \rho \in [\rho_{l-1}, \rho_l] \\ \frac{\rho_{l+1} - \rho}{\rho_{l+1} - \rho_l} & \text{if } \rho \in [\rho_l, \rho_{l+1}] \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

where $l = 1, ..., M$. Then, the inverse becomes to find the probability distribution such that:

$$\hat{\mathbb{P}}_{\boldsymbol{\rho}} = \underset{\mathbb{R}}{\operatorname{argmin}} \sum_{i,j} \left[y_{ij} - \left(\sum_l a_l \int_{\Omega_\rho} n(x_i, t_j; D, \boldsymbol{\rho}) s_l(\boldsymbol{\rho}) d\rho \right) \right]^2 \tag{5.6}$$

where a_l can be considered as weights of the hat functions. Since $p_l = a_l s_l(\boldsymbol{\rho})$ represents the probability density, it is required that $\sum_{l=1}^{M} a_l \int_{\Omega_\rho} s_l(\boldsymbol{\rho}) d\rho = 1$. The goal is to find values of a_l that minimize the distance between the computed solution and the data.

Generally we do not know whether the underlying distribution is discrete or continuous. In order to decide which method, spline approximations or discrete approximations, and how many nodes to use, we need an unbiased measure. We use the Akaike Information Criteria (AIC) [92] that penalizes for using a higher number of nodes.

5.2. Case study: Applying the PMF to infer inter-patient heterogeneity

In previous work involving subpopulation identification for *intra-tumoral* heterogeneity, the data v_{ji} was the aggregate tumor cell population for a single individual at spatial location j and time i. In our case we aim to understand inter-tumor heterogneity. Instead of identifying subpopulations of cells that grow at different rates, we want to identify subpopulations of individuals that grow at different rates. To accomplish this, we aggregate over the entire population. For example, if individuals 1 through K had a tumor cell population v_{ji}^k, the aggregate population would be $\sum_{k=1}^{K} v_{ji}^k$.

We implemented the PMF for the same 25 synthetic tumors described in section 4, of which 15 tumors are benign and 10 tumors are malignant. We used a discrete approximation to estimate the underlying distribution. In particular, we found that 380 nodes, equispaced between 0 and 0.06, resulted in the lowest AIC scores. The resulting estimation exhibited a clear difference between growth rates for benign tumors (< 0.005) and growth rates for malignant tumors (0.02–0.06). In particular, 60% of the tumors are identified as benign which reflects the true number of benign patients (15 of the 25 synthetic tumors are benign).

Figure 6 displays the 'true' underlying distribution (in purple rectangles) and the best discrete approximation (in red). The lines connecting the discrete approximations are for ease of reading the trends only. The PMF method is able to correctly determine that the underlying distributions is

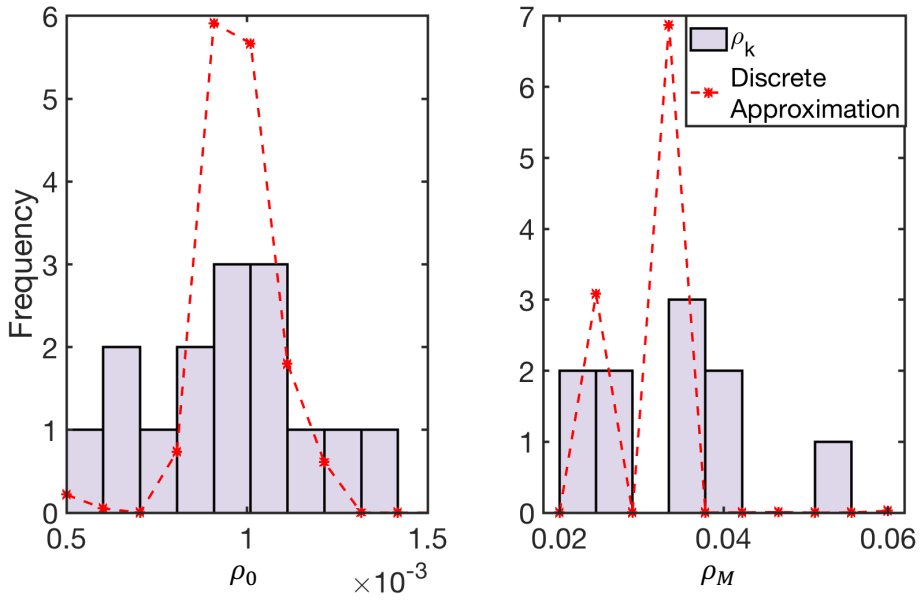

Figure 6. The true (purple) and predicted (red) growth rates for the 25 individuals using a discrete approximation. The left panel corresponds to benign patients and the right panel signifies malignant patients. The true (purple) and predicted (red) growth rates using a discrete approximation for the 15 benign individuals (left panel) and 10 malignant patients (right panel).

bi-modal with a peak in the low growth rates and a second peak in the higher growth rates. The total amount of the predicted distribution lying between the benign and malignant cases is less than 0.09%. Examining the recovered distributions in the benign case (left), we can see that the recovered distribution has slightly over-estimated the peak of the benign growth rates and the recovered distribution is slightly wider than the true underlying distribution. When examining the malignant case, the discrete approximation does a good job in approximating the true underlying distribution. We hypothesize that since the sample rate is sparse (only 10 patients), we do not recover the 'true' underlying normal distribution. Like many numerical schemes, the PMF framework is sensitive to the discrete mesh used in the approximation. In particular, the parameter mesh must be fine enough to fit the benign growth rates, which may result in overfitting of the malignant growth rates.

6. Bayesian methods

As we have seen, tumor growth models and their associated QoI can vary widely in complexity. For example, the 2-dimensional RDE model (2.1) with homogeneous coefficients (D, ρ, κ) together with Eq (2.7) leads to a QoI (2.5) with input and output dimensions $p = 3$ and $q = 1$, respectively (for each fixed time $t = T$). Alternatively, incorporating spatial heterogeneity into each of the RDE parameter fields using a lumpy-type random field model (3.3), described by a fixed lump function $\ell(\boldsymbol{x})$ and a fixed number of lumps L, requires $2L$ scalar parameters (the (x, y) location of each lump) for *each* of the parameter fields, resulting in a total of $p = 6L$ scalar parameters required to specify the fields and simulate the model output (2.6) and subsequent QoI models (2.5). If, instead of a scalar

QoI, we require a prediction of the cell density on a spatial grid or for multiple time points, the output dimension q can be very large indeed. In addition, boundary conditions can influence the output in a nontrivial way, and ad-hoc choices can lead to inaccurate model predictions.

In an ideal situation, one would have access to patient-specific estimates of the coefficient fields $D(\boldsymbol{x}), \rho(\boldsymbol{x}), \kappa(\boldsymbol{x})$ and the initial condition $n_0(\boldsymbol{x})$, from which the forward model (2.6) could be run to assist in prognostic and therapeutic decision making. The growth factor $\rho(\boldsymbol{x})$ can be related directly to glioma cell proliferation rate, and hence estimated *in vivo* using a molecular proliferation marker such as the Positron Emission Tomography (PET) tracer ^{18}F-FLT [93]. Diffusion Tensor Imaging (DTI) may offer a pathway to estimating the cellular diffusion rate $D(\boldsymbol{x})$ [94], but this technique requires further clinical validation. To our knowledge, there is no convenient clinical technique to measure either the initial condition $n_0(\boldsymbol{x})$ nor the carrying capacity $\kappa(\boldsymbol{x})$, and neither ^{18}F-FLT or DTI are currently considered standard clinical practice. Thus, the typical strategy is instead to obtain patient-specific estimates of tumor cell density $\boldsymbol{y}_k^{(meas)} = n^{(meas)}(\boldsymbol{x}_{ik}, t_{jk})$, using models that relate cellular density to more commonly available imaging data such as Diffusion Weighted MRI (DWMRI) [40, 41, 95] or FET-PET [96]. These estimates can then be combined with a parameter estimation step to make either population-level inferences about the distribution of $\boldsymbol{\beta}$, or patient-specific prognostic and therapeutic decisions. Another possibility is to modify the original growth model in such a way that the model parameters correspond more closely to directly measurable physiological processes instead of phenomenological parameters, a technique emphasized in [48].

Alternatively, one might attempt to pursue a more classical approach to statistical model calibration by supposing that a set of *calibration* data of known ground-truth (free from or with minimal measurement error) values of both model inputs $\boldsymbol{\beta}$ and true outputs $\boldsymbol{y}^{(true)}$ is available. Such data can be used to infer both the distribution of $\boldsymbol{\beta}$ and the model discrepancy $\boldsymbol{\epsilon}^{model}$. However, there is typically a paucity of such data with which to calibrate tumor growth models, in particular it is usually challenging to control or form independent estimates of the latent parameter $\boldsymbol{\beta}$ in a context that is useful for clinically relevant patient-specific inference. For instance, it may be possible to construct *in vitro* or small animal experiments to control $\boldsymbol{\beta}$ (for instance by modifying nutrient availability and using synthetic ECM such as Matrigel®), but the resulting inferred distributions may not apply to the clinical context. As discussed above, a wide range of clinical imaging data is typically available (e.g., [74] is an open-access database), but imaging data is always noisy, incomplete, and at best only indirectly related to the model parameter of interest, so classical calibration is nearly impossible with such data. With the issues outlined here and in the paragraph above, how do we proceed? A variety of Bayesian techniques are available for performing model calibration at the population level and patient-specific parameter estimation, as well as providing uncertainty quantification on parameter estimates and model outputs, even in the case when patient data is noisy and incomplete.

Broadly speaking, a Bayesian technique assumes that all parameters are random variables, including population parameters which may ostensibly be non-random. For example, in the RDE model with homogeneous coefficients $\boldsymbol{\beta} = (D, \rho, \kappa)$, we account for population heterogeneity by assuming that $\boldsymbol{\beta}$ is random with probability density $p(\boldsymbol{\beta}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a population parameter such as the population mean $\bar{\boldsymbol{\beta}} = (\bar{D}, \bar{\rho}, \bar{\kappa})$ and covariance matrix $\boldsymbol{K}_{\beta}$. While such population parameters are well-defined as large-sample averages, and hence non-random, a Bayesian technique will assign a probability distribution to them anyway. Probabilistic conditioning on available data – whether it be population-level calibration

data or patient-specific data – then specifies the distribution to be used for subsequent inference. Such a distribution is called a posterior. With the notation and setup outlined in section 3, we assume that $\boldsymbol{\beta} \sim p(\boldsymbol{\beta}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a population parameter and $p(\cdot)$ is a Probability Density Function (PDF). Bayesian techniques are capable of performing inference on $\boldsymbol{\theta}$ (to assess inter-patient heterogeneity), as well as on a patient-specific realization $\boldsymbol{\beta}_k$ (to assess intra-patient heterogeneity and quantify uncertainties). We discuss each case separately, starting with the patient-specific case, since population inference can only proceed with a collection of samples from individual patients.

6.1. The patient-specific Bayesian inference framework

In the context of patient-specific inference, we wish to use available data to make inferences about an individual patient's parameter, which we denote by $\boldsymbol{\beta}_k$ here. As throughout this review, we assume that $\boldsymbol{\beta}_k$ is a realization of the random vector $\boldsymbol{\beta}$, which we assume to have PDF $p(\boldsymbol{\beta}|\boldsymbol{\theta})$. While the true population distribution $p(\boldsymbol{\beta}|\boldsymbol{\theta})$ may be partly or completely unknown, one can assume a particular form for patient k, for instance by selecting a convenient nominal value of $\boldsymbol{\theta}$. The assumed distribution for patient k is called the *prior* and is denoted $p^{(prior)}(\boldsymbol{\beta}_k)$. Ideally, the choice of prior is informed by calibration data and not simply convenience; non-informative priors such as uniform and maximum entropy distributions are an alternate choice if such data are unavailable [45, 47].

The next component of a patient-specific Bayesian procedure is an explicit measurement model relating $\boldsymbol{y}_k^{(meas)}$ to $\boldsymbol{y}_k^{(true)}$. Such models typically take the form

$$\boldsymbol{y}_k^{(meas)} = \boldsymbol{\mathcal{H}}\left(\boldsymbol{y}_k^{(true)}\right) + \boldsymbol{\epsilon}_k^{(meas)}, \tag{6.1}$$

where $\boldsymbol{y}_k^{(meas)} \in \mathbb{R}^m$ is the measured data for patient k, and $\boldsymbol{\mathcal{H}} : \mathbb{R}^q \to \mathbb{R}^m$ is the measurement operator. For imaging data, m is typically quite large ($O(10^6 - 10^8)$, especially for volumetric images). For a well-calibrated measurement model, it is usually reasonable to assume that $\mathbb{E}\left[\boldsymbol{y}_k^{(meas)}|\boldsymbol{y}_k^{(true)}\right] = \boldsymbol{\mathcal{H}}(\boldsymbol{y}_k^{(true)})$ – that is, there is no systematic model bias for $\boldsymbol{\mathcal{H}}$. Furthermore, the complete statistics of $\boldsymbol{y}_k^{(meas)}|\boldsymbol{y}_k^{(true)}$ are typically well understood. For instance, most photon-based imaging systems obey Poisson or Gaussian statistics very closely, meaning that $\boldsymbol{y}_k^{(meas)}|\boldsymbol{y}_k^{(true)}$ is accurately described as being either Poisson with mean $\boldsymbol{\mathcal{H}}(\boldsymbol{y}_k^{(true)})$ or Gaussian with mean $\boldsymbol{\mathcal{H}}(\boldsymbol{y}_k^{(true)})$ and covariance matrix $\boldsymbol{K}^{(meas)}$ [67]. The probability density of $\boldsymbol{y}_k^{(meas)}|\boldsymbol{y}_k^{(true)}$ is denoted $p^{(meas)}(\boldsymbol{y}^{(meas)}|\boldsymbol{y}^{(true)})$; as a function of $\boldsymbol{y}^{(true)}$, this density is called the measurement *likelihood* and is written $\mathcal{L}^{(meas)}(\boldsymbol{y}^{(true)}|\boldsymbol{y}^{(meas)}) = p^{(meas)}(\boldsymbol{y}^{(meas)}|\boldsymbol{y}^{(true)})$. Solving the statistical inverse problem (6.1) is considered standard practice, with both frequentest and Bayesian techniques being available [67, 97, 98]. In both cases, one has a choice to produce either a point estimate $\hat{\boldsymbol{y}}_k^{(true)}$, or a full Bayesian posterior $p(\boldsymbol{y}_k^{(true)}|\boldsymbol{y}_k^{(meas)})$. In the latter case, the prior and likelihood are combined to write

$$\boldsymbol{y}_k^{(true)}|\boldsymbol{y}_k^{(meas)} \sim p(\boldsymbol{y}_k^{(true)}|\boldsymbol{y}_k^{(meas)}) = \frac{\mathcal{L}(\boldsymbol{y}_k^{(true)}|\boldsymbol{y}_k^{(meas)})p^{(prior)}(\boldsymbol{y}_k^{(true)})}{p^{(data)}(\boldsymbol{y}_k^{(meas)})} \tag{6.2}$$

However, as discussed throughout this review, we may be more interested in the latent model parameter $\boldsymbol{\beta}$ than we are in the state $\boldsymbol{y}_k^{(true)}$; as highlighted in [23, 24, 99] and elsewhere, the parameters $\boldsymbol{D}$ and $\boldsymbol{\rho}$ in the Fisher-KPP model correlate with prognostic outcome more so than cell density $n(\boldsymbol{x}, t)$.

Combining (6.1) with the model (2.6), we have

$$\boldsymbol{y}_k^{(meas)} = \boldsymbol{\mathcal{H}}\left(\boldsymbol{\mathcal{M}}(\boldsymbol{\beta}_k) + \boldsymbol{\epsilon}_k^{(model)}\right) + \boldsymbol{\epsilon}_k^{(meas)}. \tag{6.3}$$

If $\mathcal{H}$ is linear (a reasonable assumption for many imaging systems, for instance), this reduces to

$$\boldsymbol{y}_k^{(meas)} = \mathcal{H}\mathcal{M}(\boldsymbol{\beta}_k) + \mathcal{H}\boldsymbol{\epsilon}_k^{(model)} + \boldsymbol{\epsilon}_k^{(meas)}. \tag{6.4}$$

Equation (6.4) highlights a fundamental issue, namely that it is difficult to deconvolve the influence of model discrepancy and measurement error: The measured data may differ from the model predicted output both due to model error and measurement noise, and in the absence of noise-free data, it can be difficult to calibrate $\boldsymbol{\epsilon}^{(model)}$ [100]. This leads many to assume (possibly erroneously) that either $\boldsymbol{\epsilon}_k^{(model)} = 0$, or that it lies in the null space of $\mathcal{H}$, so that the nonlinear statistical inverse problem

$$\boldsymbol{y}_k^{(meas)} = \mathcal{H}\mathcal{M}(\boldsymbol{\beta}_k) + \boldsymbol{\epsilon}_k^{(meas)} \tag{6.5}$$

can be solved in a manner similar to (6.1) to obtain either a point estimate $\hat{\boldsymbol{\beta}}_k$ or Bayesian posterior $\boldsymbol{\beta}_k|\boldsymbol{y}_k^{(meas)}$. Denoting the composition of the measurement and model by $\mathcal{F} = \mathcal{H} \circ \mathcal{M} : \mathbb{R}^p \to \mathbb{R}^m$, we write (6.5) as

$$\boldsymbol{y}_k^{(meas)} = \mathcal{F}(\boldsymbol{\beta}_k) + \boldsymbol{\epsilon}_k^{(meas)} \tag{6.6}$$

A Bayesian solution to (6.6) will be similar to (6.2), except that we are inferring $\boldsymbol{\beta}_k$ instead of $\boldsymbol{y}_k^{(true)}$:

$$\boldsymbol{\beta}_k|\boldsymbol{y}_k^{(meas)} \sim p(\boldsymbol{\beta}_k|\boldsymbol{y}_k^{(meas)}) = \frac{\mathcal{L}(\boldsymbol{\beta}_k|\boldsymbol{y}_k^{(meas)})p^{(prior)}(\boldsymbol{\beta}_k)}{p^{(data)}(\boldsymbol{y}_k^{(meas)})} \tag{6.7}$$

In general, drawing samples from the posterior (6.7) requires the usage of a Markov Chain Monte Carlo method such as the Metropolis-Hastings algorithm or one of its more sophisticated variants [47,97]. If one can assume that $\boldsymbol{\epsilon}_k^{(meas)}$ is conditionally independent of $\boldsymbol{y}_k^{(meas)}$, then the likelihood in (6.7) can be written as

$$\mathcal{L}(\boldsymbol{\beta}_k|\boldsymbol{y}_k^{(meas)}) = p_{\boldsymbol{\epsilon}^{(meas)}}(\boldsymbol{y}_k^{(meas)} - \mathcal{F}(\boldsymbol{\beta}_k)) \tag{6.8}$$

For example, if the measurement noise is Gaussian, we would have

$$\mathcal{L}(\boldsymbol{\beta}_k|\boldsymbol{y}_k^{(meas)}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{y}_k^{(meas)} - \mathcal{F}(\boldsymbol{\beta}_k))^\intercal \boldsymbol{K}_{\boldsymbol{\epsilon}^{(meas)}}^{-1}(\boldsymbol{y}_k^{(meas)} - \mathcal{F}(\boldsymbol{\beta}_k))\right) \tag{6.9}$$

6.2. Previous literature of patient-specific Bayesian inference for tumor heterogeneity

Several groups have pursued the usage of Bayesian strategies for patient-specific model calibration in the context of mathematical oncology. For a tutorial review on this topic in the context of an ODE tumor growth model, see [101]. We provide a brief review of several approaches here.

In [102], a reaction-diffusion system that couples proliferating and migrating cells to an extracellular matrix is combined with a measurement model similar to (6.5) to formulate patient-specific tumor growth prediction as a data assimilation problem. In data assimilation, measured data $\boldsymbol{y}_{i,k}^{(meas)}$, collected for patient k at a series of time points $t = t_i$, is used to sequentially update the estimate of the state $\boldsymbol{y}_{i,k}^{(true)}$. The authors employ a technique called the Local Ensemble Transform Kalman Filter (LETKF), which employs an ensemble (i.e., sample) of state estimates to emulate the posterior distribution $p(\boldsymbol{y}^{(true)}|\boldsymbol{y}^{(meas)})$. They do not attempt to perform the simultaneous parameter estimation

problem (i.e., solving the statistical inverse problem (6.5)), instead selecting nominal values for $\boldsymbol{\beta}$ with which to perform the forward prediction $\boldsymbol{y}^{(model)} = \boldsymbol{\mathcal{M}}(\boldsymbol{\beta})$. Their approach is thus more akin to a data assimilation solution to (6.1) in the case when $\boldsymbol{y}^{(meas)}$ is time-resolved.

In [103], the authors discuss an application of the Bayesian calibration and validation framework presented in [104] to a class of diffuse interface/continuum mixture models of tumor growth, and discuss methods of evaluating model fit and selecting the 'best fit' model by comparing quantity-of-interest PDFs from calibration and validation scenarios. The QoI they aim to predict is the same as our Eq (2.7), namely the total tumor burden at some future time $t = t_3$, which we write as $y^{(QoI)}$. Using calibration data from $t = t_1$ and validation data from $t = t_2$, they form two posterior PDFs (in our notation) $p^{(calib)} = p^{(post)}(y^{(QoI)}|\boldsymbol{y}^{(meas)}(t_1))$ and $p^{(valid)} = p(y^{(QoI)}|\boldsymbol{y}^{(meas)}(t_2))$. These posterior distributions are then compared using a metric between probability distributions such as total variation to evaluate goodness of fit, the idea being that the posterior formed with the calibration data should not be significantly different than the posterior formed with the validation data. They compare three models to each other, concluding that a model that incorporates only simple proliferation is rejected in that it cannot reproduce the QoI data generated by a more sophisticated model that incorporates both proliferation and apoptosis as well as oxygen dependence; however, the model with proliferation, apoptosis and oxygen dependence *cannot* be rejected as a surrogate for a model which incorporates time-dependent parameters. While these conclusions are fully *in silico* and hence do not have direct clinical significance, they demonstrate Bayesian calibration methodologies in the context of tumor growth modeling. A later article by the same group provides an extensive review of the application of Bayesian methodologies to multiscale tumor growth modeling [105].

In [106], the authors assume a Fisher-KPP tumor growth model with homogenous diffusion and growth parameters $D(\boldsymbol{x}) \equiv D$ and $\rho(\boldsymbol{x}) \equiv \rho$, and $\kappa \equiv 1$, then use a combination of T1Gd (T1-weighted Gadolinium contrast enhanced) and T2-FLAIR (T2-weighted FLuid-Attenuated-Inversion-Recovery) MRI images to form a Bayesian posterior for the parameters (D, ρ). They choose a likelihood model based on the Hausdorff distance between segmented MRI images and the model predicted tumor, use a variety of priors including a log-uniform, and employ a variant of Hamiltonian Monte Carlo (HMC) to perform the posterior sampling. They do not account for uncertainty in the image segmentation itself, nor do they appear to account for spatial inhomogeneity in either model parameter.

The recent article [107] assesses uncertainty and robustness of a treatment response metric called Days Gained (DG), which is a QoI derived from a Fisher-KPP model by comparing a patient's actual post-treatment tumor size to a virtual untreated control, simulated using the RDE model. The evaluated robustness and uncertainty in DG by accounting for uncertainty in both the segmentation and in the time of tumor initiation, and also evaluate the effect of prior selection, comparing a Gaussian prior calibrated to a real population of glioblastoma patients to an ad-hoc uniform prior. Their results appear to indicate that DG is robust to both the uncertainties assessed as well as the choice of prior.

6.3. The population-level Bayesian inference framework

While the Bayesian techniques discussed in section 6.1 apply to the case of patient-specific uncertainty analysis, it is also useful to apply Bayesian strategies to the population inference case. As we have discussed throughout this review, a heterogeneous population is described by a parameter distribution $\mathbb{P}_{\boldsymbol{\beta}}$, which may depend on a population parameter $\boldsymbol{\theta}$. In 6.1, Bayesian analysis is applied to $\boldsymbol{\beta}$; a population Bayesian analysis would apply to $\boldsymbol{\theta}$.

Assuming a prior PDF $p^{(prior)}(\boldsymbol{\theta})$ and a collection of calibration data $\boldsymbol{Y}^{(calib)} = (\boldsymbol{y}_1^{(meas)}, \ldots, \boldsymbol{y}_K^{(meas)}) \in (\mathbb{R}^q)^K$, a Bayesian method for $\boldsymbol{\theta}$ would produce a posterior PDF $p^{calib}(\boldsymbol{\theta}|\boldsymbol{Y}^{calib})$ for the population parameter $\boldsymbol{\theta}$ by applying Bayes rule:

$$p^{(calib)}\left(\boldsymbol{\theta}|\boldsymbol{Y}^{(calib)}\right) = \frac{p^{(meas)}\left(\boldsymbol{Y}^{(calib)}|\boldsymbol{\theta}\right) p^{(prior)}(\boldsymbol{\theta})}{p^{(data)}\left(\boldsymbol{Y}^{(calib)}\right)} \tag{6.10}$$

where $p^{(data)}(Y^{(calib)}) = \int p^{(meas)}\left(\boldsymbol{Y}^{(calib)}|\boldsymbol{\theta}\right) p^{(prior)}(\boldsymbol{\theta})\, d\theta$.

6.4. Sensitivity analysis and model reduction framework

When there are many parameters to calibrate, a first step is often to reduce the dimensionality of the problem by performing a sensitivity analysis. Heuristically, sensitivity analysis tries to identify those input parameters that most influence the output values. For example, in the RDE model (2.1), of the original p parameters, it may be that only a few play a significant role in the output QoI. Sensitivity analysis techniques are usually categorized as being *local* or *global*, with the former studying the effect of parameter variation around a fixed, nominal value, and the latter evaluating the effect of parameter variation across a wide range of possible values. The simplest form of local sensitivity analysis, at least conceptually, is to compute a finite difference approximation of the partial derivatives

$$\frac{\partial \mathcal{M}_i}{\partial \beta_j} \approx \frac{1}{\delta_{ij}}\left(\mathcal{M}_i(\boldsymbol{\beta} + \delta_{ij}\boldsymbol{e}_j) - \mathcal{M}_i(\boldsymbol{\beta})\right). \tag{6.11}$$

Other techniques for local sensitivity analysis include forming and solving the forward and adjoint sensitivity equations (the 'FSAP' and 'ASAP' methods), or using automatic differentiation [45, 47].

A common approach to global sensitivity analysis is a technique called Sobol analysis [45, 47, 108]. If the effect of the parameters on the output are all statistically independent, Sobol indices measure the relative influence of each of the parameters $\beta_1, \beta_2, ..., \beta_p$. In essence, the method decomposes the total change in the output into a sum of the changes due to each individual β_i parameters, plus the changes due to every pair of parameters (β_i, β_j), plus every triple of parameters, and so on. The first order Sobol indices yield results similar to the classical Analysis of Variance (ANOVA) technique in statistics. Many times, however, the parameters are not independent, and in such a case, the Sobol approach can give misleading results.

A different methodology is called the active subspace model [109]. In essence, active subspaces approximate the gradient of the outputs $\boldsymbol{y}^{(model)}$ with respect to the inputs $\boldsymbol{\beta}$, leading to the matrix

$$\hat{\boldsymbol{C}} = (\nabla \boldsymbol{\mathcal{M}}(\boldsymbol{\beta}))\,(\nabla \boldsymbol{\mathcal{M}}(\boldsymbol{\beta}))^T\,. \tag{6.12}$$

The k largest eigenvalues of $\hat{\boldsymbol{C}}$, and the associated eigenvectors, define the most influential set of parameters. Often one chooses k large enough to capture some threshold of the total variability—say 90%—the premise being that $k \ll p$. Active subspaces generalizes the more classical technique of Principal Components Analysis (PCA), which finds linear subspaces in which the majority of variability resides. By contrast, active subspaces can produce a nonlinear manifold which contains most of the variability.

A simpler but related approach is to use Morris indices [47, 110]. Assuming the input parameters have been scaled to the unit hypercube $\boldsymbol{\beta} \in [0, 1]^p$, one selects a point at random to start and computes

the output value for this input. One then moves a small step (typically 5–10%) in the direction of one of the input dimensions and re-computes outputs. One then proceeds dimension-by-dimension until all dimensions have been sampled, keeping track of the outputs obtained. One can (approximately) integrate along this path to obtain a value of the total variability of the outputs.

Having identified the important input parameters, perhaps re-parameterizing the problem in the process, one can then explore the reduced parameter space in greater detail. A common way to proceed is as follows. First, in the less important parameter dimensions, set the parameter values to some nominal accepted values; second, one can apply a design of experiment approach that samples widely in the remaining parameter dimensions, and lastly, one can employ MCMC methods to calibrate the distribution of the important inputs that best represent any available data. Since MCMC-based calibration can be computationally demanding, it is often helpful to construct a statistical emulator, that is, a surrogate model that very quickly estimates the simulation output at untested input values and, importantly, provides an estimate of the error in this estimation procedure. Gaussian Stochastic Processes provide a robust method for emulator construction [111, 112].

As part of the calibration process, it is important to check whether the MCMC has overfit the parameters. That is, sometimes the posterior distribution of the parameter inputs are concentrated, but at an incorrect value. Such a finding would suggest an overly high degree of confidence in the parameter estimate, even though that parameter distribution does not cover the "correct'' value. The most common cause of such over-fitting is model discrepancy—that is, the model does not include all the features that nature is using. This discrepancy is the source of $\boldsymbol{\epsilon}^{(model)}$ in Eq (2.8). Although the calibration and model error cannot be distinguished independently, one can estimate the size of the error across likely input values, to estimate the extent of the model error [113]. Typically, the model error $\boldsymbol{\epsilon}^{(model)}$ is modeled as a Gaussian Process.

6.5. Case study: Patient-specific Bayesian inference

In Figure 7, we demonstrate a basic application of (6.7) in the case of a linear measurement operator $\mathcal{H}$, making the homogeneous parameter assumption that $\boldsymbol{\beta} = (\boldsymbol{D}, \boldsymbol{\rho}, \boldsymbol{\kappa})$ are spatially constant.

We simulate imaging data by assuming the measurement operator $\mathcal{H}$ from (6.1) is a linear blurring operator, that is, the imaging data consists of noisy linear functionals of the form $y_m^{meas} = (\boldsymbol{h} * \boldsymbol{n})(\boldsymbol{x}_m, t_m) + \epsilon_m^{meas}$, with $1 \leq m \leq M$. In this model, $\boldsymbol{h}$ is a Gaussian point spread (blurring) function, $*$ denotes convolution, $(\boldsymbol{x}_m, t_m)$ are the locations and times of the imaging sensors, and $\boldsymbol{\epsilon}^{meas}$ is a mean-zero Gaussian noise vector. We assume that $M = 2 \times 47^2$, i.e., two 47×47 pixel images are taken at two separate time points, namely $t = 40.6$ and $t = 81.1$ days; see Figure 8. This image data model is a reasonable approximation to many light-based microscope measurements, as well as some reconstructed tomographic images. More detailed imaging system models are discussed elsewhere [67] and are not the purpose of this review. For the experiment shown in Figure 7, the prior $p^{prior}(\boldsymbol{\beta})$ is assumed to be uniform on the set $(D, \rho) \in [0, 1e-4] \times [0, 1]$, while κ is assumed fixed at some nominal value. We employ a standard Metropolis-Hastings MCMC method to draw samples from $p^{post}(D, \rho)$, taking the maximum likelihood estimate of (D, ρ) as the initial point and a Gaussian proposal density with standard deviation $(\sigma_D, \sigma_\rho) = (1e-11, 1e-7)$, leading to an acceptance rate of around 50%. Note that in Figure 7, a marked concentration effect is taking place whereby the accepted values of (D, ρ) seem to concentrate around a 1D submanifold of the parameter space, indicating that the two images available are not sufficient to fully identify the parameter (D, ρ). If

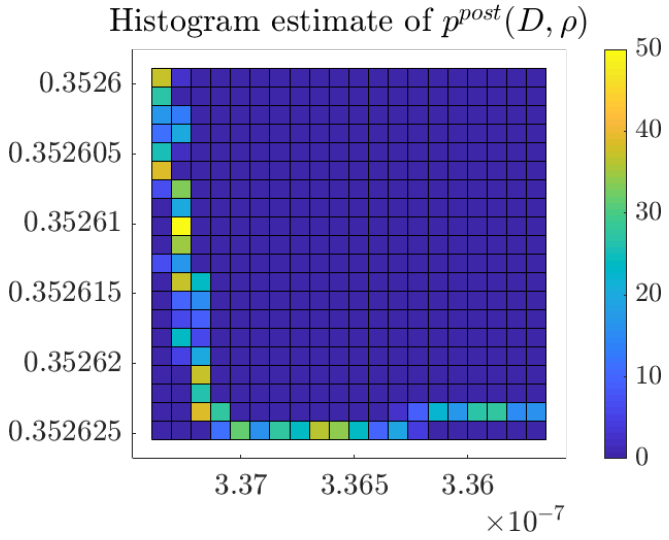

Figure 7. Histogram estimate of the posterior distribution on (D, ρ) in the homogeneous parameter case. Note that the posterior samples seem to concentrate on a 1D subset of the parameter space, demonstrating a possible identifiability issue for (D, ρ) from $n(\boldsymbol{x}, t)$ measurements alone.

additional imaging measurements were available, for instance molecular imaging of cell proliferation as discussed in the introduction to this section, this identifiability issue might be mitigated; see [70] for more information.

7. Machine learning

Machine learning has recently emerged as a powerful tool to automate tedious tasks such as image classification and segmentation [114–116]. Many of the concepts used in machine learning practice can be applied to personalized medicine—for a recent review on machine learning applications in healthcare, see [117]. One advantage of using machine learning methods is the ability to find patterns in data without requiring human input, i.e., feature engineering. However, this advantage comes with some drawbacks, such as lack of interpretability and generalizability.

Recent applications of machine learning to cancer research include data processing such as image processing, analysis of genomic data, and quantification of electronic health records. For example, image processing tasks encompass lesion detection [118, 119], cell segmentation [116, 120], and tumor segmentation [121, 122] which feature heterogeneity in image modality (phase contrast, MR, CAT, etc.), machine type, and across patients. For a recent review on deep learning methods for personalizing medicine via imaging, see [123] and the references therein. Recent work includes survival prediction given imaging data and demographic information [124–126] in a very heterogeneous patient pool. Genomic expression data has been used to predict tumor type [127] and predict survival time in conjunction with histological imaging in gliomas [128]. For a tutorial and review on deep learning models for genomic data, see [129]. Machine learning has also been applied to translating electronic health records (EHR) [130] for use in predicting sepsis [131] and diabetes [132]. Within the context of cancer, machine learning can be used to process EHRs to detect

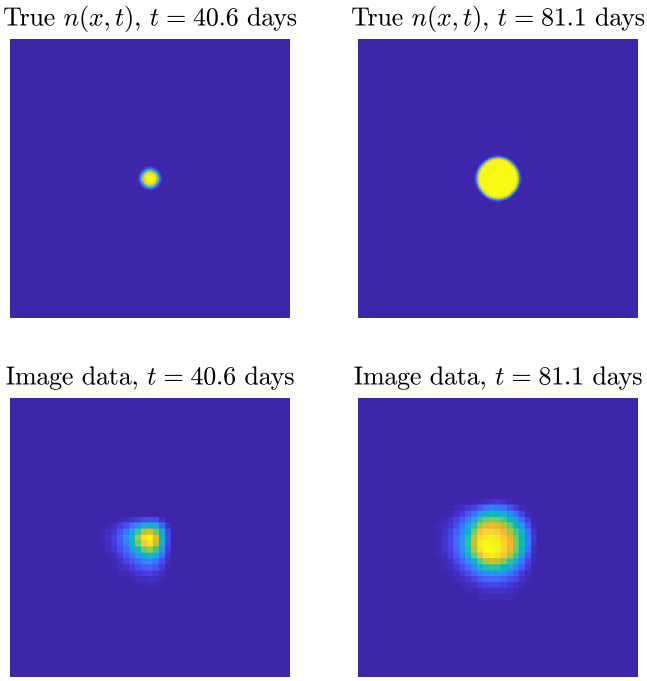

Figure 8. Simulated image data used in the patient-specific Bayesian demonstration. We assume that direct imaging of $n(x,t)$ is available in the form of $\boldsymbol{y}^{meas} = \mathcal{H}\boldsymbol{n} + \boldsymbol{\epsilon}^{meas}$ with $\mathcal{H}$ being a linear shift-invariant imaging operator and $\boldsymbol{\epsilon}^{meas}$ being a Gaussian noise. Each image is 47×47 pixels.

colorectal cancer [133] and to predict pancreatic cancer using EHR data in conjunction with PubMed keywords [134].

7.1. Previous literature of machine learning methods for tumor heterogeneity

Multi-task learning (MTL) [135], a machine learning methodology that considers learning more than one task simultaneously (e. g., predicting both the type and color for a single image of a flower) provides an opportunity to account for patient heterogeneity in prediction. By training tasks in parallel, this technique enables shared feature learning through joint modeling of multiple tasks and has also been shown to help with model regularization [135]. For a recent overview on multi-task learning approaches, see [136] and the references therein. Recent work proposed a MTL architecture in which predicting outcomes for separate individuals are considered as multiple tasks [137]. The objective of this work was to predict the health, stress, and happiness of an individual from longitudinal time-series data of a set of 343 features extracted from sensors, smartphones logs, behavioral surveys, and weather information. The authors discovered that higher prediction accuracy was achieved when a multi-task neural network model was used, in essence treating each prediction for each individual as a separate task. The key feature in this multi-task approach is that the task separation occurs by using a single neural network architecture, in which the first few layers in the network are shared among all individuals, and the last last layers branch out to predict each individual's outcome separately. By defining a separate prediction task for an individual or "user", in this way, population level information is leveraged and population features are "shared" among

individuals. Notably, it was shown that the multi-task model with users-as-tasks was significantly more accurate than the typical single-task modeling approach in which a separate neural network model is used for each individual. This multi-task learning approach could be applied to create individualized predictions of cancer growth in similar settings for which too few data points exist to train an accurate machine learning model for each patient, but data exists for a large population of patients.

More recent work has proposed combining mathematical modeling and machine learning to enhance accuracy and interpretability. The 2019 Mathematical Onocology Roadmap [138] urged the community to begin examining the ways machine learning and mathematical modeling can be integrated. For example, physics-informed neural networks (PINNs) [139] were shown to help constrain the neural networks that are used for learning equations [140, 141] to obey the laws of physics. Other examples have shown that taking a hybrid approach to prediction by combining characteristics of both mechanistic modeling and data-driven modeling can offer improvements in prediction accuracy and parameter estimation over the standalone methodologies [142, 143]. Within the context of GBM, recent work showed that constraining machine learning models with the proliferation-invasion model (Eq (2.1)) resulted in more accurate predictions of tumor cell density than mathematical models or machine learning models alone [144].

In the remainder of this section, we focus on an example machine learning methodology for generating virtual data that are similar to the real/original data: Generative networks (VAEs, GANs, etc.). Generative modeling in machine learning attempts to produce new samples from some unknown probability distribution $\mathbb{P}_{\text{data}}$ given a set of observations from that distribution. In other words, given input data ($\boldsymbol{y} \in \mathbb{R}^{\Omega}$), a model can be trained to generate data like it ($\hat{\boldsymbol{y}} \in \mathbb{R}^{\Omega}$). Thus, generative modeling, similarly to virtual populations, can provide a platform to investigate heterogeneity. Note that while conceptually similar to the process of generating virtual population data in section 3 where patient-specific virtual populations are generated based on the patient-specific distribution of parameters (derived from patient-specific data), generative models in machine learning are trained based on the patient-specific distribution of the data itself (e.g., images of tumor density). Additionally, these methods frequently assist in other ways such as increasing the amount of available training data and discovering a data set's salient features.

In most recent applications, generative models typically adopt one of two neural network approaches: Variational auto-encoders (VAEs) or generative adversarial networks (GANs). In the following subsections we provide high level overviews of each method as well as examples where they have been utilized in precision medicine. In section 7.2 and 7.3, we show examples of generative model predictions based on the virtual population data described in section 3.3 of this paper and provide open-source code for these examples. We conclude by applying these methods to the virtual population data from section 3, providing open-source code in Python 3.7 using the PyTorch deep learning library.

7.2. Case study: Variational auto-encoders for virtual population generation

Variational auto-encoders [145] are an extension of auto-encoders which specialize in compressing and decompressing data. Auto-encoders (AEs) are an unsupervised learning method that leverage neural networks to learn lower-dimensional representations of data without the need for feature engineering. Traditionally, AEs are composed of two distinct neural networks, known as an encoder

and a decoder. The encoder neural network uses input data $\boldsymbol{y} \in \mathbb{R}^{\Omega}$ (e.g., images where Ω is large) and through a series of feature extraction and down-sampling layers "compresses" each input into an ℓ-dimensional vector of latent variables $\boldsymbol{z} \in \mathbb{R}^{\ell}$ where $\ell \ll \Omega$. The decoder network uses $\boldsymbol{z}$ as an input and, through a series of feature extraction and up-sampling layers, "decompresses" the latent vector into an output $\hat{\boldsymbol{y}} \in \mathbb{R}^{\Omega}$ which is a reconstruction the original input data $\boldsymbol{y}$. The low dimensional bottleneck between the input data and their reconstructions forces the AE to learn a maximally informative compressed representation of the input data. AEs are trained end-to-end using the reconstruction error (typically mean squared error) between inputs $\boldsymbol{y}$ and reconstructions $\hat{\boldsymbol{y}}$. Concretely, given an encoder $N_E(\boldsymbol{y}; \theta_E) : \mathbb{R}^{\Omega} \to \mathbb{R}^{\ell}$ parameterized by θ_E and decoder $N_D(\boldsymbol{z}; \theta_D) : \mathbb{R}^{\ell} \to \mathbb{R}^{\Omega}$ parameterized by θ_D, the objective of auto-encoders is to minimize the cost function

$$\mathcal{J}^{\mathrm{AE}}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 \tag{7.1}$$

where $\hat{\boldsymbol{y}} = N_D(\boldsymbol{z}) = N_D(N_E(\boldsymbol{y}))$. We note that the parameters θ described in this section refer to the weights and biases of neural network models and do not describe the population parameters from section 3. AEs in this form, however, are not generative since for a fixed input, the reconstructed output is always the same.

VAEs extend auto-encoders to the class of generative models. Similarly to AEs, VAEs are also composed of encoder and decoder networks. However, rather than having the encoder output the latent vector $\boldsymbol{z}$ directly, the encoder network in a VAE instead outputs parameters which characterize a probability distribution (typically a normal distribution) from which realizations of $\boldsymbol{z}$ can be sampled. In this way, VAEs encode input data into latent *distributions* rather than latent *samples*. For example, if we wish to encode input data $\boldsymbol{y}$ into an ℓ-dimensional normal distribution from which we can easily draw samples, we can modify the encoder neural network to output two vectors $\mu \in \mathbb{R}^{\ell}$ and $\sigma \in \mathbb{R}^{\ell}$. Then, the decoder network uses as inputs a random sample $\boldsymbol{z} \sim \mathcal{N}(\mu, \Sigma)$, where Σ is the diagonal covariance matrix with diagonal elements $\sigma_1, \ldots, \sigma_{\ell}$, and outputs a reconstruction $\hat{\boldsymbol{y}}$ of the original input $\boldsymbol{y}$. In this formulation, the ℓ-dimensional normal distribution, characterized by μ and σ, is composed of ℓ independent 1-D normal distributions.

Similar to auto-encoders, variational auto-encoders are also trained end-to-end by minimizing the reconstruction error between inputs $\boldsymbol{y}$ and reconstructions $\hat{\boldsymbol{y}}$. However, in addition to accurate reconstruction quality, the optimization also accounts for similarity in latent representations. Therefore, an additional constraint is added to the VAE objective function, specifically on the latent probability distribution characterized by encoder outputs μ and σ. In practice, this constraint is the Kullback-Leibler (KL) divergence between the latent distribution $\mathcal{N}(\mu, \Sigma)$ and the standard normal distribution $\mathcal{N}(0, I)$ where I is the ℓ-dimensional identity matrix. Thus, the objective function for VAEs takes the form

$$\mathcal{J}^{\mathrm{VAE}}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 + \mathrm{KL}(\mathcal{N}(\mu, \Sigma), \mathcal{N}(0, I)) \tag{7.2}$$

where the first term represents the reconstruction error and the second term represents the KL divergence between the latent distribution and the standard normal distribution. By training the encoder and decoder networks with this dual-objective function, the VAE learns to represent high-dimensional input data in a continuous low-dimensional latent space where similar inputs are characterized by similar latent distributions. Given the clear optimization problem with well-defined inputs and outputs, training VAEs is systematic and stable, however the reconstruction quality can be

unrealistic and "blurry" due to the nature of mean squared error as an objective function. The interested reader can find a more detailed explanation of VAEs in [145, 146]. We next describe how VAEs can be used to model virtual population data. However, other notable recent uses of VAEs towards precision medicine include the best performing segmentation algorithm for the Brain Tumor Segmentation Challenge (BRATS) [147] and Stacked Sparse Autoencoders (SSAEs) for nucleus detection in H&E images [148]

We applied VAEs to modeling the virtual population data of 128 tumor cell density time series generated in section 3.3. For the encoder and decoder we employ deep convolutional neural networks (see https://github.com/jtnardin/Tumor-Heterogeneity/ for code) using a latent space of size $\ell = 100$. For training and validation data, we use the first 20 time points of each of the 128 virtual patients for a total of 2560 observations. To pre-process the data, each 256×256 realization is linearly down-sampled to 128×128 and normalized pixel-wise to $[0, 1]$. The data are then randomly partitioned into 80% training and 20% validation sets. The training data are augmented using random rotations of 90° during training while no augmentations are applied to the validation set. To train the encoder and decoder networks, we minimize the VAE objective function in Eq (7.2) using the Adam optimizer with default parameters except a learning rate of 1×10^{-4} for a total of 5000 epochs with a batch size of 64. The KL Divergence term is scaled by 1×10^{-5} to keep it from dominating the reconstruction error in the objective function. We report the results of the VAE that achieves the smallest error on the validation set.

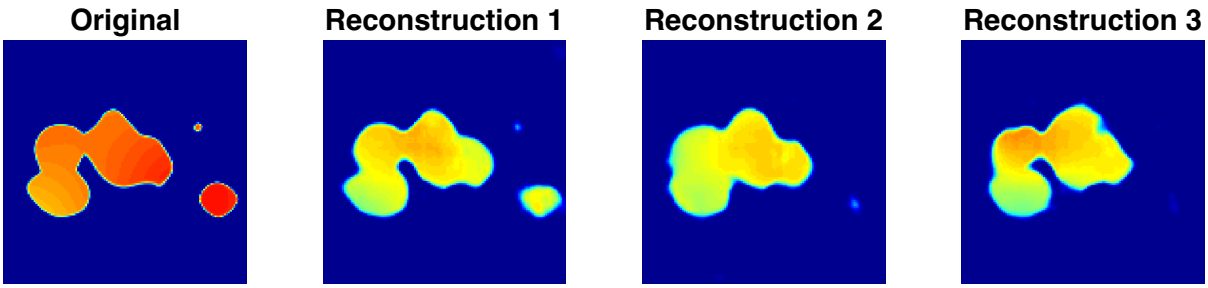

Figure 9. A virtual patient's tumor cell density (left image) is encoded into a latent distribution by the encoder network, then three random samples from the latent distribution are input to the decoder network to produce the three reconstructed images to the right of the original.

The trained VAE can be used for generating new data and producing new realizations of already existing data, which could be used synergistically with virtual populations for enhanced investigation into heterogeneity. In practice, the generated data can be used in conjunction with observed data for better mathematical model calibration or improving the performance of other machine learning models for tasks like classification or segmentation. In Figure 9 we demonstrate how the trained variational auto-encoder (VAE) can be used to produce new realizations of a given observation. An example observation (left) is selected and encoded into a latent distribution characterized by $\mu, \sigma \in \mathbb{R}^{100}$. Then the next three plots show various decoder reconstructions from samples $z \sim \mathcal{N}(\mu, \Sigma)$. Figure 9 also illustrates some of the strengths and weaknesses of VAEs. In particular, VAEs allow scientists to encode observed data into a lower-dimensional latent representation from which we can sample new

realizations that are similar to the observation itself. However the sampled reconstructions can be blurry, as shown most clearly in VAE Reconstruction 3 (Figure 9) in which the boundaries around the tumor are more diffuse.

7.3. Case study: Generative adversarial networks for virtual population generation

Generative adversarial networks (GANs) [149] are another generative modeling approach with widespread use in the field of machine learning. Like VAEs, GANs are also composed of two neural networks, called a generator and discriminator, but rather than being trained end-to-end to minimize reconstruction error, the generator and discriminator train competitively against each other. Intuitively, the generator network attempts to fool the discriminator network by producing realistic looking data while the discriminator is trained to distinguish between real and generated data. This method is formalized in the following paragraph.

The objective of GANs is to produce new samples $\hat{\boldsymbol{y}} \in \mathbb{R}^{\Omega}$ by approximating the unknown probability distribution $\mathbb{P}_{\text{data}}$ that generates the observed data $\boldsymbol{y} \in \mathbb{R}^{\Omega}$ in the training set. Since the true probability distribution is unknown, we instead wish to sample from a lower-dimensional probability distribution $\mathbb{P}_z$ that is chosen *a priori* (e.g., standard normal) and map it through a sufficiently complex function (i.e., generator neural network) to approximate the true distribution $\mathbb{P}_{\text{data}}$. Let $\boldsymbol{z} \in \mathbb{R}^{\ell}$ be a sample from $\mathbb{P}_z$, then the generator $N_G(\boldsymbol{z}; \theta_G) : \mathbb{R}^{\ell} \to \mathbb{R}^{\Omega}$ parameterized by θ_G outputs a generated data sample $\hat{\boldsymbol{y}}$. The discriminator $N_D(\boldsymbol{y}; \theta_D) : \mathbb{R}^{\Omega} \to (0, 1)$ parameterized by θ_D, where $(0, 1)$ denotes the open interval between 0 and 1, inputs real and generated data and tries to predict the binary classification (i.e., 1 for real, 0 for fake) of its inputs. In this way the discriminator represents the probability that a generated sample $\hat{\boldsymbol{y}}$ came from the true probability distribution $\mathbb{P}_{\text{data}}$.

Since GANs are composed of neural networks, they can be trained by minimizing a common objective function

$$\mathcal{J}^{\text{GAN}} = \mathbb{E}_{\boldsymbol{y} \sim \mathbb{P}_{\text{data}}}[\log N_D(\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{z} \sim \mathbb{P}_z}[\log(1 - N_D(N_G(\boldsymbol{z})))] \tag{7.3}$$

where $N_G(\boldsymbol{z}) = \hat{\boldsymbol{y}}$. The first component of Eq (7.3) trains the discriminator to maximize the probability of correctly classifying incoming real and generated data while the second component trains the generator to fool the discriminator. Once the GAN is trained, the generator network can be used to generate new samples from the learned true data distribution $\mathbb{P}_{\text{data}}$ by simply drawing samples from $\mathbb{P}_z$. Due to the adversarial objective function, training GANs can be unstable for several reasons, including mode collapse [150] where the generator outputs the same sample regardless of the input, and oscillating/vanishing gradients [151] where the adversarial network parameter updates during training unfairly benefit either the discriminator network or the generator network. Provided that a GAN is trained sufficiently well, the resulting generator model is significantly more accurate at generating realistic looking samples than VAE reconstructions. The interested reader can find a more detailed explanation of GANs in [149, 152].

GANs have been used increasingly in precision medicine applications. For example, they have been used to increase the amount of available training data for the task of liver lesion classification which boosted classification accuracy from 78.6% sensitivity and 88.4% specificity to 85.7% sensitivity and 92.4% specificity [153]. GANs have also been leveraged to address imbalanced brain tumor MRI data and to serve as an anonymization tool for patient data [154]. Several modifications of the GAN architecture exist that extend their capability to more complex domains. In particular, a GAN variant

known as CycleGAN enables unpaired image-to-image translation by combining two GANs under a single adversarial objective function [155]. This modification allows for conversion of images from one domain to another, e.g., apples to oranges, horses to zebras, etc. without the need for aligned image pairs. In precision medicine, a CycleGAN was used to translate X-ray images into synthetic CT scans for more accurate image segmentation [156].

We now demonstrate an implementation of GANs applied to the virtual population data of 128 tumor cell density time series generated in section 3.3. Similarly to the encoder and decoder for VAEs, for the generator and discriminator we employ deep convolutional neural networks (see https://github.com/jtnardin/Tumor-Heterogeneity for code) using a latent space of size $\ell = 100$. Further, we use the same 2560 observations with identical pre-processing and rotation augmentations. To train the generator and discriminator networks, we minimize the GAN objective function in Eq (7.3) using the Adam optimizer with default parameters except for learning rate 2×10^{-4} and first moment weight 0.5. These values were chosen based on [152]. The GAN is trained for a total of 50,000 iterations, where in each iteration we randomly sample and augment a batch of 64 observations. Developing metrics to know when a GAN model has converged remains the subject of ongoing research, thus we report the results from the generator that produces the best qualitative results.

GANs share many of the same applications as VAEs in which generated data can be used for model calibration and improved classification and segmentation accuracy. GANs can also work synergistically with virtual populations as a platform to investigate tumor heterogeneity. In Figure 10 we demonstrate how the trained generative adversarial network (GAN) can be used to produce new realizations of the observed data. The four plots show various generator realizations from samples $z \sim \mathcal{N}(0, I)$. Figure 10 also illustrates some of the strengths and weaknesses of GANs. In particular, we use the GAN to generate realistic looking samples, however we lack the ability to sample reconstructions from fixed observations in the training set since the GAN implementation lacks an 'encoding' step. In other words, VAEs have the ability to produce a latent sample z from an input image but when utilizing GANs one must draw z randomly each time since there is no encoder network. Therefore, GANs produce *realizations* where VAEs produce *reconstructions*.

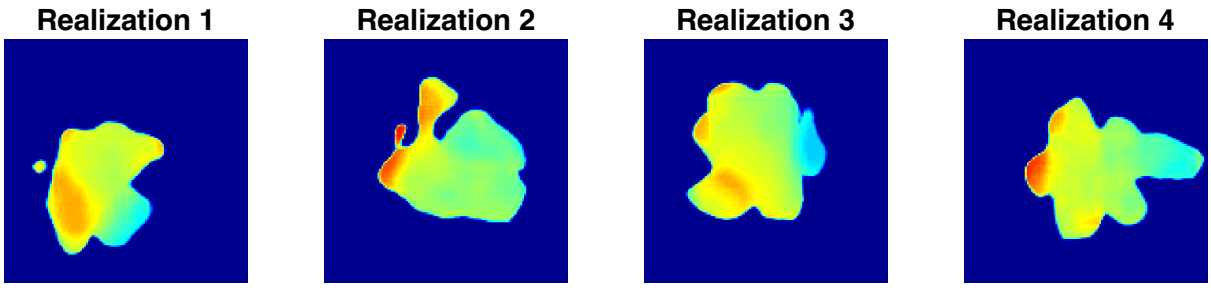

Figure 10. Plot of example virtual patient tumor cell densities obtained by inputting four random samples $z \sim \mathcal{N}(0, I)$ to the generator network of a GAN. See Figure 1 to compare the GAN generated images with examples from the virtual population data used for training.

8. Discussions

Understanding how heterogeneity between and within tumors allows cancer to harmfully spread and survive is crucial for determining strategies to develop robust patient-specific decisions for precision medicine. Several recent biological studies [157, 158] have highlighted the central role of heterogeneity for tumor survival and spread, underscoring the importance for mathematical modelers to combine models with measures of heterogeneity to help inform how heterogeneous growth and spread processes lead to tumor progression. In this review, we have detailed several methods, including virtual populations, NLME, non-parametric estimation, Bayesian statistics, and machine learning as options for modelers to use for inferring and quantifying tumor heterogeneity with the application of a mathematical model. Each of these methods have strengths and weaknesses that should be considered before implementation.

To demonstrate the strength and weakness of each method, we showcase their practical implementation in this review. We used synthetically generated data for which the true underlying distribution are known and therefore can be compared with the inferred distribution. Although these methods were not applied to clinical data in this manuscript, we included possible references and guides for the interested reader. Due to the severity of brain cancer, many open-source imaging datasets, such as the cancer imaging archive [74] consist mainly of one time point (references with one time point), which are not appropriate for all described methods. GANs are one method that can be used to analyze datasets with one time point. In order to infer parameters for underlying PDEs, multiple time points may be necessary. Multiple time-point *in vitro* [13], preclinical [159, 160], and clinical data of brain cancer could be used to quantify intertumoral heterogeneity using NLME, PMF, Bayesian, and Machine Learning Methods. Such methods could also infer heterogeneity from time-dependent *in-vitro* tumor volume data [161–163]. An interesting extension of this work would be to compare and contrast the ability of these methods to predict tumor growth in preclinical/clinical data when the underlying distribution is unknown.

Virtual populations quantify how heterogeneity arising from the distribution of model parameters, $\mathbb{P}_\beta$, leads to differences in some population-wide quantity of interest. Virtual populations are frequently used to infer $\mathbb{P}_\beta$, discover anti-cancer treatment, investigate therapeutic drug designs, and predict sensitivity of patients to different treatment regimens. We generated two virtual populations in this work: One with spatially-homogeneous model parameters and another with spatially-heterogeneous model parameters to exhibit the capabilities of the following techniques for quantifying or inferring tumor heterogeneity.

NLME is a statistical framework to infer some underlying parameter distribution for a parameterized mathematical model from distributed data. We were able to use this framework to accurately identify the underlying proliferation rate distribution from homogeneous data for benign and malignant patients with only a few virtual patient realizations (K=25). The high computational cost of this NLME did not allow us to investigate intertumoral heterogeneity through the spatially heterogeneous virtual population dataset. We suggest that extending this NLME framework to be used with a random field model is an interesting area for future research.

Non-parametric estimation within the prohorov metric framework (PMF) allows for accurate distribution inference from aggregate data without *a priori* knowledge of the underlying distribution. One drawback to non-parametric estimation is that it requires enough data in order to accurately

resolve parameter distributions. Due to potential parameter identifiability issues with estimating parameters for the random field as well as the random parameters, we only addressed inter-patient heterogeneity with non-parametric estimation in this review and did not address intratumoral heterogeneity. The PMF has previously been used to estimate intratumoral heterogeneity [89]. Optimization with respect to both intratumoral and inter-patient variability is left for future work.

Bayesian methods allow for a thorough understanding of the underlying uncertainty associated with any mechanistic parameters. This uncertainty can be investigated with a sensitivity analysis to determine which parameters most heavily influence the model output or by defining patient-specific posterior distributions for patients.

Machine learning is a recent field with promise for automating and personalizing tasks in medicine, including diagnosis and individualized therapeutic decision making. Generative modeling from machine learning allows for the reconstruction/realization of tumor images from a collection of real images. Such generated samples can be utilized to understand several underlying tumor properties, including the progression in size over time, shape, as well as variability in these measures. While we did not combine this aspect of our study with a mathematical model, the synergy of these methods with mechanistic mathematical models is a promising avenue for future research to create interpretable machine learning tools and better predict tumor invasion.

We used the Fisher-KPP model throughout this review as a model for tumor volume over space and time, which can be used in combination with spatial imaging data. We note that the model used will change between studies based upon the problem one is addressing as well as the available data. Many different mathematical models may be substituted in with the methodologies discussed within this review to aid the modeler in quantifying the heterogeneity present in their data.

Acknowledgments

SAMSI: This material was based upon work partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

This research was also supported in part by the National Institutes of Health under grant number R01EB000803 (NH) and in part by the National Science Foundation grant number DMS-1514929 (KBF).

Contributions: KF, JN, KL wrote the introduction. NH wrote the Fisher-KPP model section. NH, PN, DL, and JN wrote the virtual populations section, and NH wrote the matlab code for this section. JN, RE, and KL wrote the NLME section, and JN wrote the matlab code for this section. ER wrote the non-parametric estimation section and wrote the matlab code for this section. NH and EBP wrote the Bayesian section and NH wrote the matlab code for this section. JL, ER, and KF wrote the machine learning section, and JL wrote the python code for this section. JN wrote the discussion. All authors read and revised the final manuscript.

Conflict of interest

The authors declare no conflicts of interest.

References

1. J. Liu, H. Dang, X. W. Wang, The significance of intertumor and intratumor heterogeneity in liver cancer, *Exp. Mol. Med.*, **50** (2018), e416.
2. T. M. Grzywa, W. Paskal, P. K. Włodarski, Intratumor and intertumor heterogeneity in melanoma, *Transl. Oncol.*, **10** (2017), 956–975.
3. K. Anderson, C. Lutz, F. W. Van Delft, C. M. Bateman, Y. Guo, S. M. Colman, et al., Genetic variegation of clonal architecture and propagating cells in leukaemia, *Nature*, **469** (2011), 356–361.
4. S. Li, F. E. Garrett-Bakelman, S. S. Chung, M. A. Sanders, T. Hricik, F. Rapaport, et al., Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia, *Nat. Med.*, **22** (2016), 792–799.
5. G. C. Webb, D. D. Chaplin, Genetic variability at the human tumor necrosis factor loci., *J. Immunol.*, **145** (1990), 1278–1285.
6. A. Marusyk, K. Polyak, Tumor heterogeneity: Causes and consequences, *Biochim. Biophys. Acta, Rev. Cancer*, **1805** (2010), 105–117.
7. A. A. Alizadeh, V. Aranda, A. Bardelli, C. Blanpain, C. Bock, C. Borowski, et al., Toward understanding and exploiting tumor heterogeneity, *Nat. Med.*, **21** (2015), 846–853.
8. M. Lee, G. T. Chen, E. Puttock, K. Wang, R. A. Edwards, M. L. Waterman, et al., Mathematical modeling links wnt signaling to emergent patterns of metabolism in colon cancer, *Mol. Syst. Biol.*, **13** (2017), 912.
9. S. Terry, S. Buart, S. Chouaib, Hypoxic stress-induced tumor and immune plasticity, suppression, and impact on tumor heterogeneity, *Front. Immunol.*, **8** (2017), 1625.
10. S. Akgül, A.-M. Patch, R. C. J. D'Souza, P. Mukhopadhyay, K. Nones, S. Kempe, et al., Intratumoural heterogeneity underlies distinct therapy responses and treatment resistance in glioblastoma, *Cancers*, **11** (2019), 190.
11. T. D. Laajala, T. Gerke, S. Tyekucheva, J. C. Costello, Modeling genetic heterogeneity of drug response and resistance in cancer, *Curr. Opin. Syst. Biol.*, **17** (2019), 8–14.
12. X. Ma, J. Huang, Predicting clinical outcome of therapy-resistant prostate cancer, *Proc. Natl. Acad. Sci.*, **116** (2019), 11090–11092.
13. A. M. Stein, T. Demuth, D. Mobley, M. Berens, L. M. Sander, A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment, *Biophys. J.*, **92** (2007), 356–365.
14. T. L. Stepien, E. M. Rutter, Y. Kuang, Traveling waves of a go-or-grow model of glioma growth, *SIAM J. Appl. Math.*, **78** (2018), 1778–1801.
15. A. Martínez-González, G. F. Calvo, L. A. P. Romasanta, V. M. Pérez-García, Hypoxic cell waves around necrotic cores in glioblastoma: A biomathematical model and its therapeutic implications, *Bull. Math. Biol.*, **74** (2012), 2875–2896.
16. H. Hatzikirou, D. Basanta, M. Simon, K. Schaller, A. Deutsch, 'Go or grow': The key to the emergence of invasion in tumour progression?, *Math. Med. Biol.: A J. IMA*, **29** (2012), 49–65.

17. Z. Husain, P. Seth, V. P. Sukhatme, Tumor-derived lactate and myeloid-derived suppressor cells: Linking metabolism to cancer immunology, *Oncoimmunology*, **2** (2013), e26383.

18. A. N. Mendler, B. Hu, P. U. Prinz, M. Kreutz, E. Gottfried, E. Noessner, Tumor lactic acidosis suppresses CTL function by inhibition of p38 and JNK/c-Jun activation, *Int. J. Cancer*, **131** (2012), 633–640.

19. K. Fischer, P. Hoffmann, S. Voelkl, N. Meidenbauer, J. Ammer, M. Edinger, et al., Inhibitory effect of tumor cell-derived lactic acid on human T cells, *Blood*, **109** (2007), 3812–3819.

20. E. C. Fiedler, M. T. Hemann, Aiding and abetting: How the tumor microenvironment protects cancer from chemotherapy, *Annu. Rev. Cancer Biol.*, **3** (2019), 409–428.

21. J. K. Saggar, M. Yu, Q. Tan, I. F. Tannock, The tumor microenvironment and strategies to improve drug distribution, *Front. Oncol.*, **3** (2013), 154.

22. A. I. Minchinton, I. F. Tannock, Drug penetration in solid tumours, *Nat. Rev. Cancer*, **6** (2006), 583–592.

23. A. Baldock, R. Rockne, A. Boone, M. Neal, C. Bridge, L. Guyman, et al., From patient-specific mathematical neuro-oncology to precision medicine, *Front. Oncol.*, **3** (2013), 62.

24. A. L. Baldock, S. Ahn, R. Rockne, S. Johnston, M. Neal, D. Corwin, et al., Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas, *PLoS One*, **9** (2014), e99057.

25. S. Barish, M. F. Ochs, E. D. Sontag, J. L. Gevertz, Evaluating optimal therapy robustness by virtual expansion of a sample population, with a case study in cancer immunotherapy, *Proc. Nati. Acad. Sci.*, **114** (2017), E6277–E6286.

26. J. C. L. Alfonso, L. Berk, Modeling the effect of intratumoral heterogeneity of radiosensitivity on tumor response over the course of fractionated radiation therapy, *Radiat. Oncol.*, **14** (2019), 88.

27. F. Forouzannia, H. Enderling, M. Kohandel, Mathematical modeling of the effects of tumor heterogeneity on the efficiency of radiation treatment schedule, *Bull. Math. Biol.*, **80** (2018), 283–293.

28. C. H. Morrell, J. D. Pearson, H. B. Carter, L. J. Brant, Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer, *J. Am. Stat. Assoc.*, **90** (1995), 45–53.

29. N. Hartung, S. Mollard, D. Barbolosi, A. Benabdallah, G. Chapuisat, G. Henry, et al., Mathematical modeling of tumor growth and metastatic spreading: Validation in tumor-bearing mice, *Cancer Res.*, **74** (2014), 6397–6407.

30. J. Poleszczuk, R. Walker, E. G. Moros, K. Latifi, J. J. Caudell, H. Enderling, Predicting patient-specific radiotherapy protocols based on mathematical model choice for proliferation-saturation index, *Bull. Math. Biol.*, **80** (2018), 1195–1206.

31. A. Rafii, C. Touboul, H. A. Thani, K. Suhre, J. A. Malek, Where cancer genomics should go next: a clinician's perspective, *Hum. Mol. Genet.*, **23** (2014), R69–R75.

32. D. Zardavas, A. Irrthum, C. Swanton, M. Piccart, Clinical management of breast cancer heterogeneity, *Nat. Rev. Clin. Oncol.*, **12** (2015), 381–394.

33. A. D. Barker, C. C. Sigman, G. J. Kelloff, N. M. Hylton, D. A. Berry, L. J. Esserman, I-spy 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy, *Clin. Pharmacol. Ther.*, **86** (2009), 97–100.
34. A. S. Meyer, L. M. Heiser, Systems biology approaches to measure and model phenotypic heterogeneity in cancer, *Curr. Opin. Syst. Biol.*, **17** (2019), 35–40.
35. R. A. Gatenby, E. T. Gawlinski, A reaction-diffusion model of cancer invasion, *Cancer Res.*, **56** (1996), 5745–5753.
36. P. Tracqui, G. C. Cruywagen, D. E. Woodward, G. T. Bartoo, J. D. Murray, E. Alvord Jr, A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth, *Cell Proliferation*, **28** (1995), 17–31.
37. M. A. J. Chaplain, Reaction–diffusion prepatterning and its potential role in tumour invasion, *J. Biol. Syst.*, **3** (1995), 929–936.
38. H. L. Harpold, E. C. Alvord Jr, K. R. Swanson, The evolution of mathematical modeling of glioma proliferation and invasion, *J. Neuropathol. Exp. Neurol.*, **66** (2007), 1–9.
39. R. Rockne, J. K. Rockhill, M. Mrugala, A. M. Spence, I. Kalet, K. Hendrickson, et al., Predicting the efficacy of radiotherapy in individual glioblastoma patients in vivo: a mathematical modeling approach, *Phys. Med. Biol.*, **55** (2010), 3271.
40. T. E. Yankeelov, N. Atuegwu, D. Hormuth, J. A. Weis, S. L. Barnes, M. I. Miga, et al., Clinically relevant modeling of tumor growth and treatment response, *Sci. Transl. Med.*, **5** (2013), 187ps9–187ps9.
41. D. A. Hormuth II, J. A. Weis, S. L. Barnes, M. I. Miga, E. C. Rericha, V. Quaranta, et al., Predicting in vivo glioma growth with the reaction diffusion equation constrained by quantitative magnetic resonance imaging data, *Phys. Biol.*, **12** (2015), 046006.
42. K. R. Swanson, R. C. Rostomily, E. C. Alvord Jr, A mathematical modelling tool for predicting survival of individual patients following resection of glioblastoma: A proof of principle, *Br. J. Cancer*, **98** (2008), 113–119.
43. S. Prokopiou, E. G. Moros, J. Poleszczuk, J. Caudell, J. F. Torres-Roca, K. Latifi, et al., A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation, *Radiat. Oncol.*, **10** (2015), 159.
44. C. Vaghi, A. Rodallec, R. Fanciullino, J. Ciccolini, J. Mochel, M. Mastri, et al., Population modeling of tumor growth curves, the reduced Gompertz model and prediction of the age of a tumor, in *Mathematical and Computational Oncology* (eds. G. Bebis, T. Benos, K. Chen, K. Jahn, E. Lima), Lecture Notes in Computer Science, Springer International Publishing, Cham, (2019), 87–97.
45. T. J. Sullivan, *Introduction to Uncertainty Quantification*, Springer, 2015.
46. G. J. Lord, C. E. Powell, T. Shardlow, *Introduction to Computational Stochastic PDEs*, Cambridge, 2014.
47. R. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.
48. N. Henscheid, E. Clarkson, K. J. Myers, H. H. Barrett, Physiological random processes in precision cancer therapy, *PLoS One*, **13** (2018), e0199823.

49. J. P. O'Connor, C. J. Rose, J. C. Waterton, R. A. Carano, G. J. Parker, A. Jackson, Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome, *Clin. Cancer Res.*, **21** (2015), 249–257.

50. R. J. Allen, T. R. Rieger, C. J. Musante, Efficient generation and selection of virtual populations in quantitative systems pharmacology models, *CPT: Pharmacometrics Syst. Pharmacol.*, **5** (2016), 140–146.

51. T. R. Rieger, R. J. Allen, L. Bystricky, Y. Chen, G. W. Colopy, Y. Cui, et al., Improving the generation and selection of virtual populations in quantitative systems pharmacology models, *Prog. Biophys. Mol. Biol.*, **139** (2018), 15–22.

52. D. J. Klinke, Integrating epidemiological data into a mechanistic model of type 2 diabetes: Validating the prevalence of virtual patients, *Ann. Biomed. Eng.*, **36** (2008), 321–334.

53. B. J. Schmidt, F. P. Casey, T. Paterson, J. R. Chan, Alternate virtual populations elucidate the type i interferon signature predictive of the response to rituximab in rheumatoid arthritis, *BMC Bioinf.*, **14** (2013), 221.

54. J. Bassaganya-Riera, *Accelerated Path to Cures*, Springer, 2018.

55. N. Henscheid, *Quantifying Uncertainties in Imaging-Based Precision Medicine*, Ph.D thesis, University of Arizona, 2018.

56. J. G. Albeck, J. M. Burke, S. L. Spencer, D. A. Lauffenburger, P. K. Sorger, Modeling a snap-action, variable-delay switch controlling extrinsic cell death, *PLoS Biol.*, **6** (2008), e299.

57. Q. Wu, S. D. Finley, Predictive model identifies strategies to enhance tsp1-mediated apoptosis signaling, *Cell Commun. Signaling*, **15** (2017), 53.

58. M. R. Birtwistle, J. Rauch, A. Kiyatkin, E. Aksamitiene, M. Dobrzyński, J. B. Hoek, et al., Emergence of bimodal cell population responses from the interplay between analog single-cell signaling and protein expression noise, *BMC Syst. Biol.*, **6** (2012), 109.

59. D. Brown, R. A. Namas, K. Almahmoud, A. Zaaqoq, J. Sarkar, D. A. Barclay, et al., Trauma in silico: Individual-specific mathematical models and virtual clinical populations, *Sci. Transl. Med.*, **7** (2015), 285ra61–285ra61.

60. A. Badano, C. G. Graff, A. Badal, D. Sharma, R. Zeng, F. W. Samuelson, et al., Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial, *JAMA Net. Open*, **1** (2018), e185474–e185474.

61. W. Kainz, E. Neufeld, W. E. Bolch, C. G. Graff, C. H. Kim, N. Kuster, et al., Advances in computational human phantoms and their applications in biomedical engineering—a topical review, *IEEE Trans. Radiat. Plasma Med. Sci.*, **3** (2019), 1–23.

62. E. Roelofs, M. Engelsman, C. Rasch, L. Persoon, S. Qamhiyeh, D. De Ruysscher, et al., Results of a multicentric in silico clinical trial (rococo): Comparing radiotherapy with photons and protons for non-small cell lung cancer, *J. Thorac. Oncol.*, **7** (2012), 165–176.

63. J. K. Birnbaum, F. O. Ademuyiwa, J. J. Carlson, L. Mallinger, M. W. Mason, R. Etzioni, Comparative effectiveness of biomarkers to target cancer treatment: Modeling implications for survival and costs, *Med. Decis. Making*, **36** (2016), 594–603.

64. D. Li, S. D. Finley, The impact of tumor receptor heterogeneity on the response to anti-angiogenic cancer treatment, *Integr. Biol.*, **10** (2018), 253–269.
65. K. Mustapha, Q. Gilli, J.-M. Frayret, N. Lahrichi, E. Karimi, Agent-based simulation patient model for colon and colorectal cancer care trajectory, *Procedia Comput. Sci.*, **100** (2016), 188—-197.
66. J. P. Rolland, H. H. Barrett, Effect of random background inhomogeneity on observer detection performance, *J. Opt. Soc. Am. A*, **9** (1992), 649–658.
67. H. H. Barrett, K. J. Myers, *Foundations of Image Science*, NY: John Wiley and Sons, New York, 2004.
68. M. A. Kupinski, E. Clarkson, J. W. Hoppin, L. Chen, H. H. Barrett, Experimental determination of object statistics from noisy images, *JOSA A*, **20** (2003), 421–429.
69. F. Bochud, C. Abbey, M. Eckstein, Statistical texture synthesis of mammographic images with clustered lumpy backgrounds, *Opt. Express*, **4** (1999), 33–42.
70. N. Henscheid, Generating patient-specific virtual tumor populations with reaction-diffusion models and molecular imaging data, *Math. Biosci. Eng.*, Submitted.
71. W. Hundsdorfer, J. G. Verwer, *Numerical solution of time-dependent advection-diffusion-reaction equations*, Springer Science and Business Media, 2013.
72. S. KM, M. Prah, Data from brain-tumor-progression, *The Cancer Imaging Archive*, 2018.
73. M. Davidian, D. M. Giltinan, Nonlinear models for repeated measurement data: An overview and update, *J. Agric., Biol., and Environ. Stat.*, **8** (2003), 387–419.
74. K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, et al., The cancer imaging archive (tcia): maintaining and operating a public information repository, *J. Digital Imaging*, **26** (2013), 1045–1057.
75. S. Desmée, F. Mentré, C. Veyrat-Follet, J. Guedj, Nonlinear mixed-effect models for prostate-specific antigen kinetics and link with survival in the context of metastatic prostate cancer: A comparison by simulation of two-stage and joint approaches, *Am. Assoc. Pharm. Sci. J.*, **17** (2015), 691–699.
76. A. Król, C. Tournigand, S. Michiels, V. Rondeau, Multivariate joint frailty model for the analysis of nonlinear tumor kinetics and dynamic predictions of death, *Stat. Med.*, **37** (2018), 2148–2161.
77. E. Grenier, V. Louvet, P. Vigneaux, Parameter estimation in non-linear mixed effects models with SAEM algorithm: extension from ODE to PDE, *ESAIM: Math. Modell. Numer. Anal.*, **48** (2014), 1303–1329.
78. H. Liang, Modeling antitumor activity in xenograft tumor treatment, *Biom. J.: J. Math. Methods Biosci.*, **47** (2005), 358–368.
79. H. Liang, N. Sha, Modeling antitumor activity by using a non-linear mixed-effects model, *Math. Biosci.*, **189** (2004), 61–73.
80. A. M. Stein, D. Bottino, V. Modur, S. Branford, J. Kaeda, J. M. Goldman, et al., Bcr–abl transcript dynamics support the hypothesis that leukemic stem cells are reduced during imatinib treatment, *Clin. Cancer Res.*, **17** (2011), 6812–6821.

81. T. Ferenci, J. Sápi, L. Kovács, Modelling tumor growth under angiogenesis inhibition with mixed-effects models, *Acta Polytech. Hung.*, **14** (2017), 221–234.

82. B. Ribba, G. Kaloshi, M. Peyre, D. Ricard, V. Calvez, M. Tod, et al., A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy, *Clin. Cancer Res.*, **18** (2012), 5071–5080.

83. B. Ribba, N. H. Holford, P. Magni, I. Trocóniz, I. Gueorguieva, P. Girard, et al., A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis, *CPT: Pharmacometrics Syst. Pharmacol.*, **3** (2014), 1–10.

84. N. W. Shock, R. C. Greulich, R. Andres, D. Arenberg, P. T. Costa, E. G. Lakatta, et al., *Normal human aging. the baltimore longituidinal study of aging*, U.S. Government Printing Office Publication.

85. https://www.mathworks.com/help/stats/nlmefit.html, accessed November 16, 2019.

86. H. T. Banks, V. A. Bokil, S. Hu, A. K. Dhar, R. A. Bullis, C. L. Browdy, et al., Modeling shrimp biomass and viral infection for production of biological countermeasures, *Math. Biosci. Eng.*, **3** (2006), 635–660.

87. H. T. Banks, J. L. Davis, A comparison of approximation methods for the estimation of probability distributions on parameters, *Appl. Numer. Math.*, **57** (2007), 753–777.

88. M. Sirlanci, S. E. Luczak, C. E. Fairbairn, D. Kang, R. Pan, X. Yu, et al., Estimating the distribution of random parameters in a diffusion equation forward model for a transdermal alcohol biosensor, *Automatica*, **106** (2019), 101–109.

89. E. M. Rutter, H. T. Banks, K. B. Flores, Estimating intratumoral heterogeneity from spatiotemporal data, *J. Math. Biol.*, **77** (2018), 1999–2022.

90. H. T. Banks, K. B. Flores, I. G. Rosen, E. M. Rutter, M. Sirlanci, W. C. Thompson, The prohorov metric framework and aggregate data inverse problems for random pdes, *Commun. Appl. Anal.*, **22** (2018), 415–446.

91. H. T. Banks, W. C. Thompson, Existence and consistency of a nonparametric estimator of probability measures in the prohorov metric framework, *Int. J. Pure Appl. Math.*, **103** (2015), 819–843.

92. H. Akaike, A new look at the statistical model identification, in *Selected Papers of Hirotugu Akaike* (eds. E. Parzen, K. Tanabe, G. Kitagawa), Springer New York, New York, (1998), 215–222.

93. R. Ullrich, H. Backes, H. Li, L. Kracht, H. Miletic, K. Kesper, et al., Glioma proliferation as assessed by 3-fluoro-3-deoxy-l-thymidine positron emission tomography in patients with newly diagnosed high-grade glioma, *Clin. Cancer Res.*, **14** (2008), 2049–2055.

94. E. Stretton, E. Geremia, B. Menze, H. Delingette, N. Ayache, Importance of patient dti's to accurately model glioma growth using the reaction diffusion equation, in *2013 IEEE 10th Int. Symp. Biomed. Imaging*, IEEE, 2013, 1142–1145.

95. N. C. Atuegwu, D. C. Colvin, M. E. Loveless, L. Xu, J. C. Gore, T. E. Yankeelov, Incorporation of diffusion-weighted magnetic resonance imaging data into a simple mathematical model of tumor growth, *Phys. Med. Biol.*, **57** (2011), 225–240.

96. J. Lipková, P. Angelikopoulos, S. Wu, E. Alberts, B. Wiestler, C. Diehl, et al., Personalized radiotherapy design for glioblastoma: Integrating mathematical tumor models, multimodal scans, and bayesian inference, *IEEE Trans. Med. Imaging*, **38** (2019), 1875–1884.

97. J. Kaipio, E. Somersalo, *Statistical and Computational Inverse Problems*, Springer Science and Business Media, 2006.

98. A. M. Stuart, Inverse problems: a bayesian perspective, *Acta Numer.*, **19** (2010), 451–559.

99. P. R. Jackson, J. Juliano, A. Hawkins-Daarud, R. C. Rockne, K. R. Swanson, Patient-specific mathematical neuro-oncology: using a simple proliferation and invasion tumor model to inform clinical practice, *Bull. Math. Biol.*, **77** (2015), 846–856.

100. J. Oden, *Cse 397 Lecture Notes: Foundations of Predictive Computational Science*, 2017. Available from: https://www.oden.utexas.edu/media/reports/2017/1701.pdf.

101. J. Collis, A. J. Connor, M. Paczkowski, P. Kannan, J. Pitt-Francis, H. M. Byrne, et al., Bayesian calibration, validation and uncertainty quantification for predictive modelling of tumour growth: A tutorial, *Bull. Math. Biol.*, **79** (2017), 939–974.

102. E. J. Kostelich, Y. Kuang, J. M. McDaniel, N. Z. Moore, N. L. Martirosyan, M. C. Preul, Accurate state estimation from uncertain data and models: An application of data assimilation to mathematical models of human brain tumors, *Biol. Direct*, **6** (2011), 64.

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

104. I. Babuška, F. Nobile, R. Tempone, A systematic approach to model validation based on bayesian updates and prediction related rejection criteria, *Comput. Methods Appl. Mech. Eng.*, **197** (2008), 2517–2539.

105. J. T. Oden, E. A. B. F. Lima, R. C. Almeida, Y. Feng, M. N. Rylander, D. Fuentes, et al., Toward predictive multiscale modeling of vascular tumor growth, *Arch. Comput. Methods Eng.*, **23** (2016), 735–779.

106. M. Lê, H. Delingette, J. Kalpathy-Cramer, E. R. Gerstner, T. Batchelor, J. Unkelbach, et al., MRI based bayesian personalization of a tumor growth model, *IEEE Trans. Med. Imaging*, **35** (2016), 2329–2339.

107. A. Hawkins-Daarud, S. K. Johnston, K. R. Swanson, Quantifying uncertainty and robustness in a biomathematical model–based patient-specific response metric for glioblastoma, *JCO Clin. Cancer Inf.*, **3** (2019), 1–8.

108. I. M. Sobol, Sensitivity estimates for nonlinear mathematical models, *Math. Modell. Comput. Exp.*, **1** (1993), 407–414.

109. P. Constantine, *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*, SIAM, 2015.

110. M. D. Morris, Factorial sampling plans for preliminary computational experiments, *Technometrics*, **33** (1991), 161–174.

111. C. Rasumssen, C. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

112.X. Zhu, M. Welling, F. Jin, J. Lowengrub, Predicting simulation parameters of biological systems using a gaussian process model, *Stat. Anal. Data Mining: ASA Data Sci. J.*, **5** (2012), 509–522.

113.M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, *J. R. Stat. Soc.: S. B (Stat. Method.*, **63** (2001), 425–464.

114.Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, et al., Gradient-based learning applied to document recognition, *Proc. IEEE*, **86** (1998), 2278–2324.

115.A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems* (eds. F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger), Curran Associates, (2012), 1097–1105.

116.O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention* (eds. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi), Springer International Publishing, Cham, (2015), 234–241.

117.A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, et al., A guide to deep learning in healthcare, *Nat. Med.*, **25** (2019), 24–29.

118.Z. Cao, L. Duan, G. Yang, T. Yue, Q. Chen, An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures, *BMC Med. Imaging*, **19** (2019), 51.

119.D. Ribli, A. Horváth, Z. Unger, P. Pollner, I. Csabai, Detecting and classifying lesions in mammograms with deep learning, *Sci. Rep.*, **8** (2018), 4165.

120.E. M. Rutter, J. H. Lagergren, K. B. Flores, Automated object tracing for biomedical image segmentation using a deep convolutional neural network, in *Medical Image Computing and Computer Assisted Intervention* (eds. A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López & G. Fichtinger), Springer International Publishing, Cham, (2018), 686–694.

121.S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, preprint, arXiv:1811.02629.

122.M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, et al., Brain tumor segmentation with deep neural networks, *Med. Image Anal.*, **35** (2017), 18–31.

123.Q. Xie, K. Faust, R. Van Ommeren, A. Sheikh, U. Djuric, P. Diamandis, Deep learning for image analysis: Personalizing medicine closer to the point of care, *Crit. Rev. Clin. Lab. Sci.*, **56** (2019), 61–73.

124.X. Feng, N. Tustison, C. Meyer, Brain tumor segmentation using an ensemble of 3d U-nets and overall survival prediction using radiomic features, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (eds. A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, T. van Walsum), Springer International Publishing, Cham, (2019), 279–288.

125.E. Puybareau, G. Tochon, J. Chazalon, J. Fabrizio, Segmentation of gliomas and prediction of patient overall survival: A simple and fast procedure, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (eds. A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes & T. van Walsum), Springer International Publishing, Cham, (2019), 199–209.

126.L. Sun, S. Zhang, L. Luo, Tumor segmentation and survival prediction in glioma with deep learning, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (eds. A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes & T. van Walsum), Springer International Publishing, Cham, (2019), 83–93.

127.B. Lyu, A. Haque, Deep learning based tumor type classification using gene expression data, in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, ACM, New York, (2018), 89–96.

128.P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. V. Vega, et al., Predicting cancer outcomes from histology and genomics using convolutional networks, *Proc. Natl. Acad. Sci.*, **115** (2018), E2970–E2979.

129.J. Zou, M. Huss, A. Abid, P. Mohammadi, A. Torkamani, A. Telenti, A primer on deep learning in genomics, *Nat. Genet.*, 12–18.

130.B. Shickel, P. J. Tighe, A. Bihorac & P. Rashidi, Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis, *IEEE J. Biomed. Health Inf.*, **22** (2017), 1589–1604.

131.R. A. Taylor, J. R. Pare, A. K. Venkatesh, H. Mowafi, E. R. Melnick, W. Fleischman, et al., Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach, *Academic Emerg. Med.*, **23** (2016), 269–278.

132.T. Zheng, W. Xie, L. Xu, X. He, Y. Zhang, M. You, et al., A machine learning-based framework to identify type 2 diabetes through electronic health records, *Int. J. Med. Inf.*, **97** (2017), 120–127.

133.H. Xu, Z. Fu, A. Shah, Y. Chen, N. B. Peterson, Q. Chen, et al., Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases, in *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, (2011), 1564.

134.D. Zhao, C. Weng, Combining pubmed knowledge and ehr data to develop a weighted bayesian network for pancreatic cancer prediction, *J. Biomed. Inf.*, **44** (2011), 859–868.

135.R. Caruana, Multitask learning, *Mach. Learn.*, **28** (1997), 41–75.

136.S. Ruder, An overview of multi-task learning in deep neural networks, preprint, arXiv:1706.05098.

137.N. Jaques, S. Taylor, E. Nosakhare, A. Sano, R. Picard, Multi-task learning for predicting health, stress, and happiness, in *NIPS Workshop on Machine Learning for Healthcare*, 2016.

138.R. C. Rockne, A. Hawkins-Daarud, K. R. Swanson, J. P. Sluka, J. A. Glazier, P. Macklin, et al., The 2019 mathematical oncology roadmap, *Phys. Biol.*, **16** (2019), 041005.

139.M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.*, **378** (2019), 686–707.

140.S. H. Rudy, S. L. Brunton, J. L. Proctor, J. N. Kutz, Data-driven discovery of partial differential equations, *Sci. Adv.*, **3** (2017), e1602614.

141.J. H. Lagregren, J. T. Nardini, G. M. Lavigne, E. M. Rutter, K. B. Flores, Learning partial differential equations for biological transport models from noisy spatiotemporal data, *Proc. R. Soc. A*, **476** (2020), 20190800.

142.F. Hamilton, A. L. Lloyd, K. B. Flores, Hybrid modeling and prediction of dynamical systems, *PLoS Comput. Biol.*, **13** (2017), e1005655.

143.J. Lagergren, A. Reeder, F. Hamilton, R. C. Smith, K. B. Flores, Forecasting and uncertainty quantification using a hybrid of mechanistic and non-mechanistic models for an age-structured population model, *Bull. Math. Biol.*, **80** (2018), 1578–1595.

144.N. Gaw, A. Hawkins-Daarud, L. S. Hu, H. Yoon, L. Wang, Y. Xu, et al., Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI, *Sci. Rep.*, **9** (2019), 10063.

145.D. P. Kingma, M. Welling, Auto-encoding variational bayes, preprint, , arXiv:1312.6114.

146.C. Doersch, Tutorial on variational autoencoders, preprint, arXiv:1606.05908.

147.A. Myronenko, 3d mri brain tumor segmentation using autoencoder regularization, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (eds. A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, T. van Walsum), Springer International Publishing, Cham, (2019), 311–320.

148.J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, et al., Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images, *IEEE Trans. Med. Imaging*, **35** (2015), 119–130.

149.I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial nets, in *Advances in Neural Information Processing Systems* (eds. Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger), Curran Associates, (2014), 2672–2680.

150.L. Metz, B. Poole, D. Pfau, J. Sohl-Dickstein, Unrolled generative adversarial networks, preprint, arxiv:1611.02163.

151.M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, preprint, arxiv:1701.04862.

152.A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, preprint, arXiv:1511.06434.

153.M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, *Neurocomputing*, **321** (2018), 321—-331.

154.H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, et al., Medical image synthesis for data augmentation and anonymization using generative adversarial networks, in *Simulation and Synthesis in Medical Imaging* (eds. A. Gooya, O. Goksel, I. Oguz, N. Burgos), Springer International Publishing, Cham, (2018), 1–11.

155.J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in *The IEEE International Conference on Computer Vision (ICCV)*, 2017.

156.Y. Zhang, S. Miao, T. Mansi, R. Liao, Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation, in *Medical Image Computing and Computer Assisted Intervention* (eds. A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger), Springer International Publishing, Cham, (2018), 599–607.

157. C. E. Gast, A. D. Silk, L. Zarour, L. Riegler, J. G. Burkhart, K. T. Gustafson, et al., Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival, *Sci. Adv.*, **4** (2018), eaat7828.

158. K. Mitamura, Y. Yamamoto, N. Kudomi, Y. Maeda, T. Norikane, K. Miyake, et al., Intratumoral heterogeneity of 18f-FLT uptake predicts proliferation and survival in patients with newly diagnosed gliomas, *Ann. Nucl. Med.*, **31** (2017), 46–52.

159. D. A. Hormuth II, J. A. Weis, S. L. Barnes, M. I. Miga, V. Quaranta, T. E. Yankeelov, Biophysical modeling of in vivo glioma response after whole-brain radiation therapy in a murine model of brain cancer, *Int. J. Radiat. Oncol. Biol. Phys.*, **100** (2018), 1270–1279.

160. E. M. Rutter, T. L. Stepien, B. J. Anderies, J. D. Plasencia, E. C. Woolf, A. C. Scheck, et al., Mathematical analysis of glioma growth in a murine model, *Sci. Rep.*, **7** (2017), 2508.

161. S. Benzekry, C. Lamont, J. Weremowicz, A. Beheshti, L. Hlatky, P. Hahnfeldt, Tumor growth kinetics of subcutaneously implanted Lewis Lung carcinoma cells, 2019. Available from: https://zenodo.org/record/3572401#.Xqt0TlNKjRZ.

162. M. Mastri, A. Tracz, J. M. L. Ebos, Tumor growth kinetics of human LM2-4LUC+ triple negative breast carcinoma cells, 2019. Available from:https://zenodo.org/record/3574531#.Xqtzv1NKjRZ.

163. A. Rodallec, S. Giocometti, J. Ciccolini, R. Fanciullino, Tumor growth kinetics of human MDA-MB-231 cells transfected with dTomato lentivirus, 2019. Available from: https://zenodo.org/record/3593919#.Xqt0PlNKjRZ.

AIMS Press