Society for
Mathematical
Biology

ORIGINAL ARTICLE

Learning Equations from Biological Data with Limited Time Samples

John T. Nardini, et al. *[full author details at the end of the article]*

Abstract

Equation learning methods present a promising tool to aid scientists in the modeling process for biological data. Previous equation learning studies have demonstrated that these methods can infer models from rich datasets; however, the performance of these methods in the presence of common challenges from biological data has not been thoroughly explored. We present an equation learning methodology comprised of data denoising, equation learning, model selection and post-processing steps that infers a dynamical systems model from noisy spatiotemporal data. The performance of this methodology is thoroughly investigated in the face of several common challenges presented by biological data, namely, sparse data sampling, large noise levels, and heterogeneity between datasets. We find that this methodology can accurately infer the correct underlying equation and predict unobserved system dynamics from a small number of time samples when the data are sampled over a time interval exhibiting both linear and nonlinear dynamics. Our findings suggest that equation learning methods can be used for model discovery and selection in many areas of biology when an informative dataset is used. We focus on glioblastoma multiforme modeling as a case study in this work to highlight how these results are informative for data-driven modeling-based tumor invasion predictions.

Keywords Equation learning · Numerical differentiation · Sparse regression · Model selection · Partial differential equations · Parameter estimation · Population dynamics · Glioblastoma multiforme

✉ John T. Nardini
jtnardin@ncsu.edu

Extended author information available on the last page of the article

1 Introduction

Mathematical models are a crucial tool for inferring the mechanics underlying a scientific system of study (Nardini et al. 2016) or predicting future outcomes (Ferguson et al. 2020). The task of interpreting biological data in particular benefits from mathematical modeling, as models allow biologists to test multiple hypotheses *in silico* (Ozik et al. 2018), optimally design experiments (Walter and Pronzato 1990), or create personalized medical treatment plans for patients (Baldock et al. 2014). A common question for mathematicians and biologists alike is: *which model(s) sufficiently describe a given dataset* (Warne et al. 2019)? This is a challenging question to resolve, as there may be several candidate models that can describe the data comparably well, or the underlying mechanics may be poorly understood. This challenge of inferring a data-driven mathematical model is further complicated by common issues with biological data that inhibit our understanding of the underlying dynamics. Such challenges include large amounts of noise (Francis et al. 2003; Perretti et al. 2013), sparse time sampling (Baldock et al. 2014; Massey et al. 2020; Hawkins-Daarud et al. 2019), inter-population heterogeneity (Wang et al. 2009), and complex forms of observation noise (Banks et al. 2014; Lagergren et al. 2020)

Partial differential equations (PDEs) are used to model many spatiotemporal phenomena, ranging from the signaling proteins inside of cells (Mori et al. 2008) to the migration patterns of animal populations (Garcia-Ramos and Rodriguez 2002). Reaction–diffusion–advection equations in particular are suited to describe biological processes that involve the simultaneous transport and growth or decay of a substance over time. Such equations may be written as

$$u_t = \nabla \cdot (D(u, x, t)\nabla u) - \nabla \cdot (V(u, x, t)u) + f(u, x, t), \tag{1}$$

with system-specific initial and boundary conditions for some quantity of interest $u = u(x, t), x \in \mathbb{R}^n, t \in [t_0, t_f]$ that spreads with diffusion rate $D(u, x, t)$, migrates with advection rate $V(u, x, t)$, and grows or decays with reaction rate $f(u, x, t)$. One common example used for modeling biological phenomena is the Fisher–Kolmogorov–Petrovsky–Piskunov (Fisher–KPP) equation (Fisher 1937):

$$u_t = D\Delta u + ru\left(1 - \frac{u}{K}\right). \tag{2}$$

This equation assumes u has a constant rate of diffusion $D(u, x, t) = D \in \mathbb{R}$, grows logistically with intrinsic growth rate $r \in \mathbb{R}$ until it reaches the population carrying capacity $K \in \mathbb{R}$, and does not advect ($V(u, x, t) = 0$). Given an initial condition with compact support, solutions to the Fisher–KPP equation converge to traveling wave solutions that spread with constant speed $2\sqrt{Dr}$ (Kolmogoroff et al. 1937). One can also show that the steepness of the propagating front depends on the ratio D/r; the front becomes more steep as D/r decreases and becomes more spread out as D/r increases, see Murray (2002) for details.

The Fisher–KPP equation was initially proposed to model the spread of an advantageous gene (Fisher 1937), but has since been used to model many biological

phenomena including wound-healing experiments (Jin et al. 2016; Nardini and Bortz 2018; Warne et al. 2019) and the spread of invasive species (Urban et al. 2008). In particular, it has been shown to be an informative model for glioblastoma multiforme (GBM), an aggressive brain tumor known for its heterogeneous behavior between patients (Baldock et al. 2014; Massey et al. 2020; Rockne et al. 2015). Using two time samples of standard clinical imaging in conjunction with key assumptions about how the imaging relates to cell density, the parameters D and r can be estimated for patients based on the wave front velocity and steepness (Hawkins-Daarud et al. 2019). Inferring these two parameters from patient data aids in characterizing the inter-patient heterogeneity common to this disease, thereby enabling data-driven modeling-based prognosis. For example, Baldock et al. (Baldock et al. 2014) found that a patient-specific metric of invasiveness, given by D/r, predicts the survival benefit of gross total resection for GBM patients and, more recently, Massey et al. (2020) correlated this metric with temozolomide efficacy, a chemotherapy drug currently used in the standard of care to treat GBM. The accurate selection of a mathematical model to aid in interpreting patient GBM data is a vital tool in informing our understanding of the efficacy of potential treatment plans for this disease.

The Fisher–KPP equation thus serves as a useful model that provides insight into patient GBM dynamics through model prognosis and prediction of outcomes from therapeutic interventions (Massey et al. 2020). For any scientific process (GBM progression, population growth, species migration, etc.), determining which model accurately describes the underlying dynamics can be a challenging process. Common methods to choose such a model include deriving a multi-scale model from assumptions on individual interactions (Nardini et al. 2016), theorizing a heuristic model from expert knowledge, or utilizing model selection techniques (Bortz and Nelson 2006; Ulmer et al. 2019; Warne et al. 2019). Model selection studies often consider several plausible hypotheses to describe a given process (Akaike 1998; Bortz and Nelson 2006). Whichever of these prescribed models most parsimoniously describes the data is ultimately classified as the final selected model (Burnham et al. 2002). Common criteria to select models include the Akaike information criterion (AIC) or the Bayesian information criterion (Akaike 1974; Schwarz 1978). If one has little understanding of the underlying dynamics, however, then performing a thorough model selection study may be challenging. For example, determining the form of Eq. (1) that best describes a given dataset can become a computationally infeasible problem due to the many possible combinations of $D(u, x, t)$, $V(u, x, t)$, and $f(u, x, t)$ that one may need to consider. Furthermore, separate numerical methods may be needed for the accurate simulation of these different term combinations, and determining the effects of numerical error on statistical inference is an ongoing area of research (Nardini and Bortz 2019). A robust methodology to directly infer one or a few candidate models for a given biological dataset will be a valuable tool for mathematical modelers.

Equation learning is a recent area of research utilizing methods from machine learning to infer the mathematical model underlying a given dataset (Brunton et al. 2016; Kaiser et al. 2018; Lagergren et al. 2020; Mangan et al. 2017; Rudy et al. 2017; Zhang and Lin 2018, 2019). Brunton et al. introduced the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm, which is able to discover the governing equations underlying the chaotic Lorenz system and other systems that can be described

with ordinary differential equation models (Brunton et al. 2016). This method was extended for application to PDEs in an algorithm called PDE Functional Identification of Nonlinear Dynamics (PDE-FIND) (Rudy et al. 2017). These studies have motivated many more investigations into how modelers can infer the governing equations for experimental data, such as how methods from Bayesian inference can be used for uncertainty quantification (Zhang and Lin 2018). Lagergren et al. recently demonstrated that a neural network can be used to reduce noise in data, which in turn improves the performance of the PDE-FIND algorithm in identifying biological transport models (Lagergren et al. 2020). For example, the Lagergren et al. study found that the Fisher–KPP equation can be correctly inferred from spatiotemporal data that has been corrupted with proportional error with noise levels as high as 25%.

Biological data presents many challenges for equation learning methods. Equation learning studies have not been thoroughly tested in settings where data observation is limited to a small number of time samples. As many as 300 time samples of a spatial process were observed in inferring PDE models in Lagergren et al. (2020), and the authors of Rudy et al. (2017) considered spatiotemporal datasets with as many as 501 time samples. Rudy et al. (2017) demonstrated that the PDE-FIND algorithm can reliably infer model equations when only a small number of randomly chosen spatial values are used for equation inference, but a dense number of spatiotemporal time samples were used prior to this inference step to estimate derivatives from the data. As the sampling of biological data is often sparse, equation learning methods must be robust in inferring equations with only limited time samples before they can be widely adopted for biological data. Furthermore, many biological phenomena exhibit wide variation between realizations. To the best of our knowledge, all previous equation learning studies infer the underlying model for only a single-parameter realization without considering how the final inferred equation results change over a realistic range of parameter values. Determining how equation learning methods perform in the presence of each of these data challenges is critical if equation learning methods are to become widely adopted in the biological sciences. Such an investigation is particularly relevant for biomedical applications, including GBM growth, as clinicians may only measure the tumor volume with MR images once or twice (Baldock et al. 2014), and the estimated parameters for the same reaction diffusion model can vary widely between patients (Wang et al. 2009).

We investigate two biological questions in this work: (i) how many time samples are sufficient to learn the governing equation underlying a set of data and (ii) how do these results depend on the underlying parameter values and noise levels? Such questions are of direct interest to the biological community, as data may be expensive to collect, or model parameters may not be identifiable with the available data (Rutter et al. 2017). Obtaining MR images in oncology, for example, can be expensive, leading to only a small number of images for each patient. This challenge leaves GBM modelers with only one or two MR images from which they can estimate patient-specific values, including the metric of invasiveness (D/r) or individual diffusion and growth parameters (D and r) (Baldock et al. 2014; Hawkins-Daarud et al. 2019). Similar challenges are present in ecology where an invading species' range and population size must be estimated from partial data and may only be measured annually or at a small number of locations (Dwyer et al. 1998; Lubina and Levin 1988). Developing
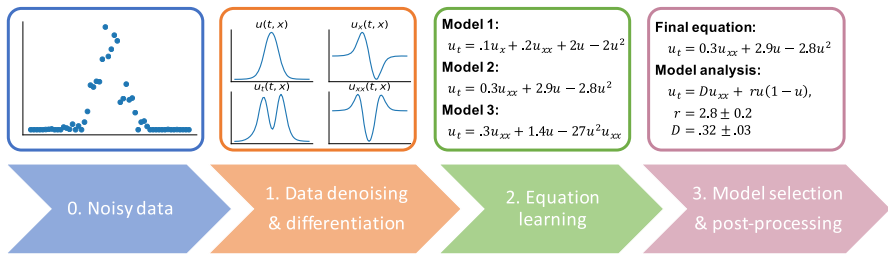
Fig. 1 Visualization of our data denoising and equation learning methodology. The steps in this methodology include 1. data denoising, where we use an ANN to smooth input noisy spatiotemporal data, which can then be used for numerical differentiation, 2. equation learning, where we infer a small number of candidate models to describe the dynamics of the given dataset, 3. model selection and post-processing, where we infer which of these models parsimoniously describes the given data and interpret the final selected model

methods to determine which datasets can reliably be used for equation learning is a crucial step before the broad adoption of such methods for mathematical modeling of experimental, clinical, or field data.

The goal of this work is to examine the success of equation learning methodologies in the presence of limited time sampling, large noise levels, and different parameter combinations. We focus on the Fisher–KPP equation in this study due to its wide use in the mathematical biology literature, but the results presented in this work will extend to many common models of biological phenomena. We begin by discussing data generation and introducing our equation learning methodology in Sect. 2. We present our results on equation learning, fit and predicted system dynamics, parameter estimation, and uncertainty quantification in Sect. 3. We further discuss these results in Sect. 4 and give final conclusions on this work and its applicability to biological studies in Sect. 5.

2 Methods

In this section we describe the datasets used throughout this study and detail the implementation of our data denoising, equation learning, and model selection methods. We discuss data generation in Sect. 2.1, data denoising in Sect. 2.2, equation learning and model selection in Sect. 2.3, and methods for parameter estimation and uncertainty quantification from the inferred model in Sect. 2.4. The equation learning methodology we adopt is summarized in Fig. 1. The code used for this methodology is available online at https://github.com/biomathlab/PDE-Learning-few-time-samples.

2.1 Data Generation

We assume data are generated from one-dimensional spatial simulations of Eq. (2) with an initial condition given by

$$u(x, t = 0) = 0.1\exp\left(-\frac{x^2}{.005}\right) \tag{3}$$

We assume data are sampled over spatiotemporal grids of the form

$$x_i = x_0 + (i - 1)\Delta x, \quad i = 1, \ldots, M, \quad \Delta x = \frac{x_f - x_0}{M - 1} \tag{4}$$

$$t_j = t_0 + (j - 1)\Delta t, \quad j = 1, \ldots, N, \quad \Delta t = \frac{t_f - t_0}{N - 1}. \tag{5}$$

The spatial grid used for all simulations here is specified by $x_0 = -17$ cm, $x_f = 17$ cm, $M = 200$. We will consider two time scales in this study by fixing $t_0 = 0.15$ years and letting either $t_f = 0.5$ years (denoted as a *short simulation*) or $t_f = 3$ years (denoted as a *long simulation*) and discuss values of N below.

We assume the data arise from the observation model

$$y_{i,j} = u(x_i, t_j) + w_{i,j}\epsilon_{i,j}; \quad i = 1, \ldots, M, \quad j = 1, \ldots, N. \tag{6}$$

In Eq. (6), the data points, $y_{i,j}$, are assumed to be observations of Eq. (2) that have been corrupted by noise from errors in the data collection process. Any negative values of $y_{i,j}$ are manually set to zero. The entire spatiotemporal dataset may be written as $\mathbf{y} = \{y_{i,j}\}_{i=1,\ldots,M}^{j=1,\ldots,N}$, and $\mathbf{y}_j$ will denote all spatial data points at the time t_j. To match a previous study (Lagergren et al. 2020), we assume the error takes the form of a statistical model with weights given by

$$w_{i,j} = \sigma u(x_i, t_j)^\gamma. \tag{7}$$

In these weights, the $\epsilon_{i,j}$ terms are independent and identically distributed (*i.i.d.*) realizations of a standard normal random variable with mean 0 and variance 1, and we will set $\sigma = 0.01$ or 0.05. We set $\gamma = 0.5$ in this work, in which Eqs. (6) and (7) form a proportional error statistical model, meaning that the variance of the data point $y_{i,j}$ is proportional to $u(x_i, t_j)$ (Banks et al. 2014). Such a statistical model has been used previously for modeling cell fluorescence data (Banks et al. 2011). In practice, however, the appropriate statistical model depends on the system under study and how data are collected. For example, a lognormal model was assumed for viral data in Bortz and Nelson (2006). We note that although we assume the constant γ is known *a priori* in Eq. (7), there exist methods for determining the value of γ when it is unknown (Banks et al. 2016). To further aid in the denoising and equation learning methods presented later on, we manually remove spatial data points from each dataset for which the value of $|y_{i,j}|$ never exceeds 10^{-4} for any time points $j = 1, \ldots, N$.

We are concerned with several data aspects in this study to investigate the performance of our methodology in the presence of biological data challenges. Namely, we vary the number of time samples N, the parameter values (D, r) parameterizing $u(x, t)$, whether data are sampled over the long or short time interval, as well as the noise level in the data. We will vary each of these values over the following domains: $N = \{3, 5, 10\}$, $(D, r) = \{(3, 3), (30, 30), (30, 3), (3, 30)\}$, $t_f = \{0.5, 3.0\}$, and
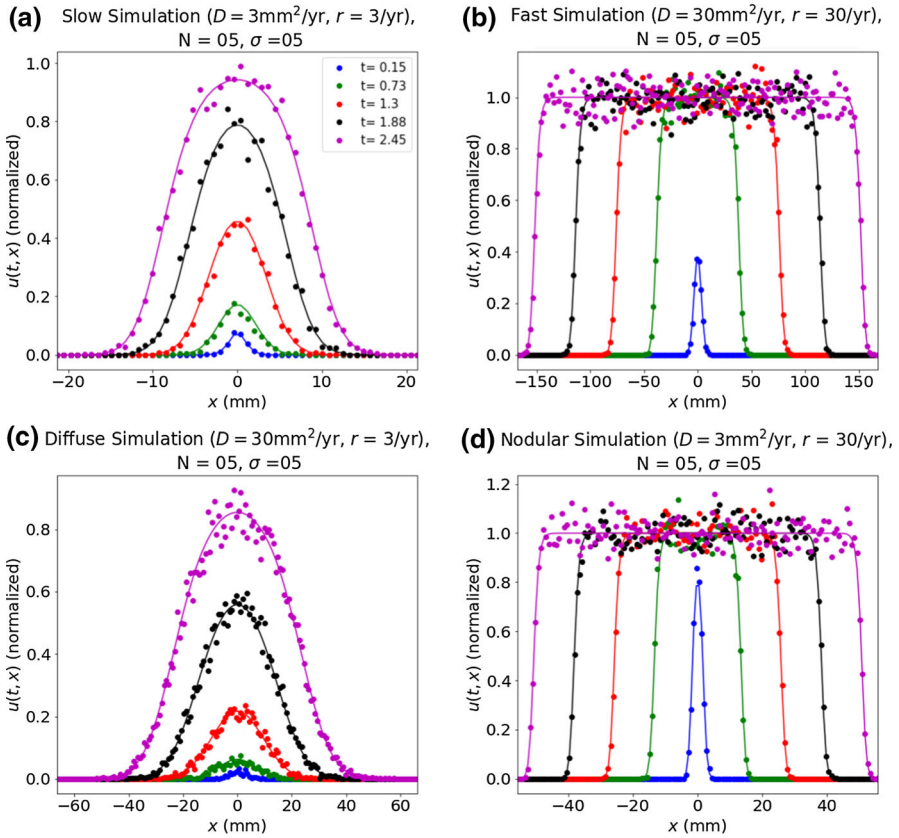
Fig. 2 (Colour figure online) Simulated noisy data of the Fisher–KPP equation with 5% noise and $N = 5$ time samples for the **a** slow simulation, **b** fast simulation, **c** diffuse simulation, and **d** nodular simulation. Solid lines represent noiseless simulations of Eq. (2), and points represent observations generated by Eq. (6). Blue plots correspond to $t = 0.15$ years, green plots correspond to $t = 0.73$ years, red plots correspond to $t = 1.3$ years, black plots correspond to $t = 1.88$ years, and magenta plots correspond to $t = 2.45$ years. Note that both the vertical and horizontal axes change between all four images based on the profile of $u(t, x)$

$\sigma = \{0.01, 0.05\}$. The units for D, r, and t_f are mm^2/year, 1/year, and year, respectively. We will refer to each of these (D, r) combinations as slow, fast, diffuse, and nodular simulations, respectively. Figure 2 depicts some resulting datasets from these four simulations over the long time interval, and we observe that the slow simulation invades the spatial domain the least, whereas the fast simulation invades the spatial domain most for the four simulations. The diffuse and nodular simulations invade a similar amount of the spatial domain, but the diffuse simulation does so with a gradually varying profile, whereas the nodular simulation invades with a steep profile. These simulation names and parameters are borrowed from a previous GBM modeling study (Hawkins-Daarud et al. 2019) to cover the ranges of D and r that have been inferred for GBM patients. Figure 3 presents histograms of measured metrics of invasiveness (D/r) and velocities from 200 GBM patients as well as where the four simulations

Table 1 Summary of the varied parameters used throughout this study

Variable	Realized values	Labels
N	3, 5, 10	3, 5, 10
(D, r)	(3, 3), (30, 30), (30, 3), (3, 30)	Slow, fast, diffuse, nodular
t_f	0.5, 3	Short, long
σ	0.01, 0.05	1%, 5%

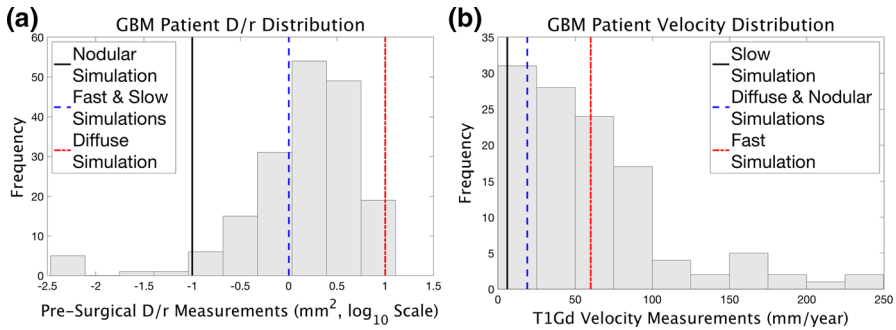Units for D are mm^2/year, r are 1/year, and t_f are year

Fig. 3 Histograms of measured GBM patient values from Neal et al. (2013a, b), including **a** presurgical D/r values and **b** T1Gd velocity

from this study fall in these distributions (Baldock et al. 2014; Neal et al. 2013a, b). Table 1 summarizes the range of parameters used in this study, as well as their labels used throughout.

2.2 Data Denoising

The first step in the equation learning process is data denoising and differentiation. Data denoising for equation learning refers to the process of approximating the solution and derivatives of an underlying dynamical system from a set of observed data. Previous methods relied on the use of finite differences for numerical differentiation in the case of noiseless data and polynomial splines in the presence of noise (Boninsegna et al. 2018; Rudy et al. 2017; Zhang and Lin 2018). Lagergren et al. recently introduced a method that leverages artificial neural networks (ANNs) to be used for robust data smoothing and numerical differentiation in the presence of large amounts of noise and heteroscedastic data (Lagergren et al. 2020).

A data denoising ANN, denoted by $h(\mathbf{x}|\theta)$, is a differentiable function with inputs consisting of spatiotemporal points, $\mathbf{x} = (x, t)$, and outputs consisting of approximations for $u(\mathbf{x})$. The ANN is parameterized by the set θ consisting of weight matrices W_i and bias vectors b_i for $i = 1, \ldots, L+1$ where L indicates the number of hidden layers or "depth" of the ANN. The size of each weight matrix and bias vector is determined by the number of neurons or "width" of the corresponding layer and its preceding layer. In particular, for the weight matrix and bias vector in the ith layer, the width of the
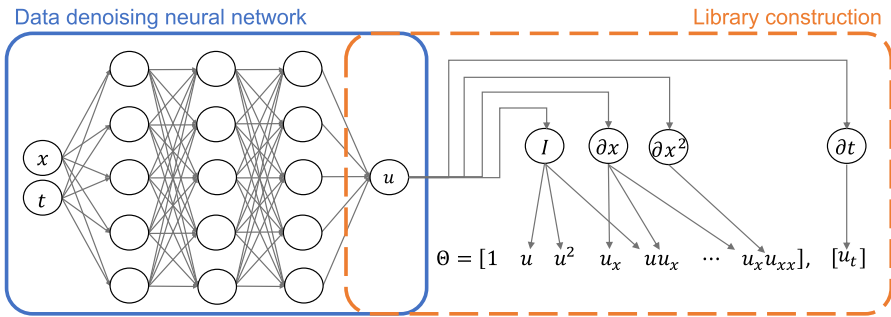
Fig. 4 Diagram of the equation learning pipeline. Left: data denoising step in which an artificial neural network (ANN) is used to approximate the solution to the dynamical system. Right: library construction step where partial derivatives of the trained ANN are used to construct the left-hand side (u_t) and library of candidate right-hand side terms (Θ) of the desired PDE

$i - 1$ layer determines the number of rows of the weight matrix, while the width of the ith layer gives the number of columns and length of the weight matrix and bias vector, respectively. "Activation functions" (e.g., the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$) are applied element-wise in between layers to introduce nonlinearity to the ANN.

With appropriately chosen activation functions, ANNs are in the class of functions known as *universal function approximators* (Hornik 1991). In practice, this means that given large enough width and depth, ANNs can approximate any continuous function on a compact domain arbitrarily well. Therefore, following (Lagergren et al. 2020), we use ANNs as surrogate models for the quantity of interest, $u(x, t)$. Backpropagation of a trained ANN can be used for numerical differentiation, which is used to construct candidate terms for the equation learning task. See Fig. 4 for a diagram of the data denoising procedure.

The ANN in this work is chosen to have three hidden layers with 256 neurons in each layer. This choice is large enough to give the ANN sufficient capacity to approximate the solution u. Concretely, this ANN can be written as

$$h(\mathbf{x}|\theta) = \phi\left(\sigma\left(\sigma\left(\sigma(\mathbf{x}W_1 + b_1)W_2 + b_2 \right) W_3 + b_3 \right) W_4 + b_4 \right) \tag{8}$$

where the weight matrices are denoted by $W_1 \in \mathbb{R}^{2 \times 256}$, $W_2, W_3 \in \mathbb{R}^{256 \times 256}$, and $W_4 \in \mathbb{R}^{256 \times 1}$, bias vectors are denoted by $b_1, b_2, b_3 \in \mathbb{R}^{256}$, and $b_4 \in \mathbb{R}$, and activation functions are denoted by $\sigma(x) = 1/(1 + e^{-x})$ and $\phi(x) = \log(1 + e^x)$. Note that ϕ (also called the "softplus" function) is specifically chosen to constrain neural network outputs to $[0, \infty)$ since we assume values of $u(\mathbf{x})$ are non-negative.

The ANN parameters θ are optimized using the first-order gradient-based Adam optimizer (Kingma and Ba 2017) with default parameters and batch optimization. The ANN is trained by minimizing the objective function given by:

$$\mathcal{L}(\theta) = \sum_{i,j}^{M,N} \left(\frac{h(x_i, t_j|\theta) - y_{i,j}}{|h(x_i, t_j|\theta)|^\gamma} \right)^2 + \sum_{i,j}^{M,N} h(x_i, t_j|\theta)^2 \Big|_{h(x_i,t_j|\theta) \notin [u_{\min}, u_{\max}]} \tag{9}$$

over the set of ANN parameters θ. The first term of Eq. (9) corresponds to the generalized least squares error between the ANN $h(\mathbf{x}|\theta)$ and the observed noisy data $\mathbf{y}$. The second term acts as a regularization penalty to ensure ANN values stay between the minimum and maximum observed values of u. We assume that u is normalized so that $u_{\min} = 0$, $u_{\max} = 1$. Note that in the small-time-sample limit (e.g. three time samples), the ANN can output unrealistic values between time samples. Therefore, for a given batch of data, the first term of Eq. (9) is evaluated on the corresponding batch of observed data values, while the second term is evaluated on a fixed 100×100 grid spanning the input domain.

To prevent the ANN from overfitting to the data, the following measures are taken. The training data are randomly partitioned into 80%/20% training and validation sets. The network parameters are iteratively updated to minimize the error in Eq. (9) on the training set. The network that results in the best error on the validation set is saved. Note that data from the validation set are never used to update network parameters. We use a small batch size of 10 to train each ANN. A small batch size acts as an additional form of regularization that (i) helps the ANN escape local minima during training and (ii) allows for better generalization (Keskar et al. 2017). We employ early stopping of 1000 (i.e., training ends if the validation error does not decrease for 1000 consecutive epochs) to ensure convergence of each ANN independently.

The trained ANN is then used in the following Equation Learning step to build a library of candidate equation terms consisting of denoised values of u and the approximated partial derivatives of u. All network training and evaluation is implemented in Python 3.6.8 using the PyTorch deep learning library.

2.3 Equation Learning, Model Selection, and Equation Post-processing

The second step of our equation learning methodology is to infer several candidate models to describe the data's dynamics. From the approximations of u and its partial derivatives from the ANN, we use the PDE-FIND algorithm (Rudy et al. 2017) to infer data-driven governing equations. This algorithm consists of first building a large library of candidate terms for the inferred model and then selecting which of these terms to include in the final model.

The PDE-FIND algorithm builds a library of potential right-hand side terms in the matrix, Θ. The columns of Θ are comprised of candidate terms for the final inferred models. We use a library consisting of ten terms given by

$$\Theta = [u_x, u_{xx}, u, u^2, uu_x, u^2u_x, uu_{xx}, u^2u_{xx}, u_x^2, uu_x^2]. \tag{10}$$

We included polynomials, partial derivatives, and the cross-products of these terms due to their wide usage in reaction–diffusion terms, but note that the choice of terms can be specified by the user and will change with different applications. For other studies, it may be more appropriate to include trigonometric terms, logarithms, convolutions of functions, etc.

From Θ, the PDE-FIND algorithm then recovers which terms should be included in the final inferred model by solving the linear system

$$u_t = \Theta\xi, \tag{11}$$

where ξ is a vector parameterizing the final inferred differential equation model. Nonzero elements of ξ parameterize the relevant right-hand side terms of the inferred model. For example, if we infer $\hat{\xi} = [0, 0.03, 0.3, -0.3, 0, 0, 0, 0, 0, 0]$ with the terms in Θ from Eq. (10), then this vector denotes that the inferred equation is the Fisher–KPP equation with $D = 0.03$ and $r = 0.3$.

Sparse linear regression methods are used to solve Eq. (11) sparsely to ensure that the final recovered model is simple and contains a small number of terms. Following our previous study (Lagergren et al. 2020), we use the adaptive forward–backward Greedy algorithm to estimate ξ (Zhang 2009). This algorithm requires estimation of the optimal tolerance with which to solve Eq. (11); to estimate this hyperparameter, we randomly split the spatiotemporal domain into 50%/50% training and validation sets. Our previous study showed that small but systematic biases from the ANN can lead to incorrect terms being inferred, so we incorporate a round of "pruning" after solving Eq. (11). During this round of pruning, we check the sensitivity of the mean-squared error $\|u_t - \Theta\hat{\xi}\|_2^2$ to each nonzero entry in ξ. The ith term of $\hat{\xi}$ is included in the final inferred equation if $\|u_t - \Theta_i\hat{\xi}_i\|_2^2$ increases by 5%, where Θ_i is a copy of Θ that is missing the ith term and $\hat{\xi}_i$ is the estimated parameter vector in the absence of this term. The 5% threshold was chosen so that the final inferred equation is sensitive to the final model terms; however, the choice to use 5% was arbitrary and not tuned to any of the datasets. A smaller threshold would increase the number of terms in the model, and a larger threshold would reduce the number of terms. The PDE-FIND algorithm is sensitive to the training-validation split of the spatiotemporal domain, so we infer 100 equations from 100 random training-validation domains and choose the three most common forms as the final inferred equations.

The third step of our equation learning methodology is model selection and post-processing of these inferred equations. To select which of the three top models best describes a given dataset, we implement the following model selection procedure. We perform forward simulations of these three top models starting with initial conditions from the first time sample, $\mathbf{y}_1$. Numerical implementation of the inferred equations is discussed in "Appendix A." From these three models, we select the model with the lowest AIC score when compared to data at the final time sample $\mathbf{y}_N$ as our selected model. The AIC here is advantageous because when two models lead to similar outputs, it will select the simpler model and avoid complex models. We include some post-processing of this model to ensure that the final model is interpretable. In particular, we ensure that the transport terms in the final recovered model can be written in a flux formulation and in turn conserve mass. For example, the diffusion term u_{xx} by itself can be re-written as $u_{xx} = (u_x)_x$ and is thus in conservation form. The term u_x^2 by itself cannot be re-written in conservation form so it would be manually removed. If the two terms u_x^2 and uu_{xx} are simultaneously recovered, then we keep both terms because $u_x^2 + uu_{xx} = (uu_x)_x$. While we could have specified such conservation equation forms in Θ manually, we chose to keep this process as flexible as possible and incorporate a small amount of post-processing for domain-specific tasks.

2.4 Parameter Estimation and Uncertainty Quantification

We will consider the performance of the PDE-FIND algorithm in parameter estimation for a given dataset by solving Eq. (11) where Θ is composed of the terms in Fisher–KPP equation. We build the two column library of terms given by $\Theta = [u(1-u), u_{xx}]$ and solve the linear system $u_t = \Theta \xi$ using the **linalg.lstsq** function in the Python Numpy subpackage. We perform this computation for 100 separate training sets comprised of 50% of the spatiotemporal domain to determine how much these parameter estimates vary. This method of randomly selecting several subsamples of the domain for uncertainty quantification is known as subagging (a sobriquet for subsample aggregating) (Buhlmann 2012). We set our parameter estimate to be the median of these 100 estimates, and we set its *normalized standard error* to be the standard error of these estimates divided by the median value. Note that typical parameter estimation approaches involving a differential equation model require explicitly solving the model over a range of parameter values. In contrast, our approach does not require solving the mathematical model because we have instead approximated the solution and its derivatives during the denoising step (see Sect. 2.2).

3 Results

We investigate the performance of our equation learning methodology in the presence of sparse time sampling over both the short and long time scales. We exhibit these results using artificial datasets simulated from the four simulations described in Table 1. Equation inference is discussed in Sect. 3.1, and parameter estimation and uncertainty quantification when the underlying equation is known are discussed in Sect. 3.2.

3.1 Equation Discovery

We test the performance of our equation learning method on artificial data that have been generated from all four simulations for $N = 3$, 5, or 10 time samples over both the short (0–0.5 years) and long (0–3 years) time scales with 1% noise (Table 2) and 5% noise levels (Table 5 in "Appendix B"). We investigate whether our equation learning methodology is able to infer the Fisher–KPP equation from noisy data and how well the final inferred equations describe the true underlying dynamics. We test the model's predicted dynamics by investigating both how the inferred models match the dynamics they were trained on and how the models predict system dynamics that they were not trained on. We further consider our methodology's ability to infer the correct underlying equation form over a finer range of (D, r) values in Sect. 3.1.3.

3.1.1 Learning Equations from Data with 1% Noise

All final inferred equations for 1% noisy data are presented in Table 2. Table cells are highlighted in bold face when the correct underlying equation form (i.e., the Fisher–KPP equation) is inferred. We also tested the ability of the inferred equations with

Table 2 Learned 1d equations from our equation learning methodology for all simulations with 1% noisy data

Slow simulation		$u_t = 3u_{xx} + 3u - 3u^2$	
σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
01	03	$u_t = -63.082u^2 + 6.394u$	$\mathbf{u_t = 2.9u_{xx} - 2.601u^2 + 2.723u}$
01	05	$\mathbf{u_t = 2.0u_{xx} - 22.572u^2 + 3.861u}$	$\mathbf{u_t = 1.8u_{xx} - 3.125u^2 + 3.061u}$
01	10	$\mathbf{u_t = 0.8u_{xx} - 44.634u^2 + 5.418u}$	$\mathbf{u_t = 2.3u_{xx} - 3.086u^2 + 3.044u}$
Diffuse simulation		$u_t = 30u_{xx} + 3u - 3u^2$	
σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
01	03	$u_t = -195.481u^2 + 6.583u$	$\mathbf{u_t = 15.3u_{xx} - 3.285u^2 + 2.485u}$
01	05	$u_t = -1.1u_x + 31.5u_{xx} + 1.303u$	$\mathbf{u_t = 21.3u_{xx} - 3.483u^2 + 3.283u}$
01	10	$u_t = 29.4u_{xx} + 2.815u$	$\mathbf{u_t = 27.9u_{xx} - 3.023u^2 + 3.003u}$
Fast simulation		$u_t = 30u_{xx} + 30u - 30u^2$	
σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
01	03	$u_t = -42.113u^2 + 43.357u + 176.8uu_{xx} - 272.1u_x^2$	$u_t = 62.4u_{xx} - 260.1u^2u_{xx}$
01	05	$\mathbf{u_t = 25.7u_{xx} - 34.379u^2 + 34.385u}$	$u_t = -53.1u_x + 69.2u_{xx} + 7.73u^2u_x$
01	10	$\mathbf{u_t = 21.1u_{xx} - 32.383u^2 + 32.471u}$	$u_t = 77.9u_{xx} - 274.1u^2u_{xx}$
Nodular simulation		$u_t = 3u_{xx} + 30u - 30u^2$	
σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
01	03	$\mathbf{u_t = 3.2u_{xx} - 27.909u^2 + 28.125u}$	$u_t = -22.315u^2 + 22.316u$
01	05	$\mathbf{u_t = 3.0u_{xx} - 30.294u^2 + 30.406u}$	$u_t = -28.583u^2 + 28.548u$
01	10	$\mathbf{u_t = 2.5u_{xx} - 30.361u^2 + 30.31u}$	$u_t = -30.215u^2 + 30.192u$

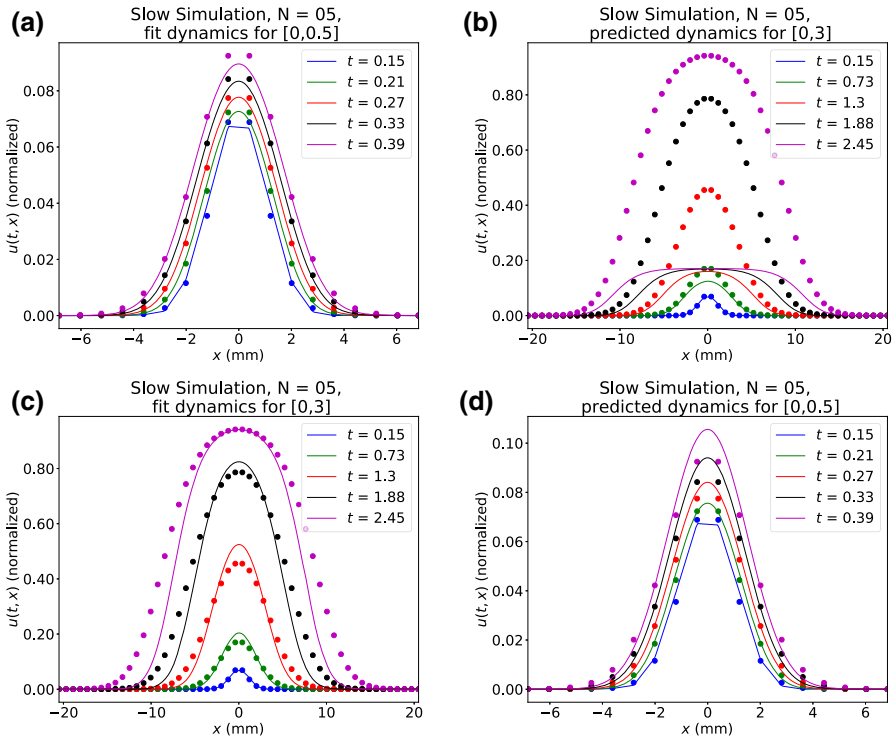Correctly inferred equation forms are mentioned in bold face

Fig. 5 Fit and predicted dynamics for the slow simulation with $N = 5$ time samples and 1% noise. **a** The simulated learned equation for the slow simulation that was inferred from data sampled over the time interval [0,0.5]. **b** The model that was inferred over the time interval [0,0.5] is used to predict the dynamics over the time interval [0,3]. **c** The simulated learned equation for the slow simulation that was inferred from data sampled over the time interval [0,3]. **d** The model that was inferred over the time interval [0,3] is used to predict the dynamics over the time interval [0,0.5]

$N = 5$ time samples to match the dynamics they were trained on and predict dynamics on a separate time interval (Figs. 5, 8, 9, 10 in "Appendix C"). In short, we are able to infer the correct underlying equation for the slow and diffuse populations on the long time interval with as few as $N = 3$ time samples. We are able to infer the correct underlying equation for the fast and nodular populations on the short time interval with as few as $N = 5$ time samples.

The slow simulation on the short time scale For noisy data sampled over the short time interval for the slow simulation, our equation learning methodology infers the Fisher–KPP equation with $N = 5$ and 10 time samples. With $N = 3$ points, the Fisher–KPP equation is not inferred. In Fig. 5a, we simulated the inferred equation for $N = 5$ time samples over the short time scale and observe that this equation accurately matches the true underlying dynamics. When this same equation is simulated over the long time scale, it does not accurately describe the true underlying dynamics (Fig. 5b).

The slow simulation on the long time scale For noisy data sampled over the long time interval for the slow simulation, our equation learning methodology infers the

Fisher–KPP equation for all values of N considered. In Fig. 5c we simulate the inferred equation for $N = 5$ time samples over the long time scale. The simulation matches the true dynamics well in many areas, although there is some disparity at $t = 1.88$ and 2.45 years for $|x| \geq 0.75$. When this same equation is simulated over the short time scale, it accurately matches the true underlying dynamics (Fig. 5d).

The diffuse simulation on the short time scale For noisy data sampled over the short time interval for the diffuse simulation, our equation learning methodology does not infer the correct underlying equation for any of the chosen values of N. In Fig. 8 in "Appendix C," we simulated the inferred equation for $N = 5$ time samples over the short time scale and observe that this equation does not match the true underlying dynamics. When this same equation is simulated over the long time scale, it does not accurately describe the true underlying dynamics.

The diffuse simulation on the long time scale For noisy data sampled over the long time interval for the diffuse simulation, our equation learning methodology infers the Fisher–KPP equation for all values of N considered. In Fig. 8, we simulate the inferred equation for $N = 5$ time samples over the long time scale. The simulation matches the true dynamics qualitatively well in many areas. When this same equation is simulated over the short time scale, it accurately matches the true underlying dynamics.

The fast simulation on the short time scale For noisy data sampled over the short time interval for the fast simulation, our equation learning methodology infers the Fisher–KPP equation for $N = 5$ and 10 and does not infer the Fisher–KPP equation for $N = 3$. In Fig. 9 in "Appendix C," we simulated the inferred equation for $N = 5$ time samples over the short time scale and observe that the inferred equation accurately describes the true underlying dynamics. When this same equation is simulated over the long time scale, it accurately matches the true underlying dynamics.

The fast simulation on the long time scale For noisy data sampled over the long time interval for the fast simulation, our equation learning methodology does not infer the correct underlying equation for any of the chosen values of N. In Fig. 9 in "Appendix C," we simulated the inferred equation for $N = 5$ time samples over the long time scale and observe that this equation does not accurately describe the true underlying dynamics. When this same equation is simulated over the short time scale, it does not accurately describe the true underlying dynamics.

The nodular simulation on the short time scale For noisy data sampled over the short time interval for the nodular simulation, our equation learning methodology infers the Fisher–KPP equation for all values of N considered. In Fig. 10 in "Appendix C," we simulated the inferred equation for $N = 5$ time samples over the short time scale and observe that the inferred equation accurately describes the true underlying dynamics. When this same equation is simulated over the long time scale, it accurately matches the true underlying dynamics.

The nodular simulation on the long time scale For noisy data sampled over the long time interval for the nodular simulation, our equation learning methodology does not infer the correct underlying equation for any of the chosen values of N. In Fig. 10 in "Appendix C," we simulated the inferred equation for $N = 5$ time samples over the

long time scale and observe that this equation does not accurately describe the true underlying dynamics. When this same equation is simulated over the short time scale, it does not accurately describe the true underlying dynamics.

3.1.2 Learning Equations from Data with 5% Noise

All final inferred equations for 5% noisy data are presented in Table 5 in "Appendix B." Table cells are highlighted in bold face when the correct underlying equation form (i.e., the Fisher–KPP equation) is inferred. Details of the results are discussed in "Appendix B." In short, we learn the correct equation for data from the slow simulation on the long time interval with $N = 5$ and 10 time samples. For data from the diffuse equation, we infer the correct equation over the long time interval with $N = 3$ time samples. For data from the fast simulation, we infer the correct equation over the short time interval with $N = 10$ time samples. For data simulated from the nodular simulation, we infer the correct equation over the short time interval with $N = 3$ time samples and over the long time interval with $N = 10$ time samples.

We also test the ability of the inferred equations with $N = 10$ time samples to match the dynamics they were trained on and predict dynamics on a separate time interval (Fig. 11 in "Appendix C"). The learned equations from data from the slow and diffuse simulations over the long time interval can accurately predict the simulations' dynamics on the short time interval. The learned equations from data from the fast and nodular simulations over the long short interval can accurately predict the simulations' dynamics on the long time interval. Further description of the prediction results is provided in "Appendix C."

3.1.3 Learning Equations for Intermediate (D, r) Values

Recall that the four considered simulations in this study correspond to values representing somewhat extreme examples of estimated values from a cohort of GBM patients (Wang et al. 2009). In this section, we investigate the performance of our equation learning methodology for intermediate values of D and r from 1% noisy data with $N = 5$ time samples. Here we are interested in if we can infer the correct equation form and score the accuracy of an inferred model form using the true positive ratio (TPR) given by:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \tag{12}$$

where "TP" stands for true positives (nonzero terms in the final inferred equation that are nonzero), "FP" stands for false positives (nonzero terms in the final inferred equation that are zero), and "FN" stands for false negatives (zero terms in the final inferred equation that are nonzero). Note that a score of TPR = 1 indicates that the correct underlying equation form has been recovered and a TPR score less than 1 indicates that the incorrect underlying equation form has been recovered.

We inferred the equations underlying such data over 25 different (D, r) combinations (Fig. 6). We let both D and r vary over five values from a log-scale between
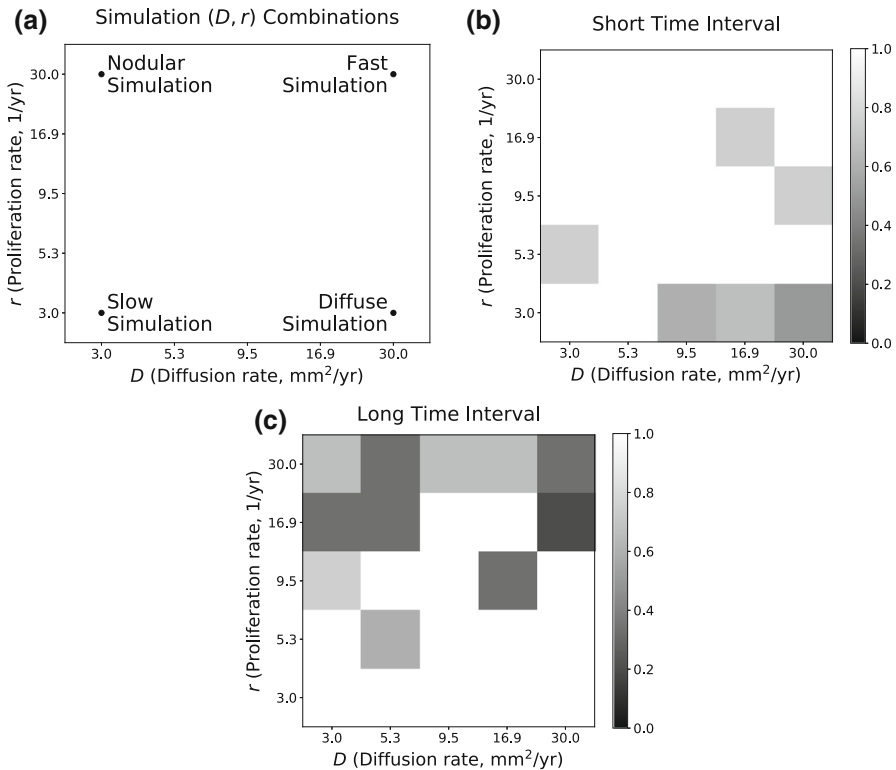
Fig. 6 TPR scores over a range of (D, r) values. **a** Depiction of where the slow, diffuse, nodular, and fast simulations fall on these plots. **b** TPR scores for the 25 (D, r) combinations for data sampled on the short time interval (0–0.5 years). **c** TPR scores for the 25 (D, r) combinations for data sampled on the long time interval (0–3 years)

their lower and upper values from Table 1. Over the short time interval, the correct underlying form is often inferred for larger values of r ($r = 5.3, 9.5, 16.9, 30.0$/year), but is not usually inferred for $r = 3.0$/year. Over the long time interval, the correct underlying form is often inferred for smaller values of r ($r = 3.0, 5.3, 9.5$/year), but is not usually inferred for $r = 16.9$ or 30.0/year. These results suggest that simulations with higher growth rates ($r > 3.0/yr.$) can be correctly inferred from data sampled over the short time interval, while simulations with lower growth rates ($r \leq 9.5/yr.$) can be correctly inferred from data sampled over the long time interval.

3.2 Parameter Estimation and Uncertainty Quantification

We investigated parameter estimation accuracy and uncertainty quantification using the PDE-FIND algorithm when we know that the underlying form is the Fisher–KPP equation. Results using data from the short time interval are presented in Sect. 3.2.1 and the long time interval in Sect. 3.2.2.

Table 3 1d parameter estimation and uncertainty quantification results for all simulations with 5% noisy data on a short time interval (0–0.5 years)

	Parameter	True	Median	% Error	Normalized SE
Slow simulation	D	3	0.928	69.1	0.25
	r	3	2.089	30.4	0.059
Fast simulation	D	30	31.422	4.7	0.084
	r	30	29.84	0.5	0.012
Diffuse simulation	D	30	22.462	25.1	0.132
	r	3	2.345	21.8	0.062
Nodular simulation	D	3	3.595	19.8	0.227
	r	30	28.588	4.7	0.027

3.2.1 Parameter Estimation and Uncertainty Quantification Over the Short Time Interval

We investigated the performance of the parameter estimation and uncertainty quantification methodology from Sect. 2.4, which is similar in concept to the PDE-FIND algorithm, but we assume the true underlying equation is known *a priori*. We use data with 5% noise and $N = 5$ time samples on the short time interval (Table 3) and attempt to estimate D and r. The fast simulation has the most accurate parameter estimation results on this time interval with error rates for D and r of 4.7% and 0.5%, respectively. The nodular simulation has error rates for D and r of 19.8% and 4.7%, respectively. The error rates for r from the slow and diffuse simulations have error rates of 30.4% and 21.8%, respectively. This methodology estimates D with an error rate of 69.1% for the slow simulation and 25.1% for the diffuse simulation. The normalized standard errors, defined as the standard error of the parameter estimates divided by their median value, for r fall between 0.012 and 0.062 for all four simulations, and the more inaccurate estimates tend to exhibit higher normalized standard errors here. The normalized standard errors for D are higher, as they fall between 0.084 and .25 for all four simulations, demonstrating that this method exhibits higher variation in estimating D than in estimating r. The fast simulation yields the smallest normalized standard error of 0.084, while the slow simulation yields the highest normalized standard error of 0.25.

3.2.2 Parameter Estimation and Uncertainty Quantification Over the Long Time Interval

We investigated the performance of the PDE-FIND algorithm for parameter estimation and uncertainty quantification with 5% noise and $N = 5$ time samples on the long time interval (Table 4) for the parameters D and r. The smallest parameter estimate error rates are found for the diffuse simulation with error rates of 4.4 and 11.8% for D and r, respectively. PDE-FIND estimates r with small error rates between 6.0 and 9.4% error for the remaining simulations. The error rates for the remaining simulations for D fall between 25.4 and 32.7%. The diffuse simulation yields the lowest normalized standard

Table 4 1d parameter estimation and uncertainty quantification results for all simulations with 5% noisy data on a long time interval (0–3 years)

	Parameter	True	Median	% Error	Normalized SE
Slow simulation	D	3	2.063	31.2	0.168
	r	3	2.786	7.1	0.02
Fast simulation	D	30	37.624	25.4	0.248
	r	30	28.189	6.0	0.047
Diffuse simulation	D	30	28.681	4.4	0.101
	r	3	2.646	11.8	0.012
Nodular simulation	D	3	2.02	32.7	0.248
	r	30	27.166	9.4	0.023

error for r of 0.012, while the fast simulation yields the highest normalized standard error of 0.047. The diffuse simulation yields the smallest normalized standard error for D of 0.101, while the fast and nodular simulations yield the highest normalized standard errors for D of 0.248.

4 Discussion

We investigated the performance of our equation learning methodology in equation inference, dynamics prediction, and parameter estimation from noisy data over a range of common challenges presented by biological data. These challenges include a wide variation in parameter values, sparse data sampling, and large amounts of noise. We used artificial data that have been generated from the Fisher–KPP equation throughout this study due to the broad applicability of this model for many biological phenomena, including tumor progression and species invasion. The diffusion and proliferation values considered in this work correspond to the ranges of measured values from GBM patients (Wang et al. 2009).

We observe in Table 2 that this methodology successfully recovers the correct underlying equation for the slow and diffuse simulations when data are observed on a long time interval of 0–3 years. The correct underlying form can be recovered with as few as three time samples on this time interval, and similar accuracy is achieved with five and ten time samples. Over the shorter time interval of 0–0.5 years, however, this methodology often either infers an incorrect equation form or infers the correct equation form with poor parameter estimates for these two simulations. With ten time samples from the slow simulation, for example, this methodology infers an equation with a carrying capacity that is ten times smaller than the true carrying capacity. Similarly, the inferred equation for the diffuse simulation will grow unbounded. These two simulations share a low proliferation rate of 3/year, suggesting that these simulations only exhibit their linear dynamics on the short interval and thus require the longer time interval for equation learning methods to accurately infer their nonlinear dynamics. These conclusions are supported by observing that the inferred equations from data

that are sampled over the longer time interval can still accurately predict the system dynamics on the shorter time interval.

Alternatively, our equation learning methodology successfully recovers the correct underlying equations for the fast and nodular simulations over the shorter time interval of 0–0.5 years. These two simulations share a high growth rate (30/year), suggesting that the systems' nonlinear dynamics are sufficiently sampled over the shorter time interval for successful equation recovery from equation learning methods. The inferred equations on this shorter time interval generalize well to predict the true underlying dynamics on the long time interval of 0–3 years. Our equation learning methodology does not infer the correct underlying equation for these two simulations when data are sampled over the long time interval. We observe in Fig. 2 that these simulations appear to have converged to their traveling wave profiles between the first and second time samples of the long time interval. Traveling waves are a known problem for the PDE-FIND algorithm because multiple reaction–diffusion–advection equations will lead to traveling wave solutions. A previous study proposed simulating bimodal data to distinguish between the advection equation and Korteweg–de Vries Eq. (Rudy et al. 2017). Here, we propose that sampling the dynamics before a traveling wave profile has been attained can lead to accurate equation inference for the Fisher–KPP equation.

These results point to some limitations of current equation learning methodologies, namely that the learned equation may depend on the number of time samples and when the data are sampled. Such considerations are particularly relevant when the true underlying equation is not known *a priori*. When the correct equation was inferred throughout this study, we observe that the inferred equation form, as well as the inferred parameters, tends not to change much with the number of data points. Simulations of correctly inferred equations were also able to accurately match data that were sampled on a different time interval. These findings suggest that future equation learning studies should thus consider testing the sensitivity of an inferred equation to the number of time samples and the ability to accurately predict unobserved system dynamics.

For low noise ($\sigma = 1\%$) data, the final inferred equation form appears robust to the number of time samples. When data are observed over an appropriate time interval where the underlying dynamics exhibit linear and nonlinear dynamics, then the final equation form typically does not change between three and ten time samples (Table 2). An exception is the fast simulation, which needed five or ten time samples to recover the correct underlying equation. With a larger amount of noise ($\sigma = 5\%$), the final inferred equation appears more sensitive to the number of observed time samples. For example, the slow simulation required five or ten time samples to infer the correct underlying equation and the fast simulation required ten time samples to infer the correct underlying equation. Interestingly, our equation learning methodology inferred the correct equation for the diffuse and nodular simulations with three time samples with 5% noise but inferred incorrect equation forms for these two simulations with five or ten time samples. We note, however, that the extra terms for the diffuse simulation here correspond to backward negative nonlinear diffusion and advection which a user may manually neglect if found in practice. Furthermore, we could have had more success recovering the correct equation form by tuning hyperparameters (such as the pruning percentage) here, but we instead focus on a flexible technique that provides users with interpretable equations for users to then alter if needed.

The four simulations considered throughout this work correspond to the outer ranges of diffusion and proliferation observed in GBM patients. We further demonstrated in Fig. 6 that our equation learning methodology is successful in recovering the correct equation learning methodology for many intermediate (D, r) values. Typically, our methodology has more success in accurate equation inference for larger values of r ($r \geq 5.3$/year) on the short time interval and for smaller values of r ($r \leq 9.5$/year) on the longer time interval for 1% noisy data.

Equation learning is promising technique for parameter estimation and uncertainty quantification due to having a low computational expense. For example, solving Eq. (11) for ξ does not require solving a differential equation, whereas typical parameter estimation routines require numerous differential equation model simulations (Banks et al. 2014). We observe that parameter estimation appears more accurate for the slow and diffuse simulations over the long time interval than for the short time interval. Similarly, parameters are more accurate for the fast and nodular simulations over the short time interval than the long time interval. Such methods could be used to obtain computationally inexpensive initial parameter estimates for frequentist parameter estimation approaches or for prior distribution specification for Bayesian inference routines (Hawkins-Daarud et al. 2019). Parameter uncertainty has been proposed previously as a measure to infer the top equation form from several plausible candidates (Zhang and Lin 2018). In support of this proposal, we observe in our work that smaller amounts of parameter uncertainty result when data are sampled over an appropriate time interval.

5 Conclusion

Methods from equation learning have the potential to become invaluable tools for researchers in developing data-driven models, and in turn understanding the mechanics underlying a biological process, estimating parameters, and predicting previously unknown system dynamics. Before such methods become commonplace in the mathematical biology field, however, it is crucial to thoroughly investigate the performance of such methods in the presence of common challenges that biological data present. We have scrutinized the performance of a state-of-the-art equation learning methodology in the presence of such challenges in this work. This equation learning methodology is composed of a data denoising step with an ANN, equation learning through the PDE-FIND algorithm (Lagergren et al. 2020; Rudy et al. 2017), and a final step comprised of model selection and post-processing to ensure the inferred equation is simple and biologically interpretable.

The biological data challenges considered in this work include sparse sampling of data, a small number of time samples, parameter heterogeneity, and large noise levels. Our equation learning method can recover the correct equation from data with a small number of time samples when the data are sampled over an appropriate time interval. Our methodology can also reliably predict previously unobserved dynamics when trained over an appropriate time interval. When this methodology is not trained over an appropriate time interval, however, the inferred equation and predicted dynamics are not reliable. Determining when to sample the data for accurate inference in this
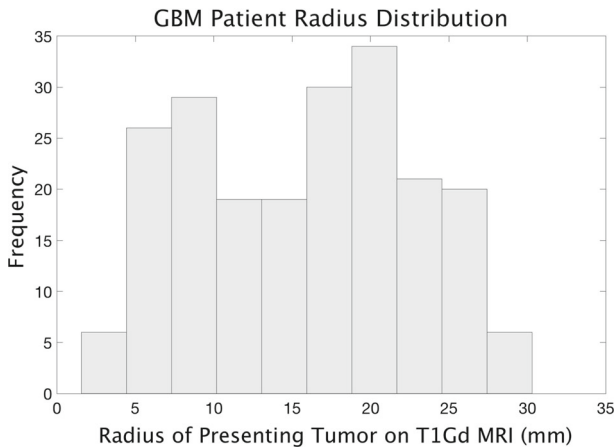
Fig. 7 Histogram of measured GBM patient radii when first presenting in the clinic obtained from T1Gd imaging

work crucially depends on the intrinsic growth rate of populations: fast-growing populations require data sampling on a short time interval and slow-growing populations require sampling on a long time interval for accurate inference. When sampled over *an appropriate* time interval, datasets exhibited a combination of both initial linear model dynamics and long-term nonlinear model dynamics. Such results suggest that an informative time interval for equation learning methods should include both of these phases for accurate inference. Noisier data require more time samples for accurate recovery: we observed in this work that three time samples were often sufficient for accurate inference of 1% noisy data, but ten time samples were required for such inference on 5% noisy data. Deciphering when equation learning methodologies are reliable in the presence of practical data challenges is important for biological research to ensure these methods are not used to make incorrect inference for a system under study. The challenges addressed in this study are prevalent in biological research, where expensive or challenging data collection may limit us to sparse datasets (Baldock et al. 2014) or measurement is often corrupted by noise (Perretti et al. 2013).

The values of diffusion and intrinsic growth considered in this work correspond to the ranges of these values that have been measured from GBM patients' MR images (Wang et al. 2009). In Figs. 3 and 7 , we depict histograms of measured D/r ratios, T1Gd velocities, and tumor radii from GBM patients before surgery. If one is interested in inferring the dynamics from patient data to inform treatment decisions for improved prognosis or predicted dynamics, then patient-estimated D/r tumor velocity measurements can be combined with the results from this study to determine how reliable the inferred equation may be. A patient's tumor could be matched to one of the four simulations considered in this work based on their measured D/r and velocity estimates (for example, if the tumor has a high D/r value and low velocity, then this tumor may correspond to the slow simulation). If patient data have been observed on an appropriate time interval for inference (a longer time interval for slow and diffuse simulations, a shorter time interval for fast and nodular simulations), then one may

have a high degree of confidence in the inferred equation and dynamics. If patient data have not been observed on an appropriate time interval for inference, then the inferred equation and dynamics might not be sufficiently supported by the data. We also observed in this work that if the population dynamics are primarily observed after the population has reached confluence, then the inferred dynamics may not be reliable. The tumor radius histogram in Fig. 7 can be used to further determine whether or not observed patient data have already reached its carrying capacity and is suitable for inference from equation learning methods.

We focused on inference of the Fisher–KPP equation in this work due to its wide use in the biological literature (Hastings et al. 2005; Hawkins-Daarud et al. 2019; Nardini and Bortz 2018). Previous equation learning studies have also successfully inferred other common models in biology, including transport equations (Lagergren et al. 2020), pattern-forming reaction–diffusion equations (Rudy et al. 2017), and compartmental epidemiological models (Mangan et al. 2017). We expect that the results, methodology, and open-source code presented in this work will enable other researchers to assess whether equation learning methods can be applied to a wide array biological data.

This work proposes many areas for future research for mathematicians and data scientists working on computational methods and modeling for biological data. We considered PDE-FIND's performance for a fixed library consisting of ten terms in this work. Future work should consider the performance and computational cost of this method with an increasing amount of terms. We observed that changing the parameters of a mathematical model influences the spatiotemporal domain over which the PDE-FIND algorithm can successfully recover the correct underlying equation. Future research should aim to infer what an informative domain may be for general ODE and PDE systems as well as methods to determine whether datasets contain high or low information content for equation learning. As an example, only sampling near an equilibrium solution will likely lead to a learned model of $u_t = 0$, whereas only sampling far away from stable equilibria may neglect nonlinear dynamics in the final learned model. It will be interesting to further investigate what combinations of transient and long-term dynamics are sufficient for accurate inference from biological systems data, which may be difficult or expensive to collect. Improvements in data denoising and equation learning methods will further inform this work. A recent neural network architecture, termed a "Physics-informed neural network," simultaneously denoises data and estimates the parameters under the assumption that the model is known (Raissi and Karniadakis 2018). The simultaneous denoising of data and equation inference will likely improve both methods and is thus an important area for future research.

A Simulating a Learned Equation

To simulate the inferred equation represented by the sparse vector $\hat{\xi}$, we begin by removing all zero terms from $\hat{\xi}$ as well as the corresponding terms from Θ. We can now define our inferred dynamical systems model as

$$u_t = \sum_i \xi_i \Theta_i. \tag{13}$$

We use the method of lines approach to simulate this equation, in which we discretize the right-hand side in space and then integrate along the t dimension. The Scipy integration subpackage (version 1.4.1) is used to integrate this equation over time using an explicit fourth-order Runge–Kutta method. We ensure that the simulation is stable by enforcing the CFL condition for an advection equation with speed $2\sqrt{Dr}$ is satisfied, e.g., $2\sqrt{Dr}\Delta t \leq \Delta x$. Some inferred equations may not be well-posed, e.g., $u_t = -u_{xx}$. If the time integration fails at any point, we manually set the model output to 10^6 everywhere to ensure this model is not selected as a final inferred model.

For the final inferred columns of $\Theta = [\Theta_1, \Theta_2, \ldots, \Theta_n]$, we define nonlinear stencils, A_{Θ_i} such that $A_{\Theta_i} u \approx \Theta_n$. As an example, we an upwind stencil (LeVeque 2007) for first-order derivative terms, such as A_{u_x}, so that $A_{u_x} u \approx u_x$. We use a central difference stencil for $A_{u_{xx}}$. For multiplicative terms, we define the stencil for A_{uu_x} as $A_{uu_x} v = u \odot (A_{u_x} v)$, where $\odot$ denotes element-wise multiplication so that $A_{uu_x} u \approx uu_x$. Similarly, we set $A_{u_x u_{xx}} = A_{u_x} A_{u_{xx}}$, etc.

B Learning the 1d Fisher–KPP Equation with 5% Noisy Data

In Table 5, we present the inferred equations for all 1d datasets considered with $\sigma = 0.05$.

The slow simulation on the short time interval For noisy data sampled over the short time interval for the slow simulation, our equation learning methodology does not infer the correct underlying equation for any values of N considered. Simulating the inferred equation for $N = 10$ time samples over the short time scale does not lead to an accurate description of the true underlying dynamics on the short time interval or prediction of the true dynamics on the long time interval.

The slow simulation on the long time interval Over the long time interval, our equation learning methodology does infer the Fisher–KPP equation with $N = 10$ time samples. Simulating the inferred equation for $N = 10$ time samples over the long time scale accurately matches the true underlying dynamics on the long time interval and accurately predicts the true dynamics on the short time interval.

The diffuse simulation on the short time interval For noisy data sampled over the short time interval for the diffuse simulation, our equation learning methodology does not infer the correct underlying equation for any values of N considered. Simulating the inferred equation for $N = 10$ time samples over the short time scale does not lead to

Table 5 Learned 1d equations from our equation learning methodology for all simulations with 5% noisy data

Slow simulation

$u_t = 3u_{xx} + 3u - 3u^2$

σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
05	03	$u_t = -101.718u^2 + 9.941u$	$u_t = 3.3u_{xx} - 2.26u^2 + 2.501u - 0.6uu_x$
05	05	$u_t = 1.637u$	$\mathbf{u_t = 1.9u_{xx} + 2.788u - 2.8u^2}$
05	10	$u_t = -35.4u^2 + 4.936u - 15.2uu_x$	$\mathbf{u_t = 2.2u_{xx} - 2.913u^2 + 3.002u}$

Diffuse simulation

$u_t = 30u_{xx} + 3u - 3u^2$

σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
05	03	$u_t = 2349.5u_x^2$	$\mathbf{u_t = 24.2u_{xx} - 4.034u^2 + 3.071u}$
05	05	$u_t = -422.687u^2 + 9.883u$	$u_t = 29.2u_{xx} - 3.647u^2 + 3.254u - 40.1uu_{xx} - 98.6u_x^2$
05	10	$u_t = 3551.9u_x^2$	$u_t = -0.54u_x + 28.4u_{xx} - 3.181u^2 + 2.996u + 2.9u^2 u_x - 36.5uu_{xx} + 22.2u_x^2$

Fast simulation

$u_t = 30u_{xx} + 30u - 30u^2$

σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
05	03	$u_t = -30.113u^2 + 29.593u$	$u_t = -155.4u^2 u_{xx} + 644.4u_x^2$
05	05	$u_t = 40.1u_{xx} - 28.118u^2 + 28.84u + 5.86u^2 u_x$	$u_t = 60.8u_{xx} - 220.3u^2 u_{xx}$
05	10	$\mathbf{u_t = 23.5u_{xx} - 29.686u^2 + 29.83u}$	$u_t = -26.923u^2 + 26.964u$

Nodular simulation

$u_t = 3u_{xx} + 30u - 30u^2$

σ	N	Learned equation (0–0.5 years)	Learned equation (0–3 years)
05	03	$\mathbf{u_t = 3.2u_{xx} - 21.375u^2 + 22.857u}$	$u_t = -34.4u^2 u_{xx} + 73.3u_x^2$
05	05	$u_t = 6.7u_{xx}$	$u_t = -26.496u^2 + 26.634u$
05	10	$u_t = 7.4u_{xx}$	$\mathbf{u_t = 2.9u_{xx} - 24.297u^2 + 24.427u}$

Correctly inferred equation forms are mentioned in bold face

an accurate description of the true underlying dynamics on the short time interval or prediction of the true dynamics on the long time interval.

The diffuse simulation on the long time interval Over the long time interval, our equation learning methodology does infer the Fisher–KPP equation with $N = 3$ time samples. Simulating the inferred equation for $N = 10$ time samples over the long time scale accurately matches the true underlying dynamics on the long time interval and accurately predicts the true dynamics on the short time interval (Fig. 11 in "Appendix C").

The fast simulation on the short time interval For noisy data sampled over the short time interval for the fast simulation, our equation learning methodology infers the Fisher–KPP equation with $N = 10$ time samples. Simulating the inferred equation for $N = 10$ time samples over the short time scale accurately matches the true underlying dynamics on the short time interval and accurately predicts the true dynamics on the long time interval (Fig. 11 in "Appendix C").

The fast simulation on the long time interval Over the long time interval, our equation learning methodology does not infer the correct underlying equation for any values of N considered. Simulating the inferred equation for $N = 10$ time samples over the short long scale does lead to an accurate description of the true underlying dynamics on the long time interval or prediction of the true dynamics on the short time interval.

The nodular simulation on the short time interval For noisy data sampled over the short time interval for the nodular simulation, our equation learning methodology infers the Fisher–KPP equation with $N = 3$ time samples. Simulating the inferred equation for $N = 10$ time samples over the short time scale does not lead to an accurate description of the true underlying dynamics on the short time interval or prediction of the true dynamics on the long time interval.

The nodular simulation on the long time interval Over the long time interval, our equation learning methodology infers the Fisher–KPP equation with $N = 10$ time samples. Simulating the inferred equation for $N = 10$ time samples over the long time scale accurately matches the true underlying dynamics on the long time interval and accurately predicts the true dynamics on the short time interval.

C Fit and Predicted Dynamics

The fit and predicted system dynamics for the diffuse, fast, and nodular *s* with 1% noise and $N = 5$ time samples are depicted in Figs. 8, 9, and 10, respectively. The fit and predicted dynamics for the diffuse and fast *s* with 5% noise and $N = 10$ time samples are depicted in Fig. 11.
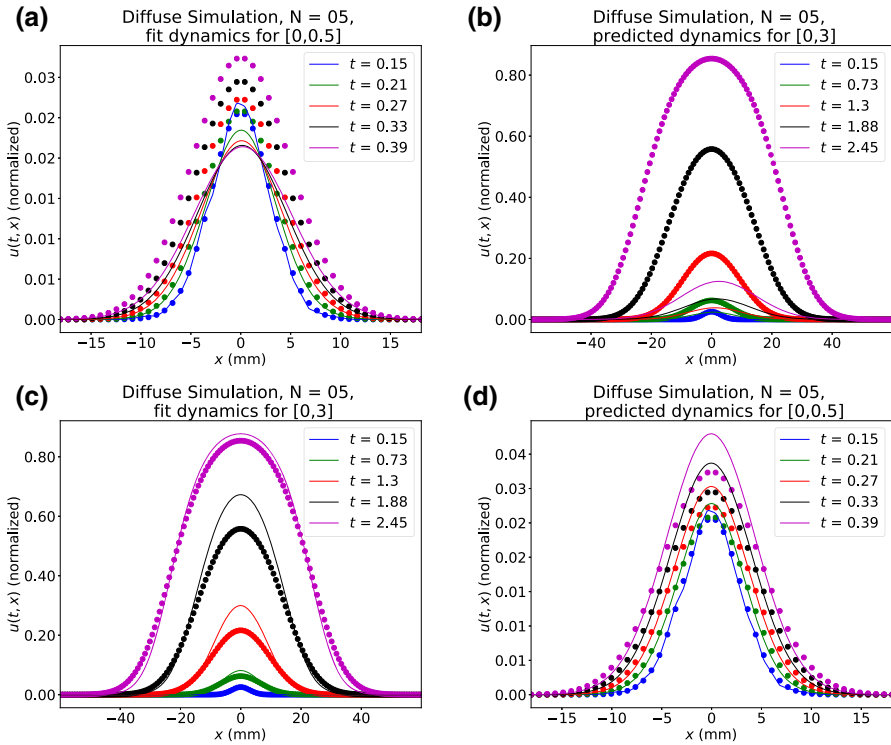
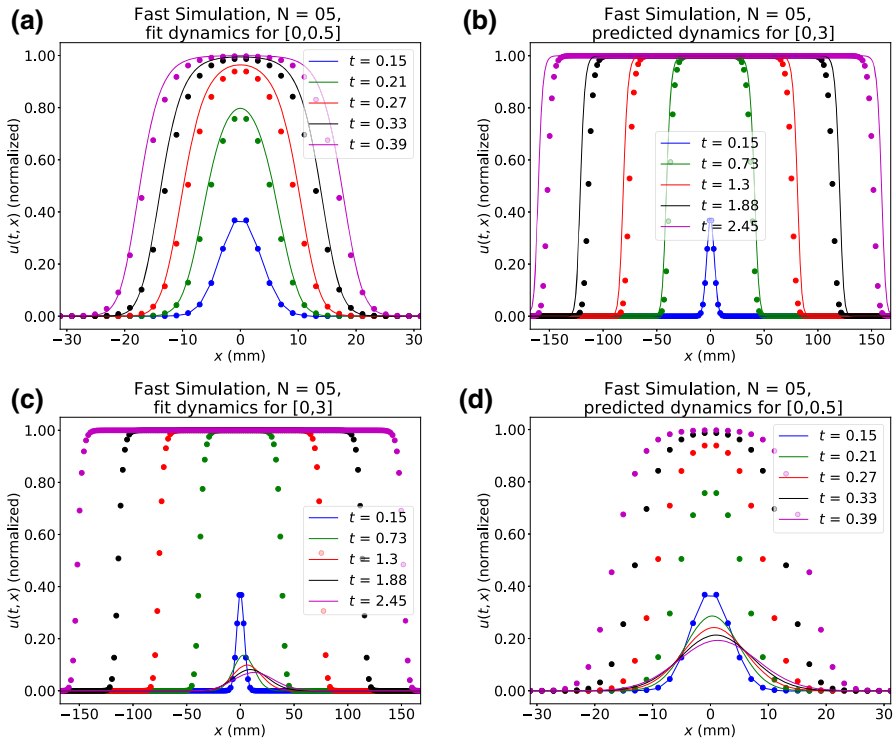**Fig. 8** Fit and predicted dynamics for the fast with $N = 5$ time samples and 1% noise. **a** The simulated learned equation for the fast that was inferred from data sampled over the time interval [0,0.5]. **b** The model that was inferred over the time interval [0,0.5] is used to predict the dynamics over the time interval [0,3]. **c** The simulated learned equation for the fast that was inferred from data sampled over the time interval [0,3]. **d** The model that was inferred over the time interval [0,3] is used to predict the dynamics over the time interval [0,0.5]. Simulated models are shown in solid lines, and the true underlying dynamics are shown by dots

Fig. 9 Fit and predicted dynamics for the diffuse with $N = 5$ time samples and 1% noise. **a** The simulated learned equation for the diffuse that was inferred from data sampled over the time interval [0,0.5]. **b** The model that was inferred over the time interval [0,0.5] 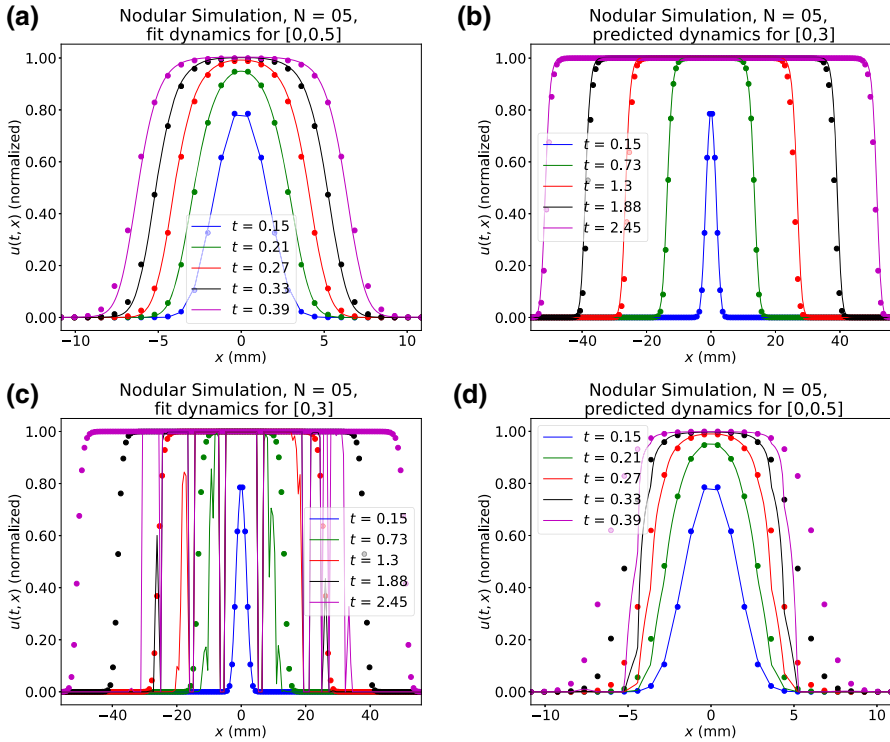is used to predict the dynamics over the time interval [0,3]. **c** The simulated learned equation for the diffuse that was inferred from data sampled over the time interval [0,3]. **d** The model that was inferred over the time interval [0,3] is used to predict the dynamics over the time interval [0,0.5]

Fig. 10 Fit and predicted dynamics for the nodular with $N = 5$ time samples and 1% noise. **a** The simulated learned equation for the nodular that was inferred from data sampled over the time interval [0,0.5]. **b** The model that was inferred over the time interval [0,0.5] is used to predict the dynamics over the time interval [0,3]. **c** The simulated learned equation for the nodular that was inferred from data sampled over the time interval [0,3]. **d** The model that was inferred over the time interval [0,3] is used to predict the dynamics over the time interval [0,0.5]. While the simulations in part **c** may appear to be the result of an unstable numerical simulation, it instead is the result of a noisy initial condition combined with a an inferred ODE model of the form $u_t = -28.58u^2 + 28.55u$. Small bumps in the initial condition grow to confluence over time as depicted in this figure
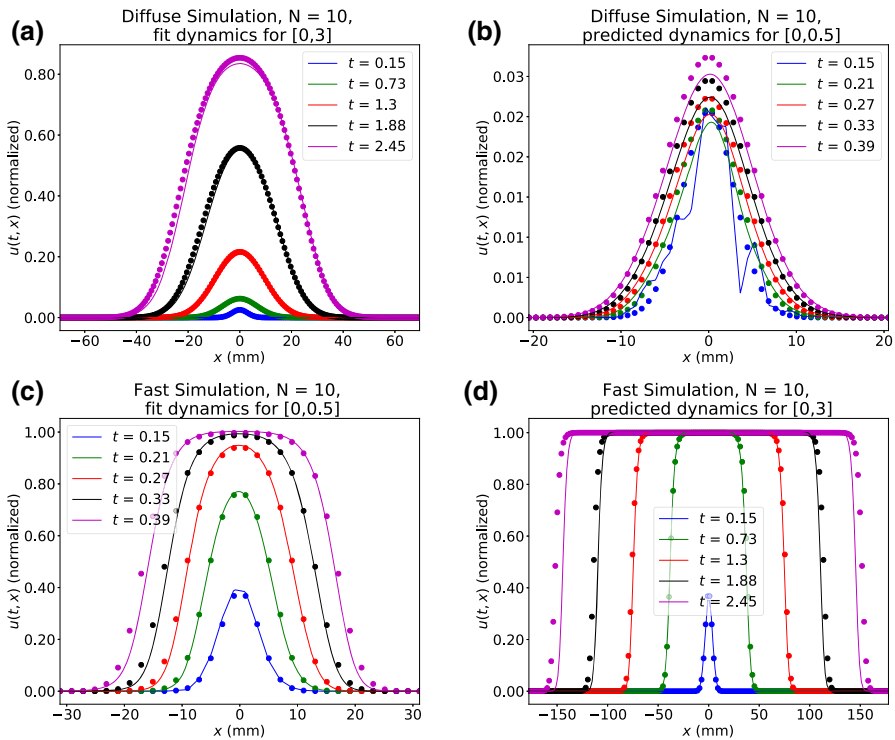
Fig. 11 Sample fit and predicted dynamics for *s* with $N = 10$ time samples and 5% noise. **a** The simulated learned equation for the diffuse that was inferred from data sampled over the time interval [0,3]. **b** The model that was inferred over the time interval [0,3] is used to predict the dynamics over the time interval [0,0.5]. **c** The simulated learned equation for the fast that was inferred from data sampled over the time interval [0,0.5]. **d** The model that was inferred over the time interval [0,0.5] is used to predict the dynamics over the time interval [0,3]

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723. https://doi.org/10.1109/TAC.1974.1100705

Akaike H (1998) Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G (eds) Selected papers of Hirotugu Akaike. Springer series in statistics. Springer, New York, pp 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15

Baldock AL, Ahn S, Rockne R, Johnston S, Neal M, Corwin D, Clark-Swanson K, Sterin G, Trister AD, Malone H, Ebiana V, Sonabend AM, Mrugala M, Rockhill JK, Silbergeld DL, Lai A, Cloughesy T, Ii GMM, Bruce JN, Rostomily RC, Canoll P, Swanson KR (2014) Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas. PLoS ONE 9(10):e99057. https://doi.org/10.1371/journal.pone.0099057

Banks HT, Sutton KL, Clayton Thompson W, Bocharov G, Roose D, Schenkel T, Meyerhans A (2011) Estimation of Cell Proliferation Dynamics Using CFSE Data. Bull Math Biol 73(1):116–150. https://doi.org/10.1007/s11538-010-9524-5

Banks HT, Hu S, Thompson WC (2014) Modeling and inverse problems in the presence of uncertainty. Chapman and Hall, Boca Raton

Banks HT, Catenacci J, Hu S (2016) Use of difference-based methods to explore statistical and mathematical model discrepancy in inverse problems. J Inverse Ill Posed Probl 24(4):413–433

Boninsegna L, Nüske F, Clementi C (2018) Sparse learning of stochastic dynamical equations. J Chem Phys 148(24):241723

Bortz DM, Nelson PW (2006) Model selection and mixed-effects modeling of HIV infection dynamics. Bull Math Biol 68(8):2005–2025. https://doi.org/10.1007/s11538-006-9084-x

Brunton SL, Proctor JL, Kutz JN (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. PNAS 113(15):3932–3937. https://doi.org/10.1073/pnas.1517384113

Buhlmann P (2012) Bagging, boosting and ensemble methods. In: Gentle JE, Hrdle WK, Mori Y (eds) Handbook of computational statistics: concepts and methods. Springer handbooks of computational statistics. Springer, Berlin, pp 985–1022. https://doi.org/10.1007/978-3-642-21551-3_33

Burnham KP, Anderson DR, Burnham KP (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd, ed edn. Springer, New York

Dwyer G, Elkinton JS, Hajek AE (1998) Spatial scale and the spread of a fungal pathogen of gypsy moth. Am Nat 152(3):485–494. https://doi.org/10.1086/286185

Ferguson N.M, Laydon D, Nedjati-Gilani et al. G (2020) Impact of non-pharmaceutical interventions NPIs to reduce COVID-19 mortality and healthcare demand. pre-print. https://doi.org/10.25561/77482

Fisher RA (1937) The wave of advance of advantageous genes. Ann Eugen 7:353–369

Francis CRIC, Hurst RJ, Renwick JA (2003) Quantifying annual variation in catchability for commercial and research fishing. Fish Bull 101(2):293–304

Garcia-Ramos G, Rodriguez D (2002) Evolutionary speed of species invasions. Evolution 56(4):661–668. https://doi.org/10.1554/0014-3820(2002)056[0661:ESOSI]2.0.CO;2

Hastings A, Cuddington K, Davies KF, Dugaw CJ, Elmendorf S, Freestone A, Harrison S, Holland M, Lambrinos J, Malvadkar U, Melbourne BA, Moore K, Taylor C, Thomson D (2005) The spatial spread of invasions: new developments in theory and evidence. Ecol Lett 8(1):91–101. https://doi.org/10.1111/j.1461-0248.2004.00687.x

Hawkins-Daarud A, Johnston SK, Swanson KR (2019) Quantifying uncertainty and robustness in a biomathematical modelbased patient-specific response metric for glioblastoma. JCO Clin Cancer Inform 3:1–8. https://doi.org/10.1200/CCI.18.00066

Hornik K (1991) Approximation capabilities of multilayer feedforward networks. Neural Netw 4(2):251–257. https://doi.org/10.1016/0893-6080(91)90009-T

Jin W, Shah ET, Penington CJ, McCue SW, Chopin LK, Simpson MJ (2016) Reproducibility of scratch assays is affected by the initial degree of confluence: experiments, modelling and model selection. J Theor Biol 390:136–145. https://doi.org/10.1016/j.jtbi.2015.10.040

Kaiser E, Kutz JN, Brunton SL (2018) Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc R Soc A 474(2219):20180335

Keskar N.S, Mudigere D, Nocedal J, Smelyanskiy M, Tang P.T.P (2017) On large-batch training for deep learning: Generalization gap and sharp minima. arXiv:1609.04836 [cs, math]

Kingma DP, Ba J (2017) Adam: a method for stochastic optimization. arXiv:1412.6980 [cs]

Kolmogoroff A, Petrovsky I, Piscounoff N (1937) Etude de l'equation de la diffusion avec croissance de la quantite de matiere et son application a un probleme biologique. Moscow Univ Bull Math 1:1–25

Lagergren JH, Nardini JT, Michael Lavigne G, Rutter EM, Flores KB (2020) Learning partial differential equations for biological transport models from noisy spatio-temporal data. Proc R Soc A 476(2234):20190800. https://doi.org/10.1098/rspa.2019.0800

LeVeque RJ (2007) Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. Society for Industrial and Applied Mathematics, Philadelphia

Lubina JA, Levin SA (1988) The spread of a reinvading species: range expansion in the California sea otter. Am Nat 131(4):526–543

Mangan NM, Kutz JN, Brunton SL, Proctor JL (2017) Model selection for dynamical systems via sparse regression and information criteria. Proc R Soc A 473(2204):20170009

Massey SC, White H, Whitmire P, Doyle T, Johnston SK, Singleton KW, Jackson PR, Hawkins-Daarud A, Bendok BR, Porter AB, Vora S, Sarkaria JN, Hu LS, Mrugala MM, Swanson KR (2020) Image-based metric of invasiveness predicts response to adjuvant temozolomide for primary glioblastoma. PLoS ONE 15(3):e0230492. https://doi.org/10.1371/journal.pone.0230492

Mori Y, Jilkine A, Edelstein-Keshet L (2008) Wave-pinning and cell polarity from a bistable reaction-diffusion system. Biophys J 94(9):3684–3697. https://doi.org/10.1529/biophysj.107.120824

Murray JD (2002) Mathematical biology I. An introduction, 3rd edn. Springer, New York

Nardini J, Bortz D (2018) Investigation of a structured Fisher's equation with applications in biochemistry. SIAM J Appl Math 78(3):1712–1736. https://doi.org/10.1137/16M1108546

Nardini JT, Bortz DM (2019) The influence of numerical error on parameter estimation and uncertainty quantification for advective PDE models. Inverse Prob 35(6):065003. https://doi.org/10.1088/1361-6420/ab10bb

Nardini JT, Chapnick DA, Liu X, Bortz DM (2016) Modeling keratinocyte wound healing: cell-cell adhesions promote sustained migration. J Theor Biol 400:103–117. https://doi.org/10.1016/j.jtbi.2016.04.015

Neal ML, Trister AD, Ahn S, Baldock A, Bridge CA, Guyman L, Lange J, Sodt R, Cloke T, Lai A, Cloughesy TF, Mrugala MM, Rockhill JK, Rockne RC, Swanson KR (2013a) Response classification based on a minimal model of glioblastoma growth is prognostic for clinical outcomes and distinguishes progression from pseudoprogression. Cancer Res 73(10):2976–2986. https://doi.org/10.1158/0008-5472.CAN-12-3588

Neal ML, Trister AD, Cloke T, Sodt R, Ahn S, Baldock AL, Bridge CA, Lai A, Cloughesy TF, Mrugala MM, Rockhill JK, Rockne RC, Swanson KR (2013b) Discriminating survival outcomes in patients with glioblastoma using a simulation-based, patient-specific response metric. PLoS ONE 8(1):1–7. https://doi.org/10.1371/journal.pone.0051951

Ozik J, Collier N, Wozniak JM, Macal C, Cockrell C, Friedman SH, Ghaffarizadeh A, Heiland R, An G, Macklin P (2018) High-throughput cancer hypothesis testing with an integrated PhysiCell-EMEWS workflow. BMC Bioinform 19(18):483. https://doi.org/10.1186/s12859-018-2510-x

Perretti C, Munch S, Sugihara G (2013) Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. PNAS 110:5253–5257

Raissi M, Karniadakis GE (2018) Hidden physics models: Machine learning of nonlinear partial differential equations. J Comput Phys 357:125–141. https://doi.org/10.1016/j.jcp.2017.11.039

Rockne RC, Trister AD, Jacobs J, Hawkins-Daarud AJ, Neal ML, Hendrickson K, Mrugala MM, Rockhill JK, Kinahan P, Krohn KA, Swanson KR (2015) A patient-specific computational model of hypoxia-modulated radiation resistance in glioblastoma using 18f-FMISO-PET. J R Soc Interface 12(103):20141174. https://doi.org/10.1098/rsif.2014.1174

Rudy SH, Brunton SL, Proctor JL, Kutz JN (2017) Data-driven discovery of partial differential equations. Sci Adv 3(4):e1602614. https://doi.org/10.1126/sciadv.1602614

Rutter EM, Stepien TL, Anderies BJ, Plasencia JD, Woolf EC, Scheck AC, Turner GH, Liu Q, Frakes D, Kodibagkar V et al (2017) Mathematical analysis of glioma growth in a murine model. Sci Rep 7(2508):1–16

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464. https://doi.org/10.1214/aos/1176344136. Publisher: Institute of Mathematical Statistics

Ulmer M, Ziegelmeier L, Topaz CM (2019) A topological approach to selecting models of biological experiments. PLoS ONE 14(3):e0213679. https://doi.org/10.1371/journal.pone.0213679

Urban MC, Phillips BL, Skelly DK, Shine R, Wiens AEJJ, DeAngelis EDL (2008) A toad more traveled: the heterogeneous invasion dynamics of cane toads in Australia. Am Nat 171(3):E134–E148. https://doi.org/10.1086/527494

Walter E, Pronzato L (1990) Qualitative and quantitative experiment design for phenomenological modelsA survey. Automatica 26(2):195–213. https://doi.org/10.1016/0005-1098(90)90116-Y

Wang CH, Rockhill JK, Mrugala M, Peacock DL, Lai A, Jusenius K, Wardlaw JM, Cloughesy T, Spence AM, Rockne R, Alvord EC, Swanson KR (2009) Prognostic significance of growth kinetics in newly diagnosed glioblastomas revealed by combining serial imaging with a novel biomathematical model. Cancer Res 69(23):9133–9140. https://doi.org/10.1158/0008-5472.CAN-08-3863

Warne DJ, Baker RE, Simpson MJ (2019) Using experimental data and information criteria to guide model selection for reactiondiffusion problems in mathematical biology. Bull Math Biol 81(6):1760–1804. https://doi.org/10.1007/s11538-019-00589-x

Zhang T (2009) Adaptive forward-backward greedy algorithm for sparse learning with linear models. In: Advances in neural information processing systems, pp 1921–1928

Zhang S, Lin G (2018) Robust data-driven discovery of governing physical laws with error bars. Proc R Soc A 474(2217):20180305

Zhang S, Lin G (2019) Robust subsampling-based sparse Bayesian inference to tackle four challenges (large noise, outliers, data integration, and extrapolation) in the discovery of physical laws from data. arXiv:1907.07788 [cs, stat]

Affiliations

John T. Nardini[1,2] · **John H. Lagergren[1]** · **Andrea Hawkins-Daarud[3]** · **Lee Curtin[3]** · **Bethan Morris[4]** · **Erica M. Rutter[5]** · **Kristin R. Swanson[3]** · **Kevin B. Flores[1]**

[1]   Department of Mathematics, North Carolina State University, Raleigh, NC, USA

[2]   The Statistical and Applied Mathematical Sciences Institute, Durham, NC, USA

[3]   Mathematical NeuroOncology Laboratory, Precision Neurotherapeutics Innovation Program, Mayo Clinic, Phoenix, AZ, USA

[4]   Centre for Mathematical Medicine and Biology, University of Nottingham, Nottingham, UK

[5]   Department of Applied Mathematics, University of California, Merced, Merced, CA, USA