Society for
Mathematical
Biology

ORIGINAL ARTICLE

A Continuation Technique for Maximum Likelihood Estimators in Biological Models

Tyler Cassidy[1] (ID)

Abstract
Estimating model parameters is a crucial step in mathematical modelling and typically involves minimizing the disagreement between model predictions and experimental data. This calibration data can change throughout a study, particularly if modelling is performed simultaneously with the calibration experiments, or during an on-going public health crisis as in the case of the COVID-19 pandemic. Consequently, the optimal parameter set, or maximal likelihood estimator (MLE), is a function of the experimental data set. Here, we develop a numerical technique to predict the evolution of the MLE as a function of the experimental data. We show that, when considering perturbations from an initial data set, our approach is significantly more computationally efficient that re-fitting model parameters while producing acceptable model fits to the updated data. We use the continuation technique to develop an explicit functional relationship between fit model parameters and experimental data that can be used to measure the sensitivity of the MLE to experimental data. We then leverage this technique to select between model fits with similar information criteria, a priori determine the experimental measurements to which the MLE is most sensitive, and suggest additional experiment measurements that can resolve parameter uncertainty.

Keywords Parameter estimation · Numerical continuation · Model calibration · Experimental design

1 Introduction

As quantitative modeling becomes more prevalent across biology and medicine (Altrock et al. 2015; Perelson 2002; Sanche et al. 2020), mathematical models are increasingly being developed during the experimental data collection that will inform model parameters. This cooperation facilitates the use of mathematical modelling to

✉ Tyler Cassidy
  t.cassidy1@leeds.ac.uk

[1] School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

inform experimental design and suggest potential intervention strategies (Zhang et al. 2022; Sanche et al. 2020; Cárdenas et al. 2022; Luo et al. 2022). The COVID-19 pandemic is a striking example of the resulting feedback loop, where mathematical models suggest intervention strategies that influence the evolving public health crisis before being re-calibrated to new data. (Holmdahl and Buckee 2020; Thompson 2020; Davies et al. 2020).

Each updated data set requires re-calibration of the model typically through computationally expensive optimization techniques. To reduce this computational cost of the re-calibration step, it is common to use the existing parameters as a starting point when performing parameter fitting to incoming experimental data sets. This approach recycles optimization work but does not utilize leverage the relationship between the initial and updated experimental data set. Here, we present a computational method to incorporate information about evolving data sets during the model validation and parameter estimation steps.

Specifically, for given model parameters and an initial experimental data set, we develop a method to predict the best-fit parameter set to an updated experimental data set. Our approach can be viewed as a numerical continuation technique (Dhooge et al. 2008; De Souza and Humphries 2019). However, rather than studying the dynamical properties of the mathematical model as a function of model parameters, we consider the evolution of best-fit model parameters as a function of the experimental data. We use the necessary condition for a local optima to write the best-fit parameters as an implicit function of the experimental data. We can then predict best-fit parameter sets for evolving experimental data without performing any optimization. Avoiding optimization leads to significant computational savings and we demonstrate these gains via two examples. In both these examples, our prediction method produces comparable model fits to randomly perturbed data sets as optimization techniques without the computational cost of solving the inverse optimization problem.

While our approach does lead to increased computational efficiency, the more immediate application of our work may be in experimental design. Specifically, we identify an explicit relationship between individual best-fit parameter values and individual experimental data points through our continuation approach. We can therefore quantify which experimental measurements are the most informative for determining best-fit parameters and measure the sensitivity of parameter estimates to perturbations in data. The role of experimental design in model selection and parameterization has been extensively studied (Silk et al. 2014; Cárdenas et al. 2022; Li and Vu 2015, 2013). In particular, Li and Vu (2015) studied how correlations between best-fit model parameters can impact practical and structural identifiability of model parameters while Silk et al. (2014) and Cárdenas et al. (2022) explored how experimental design impacts model selection from a class of possible mathematical models. Conversely, our contribution explicitly relates individual experimental measurements with individual best-fit parameter estimates. We explicitly link our continuation technique to the Fisher information matrix commonly used in optimal experimental design (Kreutz and Timmer 2009; Braniff et al. 2019). Taken together, our approach allows the increased confidence in model parametrization from optimal experimental design to be mapped directly to individual model parameters. Accordingly, we can therefore design experiments to address specific uncertainties in parameter estimates.

Furthermore, our work offers a distinct step towards understanding how robust parameter estimates are to evolving data. Many existing computational methods quantify confidence in parameterization; formal parameter sensitivity analyses (Marino et al. 2008; Maiwald et al. 2016; Zi 2011), virtual population approaches (Allen et al. 2016; Cassidy and Craig 2019; Jenner et al. 2021), or parameter identifiability analysis (Castro and de Boer 2020), often via profile likelihood computation (Raue et al. 2009, 2014; Kreutz et al. 2012), quantify how robust model predictions are to parameter variation. In particular, these techniques view the experimental data as fixed up to experimental noise and focus on the relationship between model parameters and model predictions. We offer a complementary approach to existing sensitivity analysis by explicitly studying how the best-fit parameters vary due to changes in calibration data. As we will see, our approach encodes information from local sensitivity analysis when calculating the functional relationship between the best-fit parameters and the calibration data. Consequently, while classical sensitivity analysis quantifies variability in model output due to change in model parameters, our approach considers changes in model parameters, and thus model predictions, as a function of the calibration data. We demonstrate this mapping of experimental data to best-fit parameter via an example drawn from mathematical oncology (Cassidy et al. 2021). These results, when combined with existing information criteria like the AIC or BIC (Kass and Raftery 1995), allow for modellers to quantify the robustness of best-fit parameter estimates when comparing different model fits to experimental data.

The remainder of the article is structured as follows. We begin by defining the optimization problem in Sect. 2.1. We develop the continuation method in Sect. 2.2, discuss our numerical implementation in 2.3, and explore the connection between our continuation approach and optimal experimental design in 3.1. We then turn to two examples from mathematical biology to illustrate the utility of our technique in Sect. 3.2 before finishing with a brief discussion.

2 Methods

2.1 Formulation of the Optimization Problem

Here, we introduce the framework of the underlying optimization problem. We focus on ordinary differential equation (ODE) models representing biological processes, as these models are common throughout mathematical biology. However, our approach extends to partial differential equation or delay differential equation models directly. We consider a generic ODE based model throughout the remainder of this work.

Let the model states be given by $x(t) \in \mathbb{R}^n$ with model parameters denoted by $\theta \in \Omega \subset \mathbb{R}^p$ where Ω is a subset of biologically plausible parameter values. We allow the initial condition $x(0)$ to depend on the model parameters θ. Taken together, we consider the differential equation model

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = f(x, \theta); \quad x(0) = x_0(\theta) \tag{1}$$

where f is continuously differentiable in x and θ.

We consider calibration data $\{\phi_i\}_{i=1}^{d \times m}$ representing m measurements each taken at d time points $\{t_i\}_{i=1}^{d}$. It is possible that model species are not directly comparable against the calibration data so we define the m model observables by

$$y_i(\theta) = h(x(t_i, \theta), \theta) \in \mathbb{R}^{d \times m}.$$

In what follows, we consider $m = 1$ for notational simplicity although the analysis extends for $m \geqslant 2$.

Likelihood function and objective function

Remark 1 The methods that follow do not assume a specific objective function. However, we do assume that the objective function is twice continuously differentiable as is commonly the case. For simplicity, we present the remainder of our results using the common log-likelihood formulation (Stapor et al. 2018; Maiwald et al. 2016).

The likelihood describes the probability of observing experimental data ϕ as a function of θ and is given by

$$\mathcal{L}(y(\theta), \phi) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi \sigma_i^2}} \exp \left[-\frac{(y_i(\theta) - \phi_i^*)^2}{\sigma_i^2} \right] \tag{2}$$

The experimental error at each measurement point, σ_i, can be estimated as an additional model parameter or fixed to a known value. Here, we follow Sharp et al. (2022) and take σ_i fixed at a known constant value, although it is possible to include σ_i in the vector of unknown parameters θ. The maximum likelihood estimator (MLE) θ^*, and thus best-fit model parameters for the given experimental data ϕ, is defined by the solution of the inverse problem

$$\theta^* = \text{argmax}_{\theta \in \Omega} \mathcal{L}(\theta, \phi^*). \tag{3}$$

As the differential equations defining $y(\theta)$ rarely have explicit solutions, the likelihood (2) is difficult to evaluate analytically. It is therefore standard to minimize the negative log-likelihood $G(\theta, \phi) = -\log\left(\mathcal{L}(y(\theta), \phi^*)\right)$ given by

$$G(\theta, \phi) = \sum_{i=1}^{d} \log\left(\sqrt{2\pi \sigma_i^2}\right) + \frac{(y_i(\theta) - \phi_i^*)^2}{\sigma_i^2}. \tag{4}$$

Under the assumption that $\sigma_i = \sigma$ is fixed, the error term $\log\left(\sqrt{2\pi \sigma^2}\right)$ and denominator of $G(\theta, \phi)$ are constant and do not influence the solution of the optimization problem. The maximum likelihood estimator θ^* is the parameter set that minimizes $G(\theta, \phi^*)$. A number of computational techniques exist to minimize $G(\theta, \phi)$ and thus

calculate θ^*. These optimization techniques typically require simulating the mathematical model (1) at each optimization step. Further complicating the optimization problem, $G(\theta, \phi^*)$ is often non-convex with multiple local minima.

2.2 Continuation of Maximal Likelihood Estimator

Model fitting is increasingly performed concurrently with experiments (Luo et al. 2022) or obtained from an evolving real-world scenario, as in epidemic modelling (Sanche et al. 2020). In both of these cases, the calibration data ϕ evolves and should not be considered as known and constant. In (4), we explicitly wrote the objective function $G(\theta, \phi)$ as a function of the model parameters θ and the experimental data ϕ. The MLE θ^* is an implicit function of the experimental data ϕ defined by the solution of (3). We are interested in this implicit function $\theta^*(\phi)$. Most existing optimization techniques consider the calibration data fixed and omit this dependence.

Here, we develop a continuation type technique to compute the evolution of $\theta^*(\phi)$ numerically as a function of ϕ from an initial solution of the optimization problem. We calculate the evolution of $\theta^*(\phi)$ as the calibration data varies to generate a curve of potential MLEs in (ϕ, θ^*) space by developing a numerical continuation technique.

Numerical continuation methods are specialized numerical methods to compute branches of implicitly defined curves. A standard application of these continuation type techniques in mathematical biology is numerical bifurcation analysis (Dhooge et al. 2008; Sanche et al. 2022). These numerical bifurcation techniques compute equilibrium systems of a non-linear dynamical system as a function of model parameters but can be used to detect much richer dynamical behaviour (De Souza and Humphries 2019).

Often, continuation techniques leverage "predictor-corrector" algorithms. Predictor-corrector algorithms predict the solution to a non-linear system of equations using the implicit function theorem (IFT). The IFT is a crucial tool in numerical continuation as it maps a continuation condition to an implicitly defined multivariable function. The IFT states

Theorem 1 (Implicit function theorem) *Let $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function. Assume that*

$$F(x_0, y_0) = 0$$

where $x_0 \in \mathbb{R}^m$ and $y_0 \in \mathbb{R}^n$ and

$$\det\left(D_x F(x_0, y_0)\right) \neq 0$$

where $D_x F(x_0, y_0)$ is the $m \times m$ Jacobian matrix obtained by taking partial derivatives of F with respect to x at the point (x_0, y_0).

Then, there exists an open set $S \subset \mathbb{R}^n$ with $y_0 \in S$ and a curve $g(y)$ such that $F(g(y), y) = 0$ for all $y \in S$. Furthermore, $g(y)$ is continuously differentiable with

$$D_y g(y) = -[D_x F]^{-1} D_y F. \tag{5}$$

To illustrate how the IFT facilitates numerical continuation, assume that $F(x_0, y_0)$ satisfies the hypothesis of the theorem. Let y be the continuation parameter and we search for solutions of the continuation equation, $F(x, y) = 0$, for y in a neighborhood of y_0. The IFT ensures that, for a small perturbation $y_1 = y_0 + \Delta y$ of the continuation parameter, there exists a function $g(y)$ such that $F(g(y_1), y_1) = 0$. Calculating $g(y_1)$ comprises the continuation step of numerical continuation techniques (Meijer et al. 2012). In practice, the continuation step uses the initial solution $x_0 = g(y_0)$ to predict $\hat{x}_1 = g(y_0) + \xi \Delta y$ where ξ is tangent to the solution curve $g(y)$, although more complex approaches are possible (Meijer et al. 2012; Dhooge et al. 2008). The prediction, $\hat{x}_1$, is then used as a starting point to calculate $g(y_1)$ using standard root finding techniques during the correction step (Meijer et al. 2012).

Here, we develop a "prediction-correction" strategy to predict the behaviour of the solution $\theta^*(\phi)$ of the inverse problem (3) as a function of the data ϕ. We focus on the "predictor" step, as the corrector step, if necessary, can utilize existing numerical optimization techniques to calculate the MLE from the predicted MLE.

One of the major steps in developing a continuation method is properly defining the continuation equation. We recall that we are concerned with predicting the evolution of the MLE, which is defined as the minimizer of the log-likelihood. As the log-likelihood (4) is continuously differentiable, local optimal must satisfy

$$D_\theta G(\theta^*, \phi) = 0, \tag{6}$$

so we necessarily have

$$\theta^*(\phi) \in \{\theta \in \Omega | D_\theta G(\theta^*, \phi) = 0\}.$$

However, the optimality condition (6) is a necessary, but not sufficient, condition for θ^* to be a MLE. Models that are not structurally identifiable (Raue et al. 2014) have manifolds in parameter space on which this optimality constraint holds but are not necessarily MLEs. We discuss the relationship between our approach and profile likelihood classifications of structural identifiability in Appendix A.

Now, let θ_0^* be the MLE for calibration data ϕ_0. Further, let the Hessian $D_\theta^2 G(\theta, \phi)$ be invertible at $(\theta_0^*, \phi_0) \in \mathbb{R}^p \times \mathbb{R}^d$ and consider the function

$$D_\theta G(\theta^*, \phi) : \mathbb{R}^p \times \mathbb{R}^d \to \mathbb{R}^p.$$

We take (6) as the continuation equation. As θ_0^* is the MLE corresponding to the calibration data ϕ_0, (6) necessarily holds at (θ_0^*, ϕ_0). As we have assumed that the Hessian $D_\theta^2 G(\theta, \phi)$ is invertible, we can directly apply the IFT to determine a branch of solutions in (θ, ϕ) space of (6). The IFT ensures the existence of a function $\Psi(\phi)$ in a neighbourhood of ϕ_0 with $\Psi(\phi_0) = \theta^*(\phi_0)$ such that

$$D_\theta G(\Psi(\phi), \phi) = 0.$$

It is natural to consider $\Psi(\phi)$ as the predicted MLE $\theta^*(\phi)$ for ϕ in a neighbourhood of ϕ_0. However, computing $\Psi(\phi)$ analytically is functionally impossible. We therefore

expand $\Psi(\phi)$ as a function of the calibration data ϕ using Taylor series

$$\Psi(\phi + \Delta\phi) = \Psi(\phi) + D\Psi(\phi)\Delta\phi + \mathcal{O}(\Delta\phi^2). \tag{7}$$

where $\phi + \Delta\phi$ is the updated calibration data and the IFT ensures that the function $\Psi(\phi)$ is continuously differentiable. We calculate $D\Psi(\phi_0)$ to predict Ψ starting from the known solution $\Psi(\phi_0) = \theta_0^*$.

We use the explicit expression given in (5) to calculate $D\Psi(\phi_0)$. In the notation of the IFT, $F(\theta_0, \phi_0) = D_\theta G(\theta_0, \phi_0)$, so $F_\theta = D_\theta^2 G(\theta_0, \phi_0)$ and $F_\phi = D_{\theta,\phi}^2 G(\theta_0, \phi_0)$. Then, (5) directly implies that

$$D\Psi(\phi) = -\left[D_\theta^2 G(\Psi(\phi), \phi)\right]^{-1} D_{\theta,\phi}^2 G(\Psi(\phi), \phi). \tag{8}$$

We then use $D\Psi$ to evaluate (7) and thus perform the continuation step to approximate $\Psi(\phi + \Delta\phi)$.

2.3 Numerical Implementation

We now show how to calculate finite difference approximations to the derivatives included in (8). For θ_n denoting the n-th parameter, we calculate

$$\frac{\partial G(\theta, \phi)}{\partial \theta_n} = \sum_{i=1}^{d} 2\left(y_i(\theta) - \phi_i\right) \frac{\partial y_i(\theta)}{\partial \theta_n}$$

and so

$$\left[D_{\theta,\phi}^2 G(\Psi(\phi), \phi)\right]_{(n,i)} = -2\frac{\partial y_i(\theta)}{\partial \theta_n}. \tag{9}$$

The derivatives $\partial_{\theta_n} y_i(\theta)$ can be calculated through finite difference schemes (Zi 2011)

$$\frac{\partial y_i(\theta)}{\partial \theta_n} = \frac{y_i(\theta + \Delta\theta_n) - y_i(\theta - \Delta\theta_n)}{2\Delta\theta_n} + \mathcal{O}\left((\Delta\theta_n)^2\right),$$

where $\Delta\theta_n$ is a small perturbation in the n-th parameter. Computing $D_{\theta,\phi}^2 G(\Psi(\phi), \phi)$ requires $2p$ model simulations for p model parameters. We note that $\partial_{\theta_n} y_i(\theta)$ is commonly used to perform local sensitivity analysis (Li and Vu 2013) and that more accurate finite difference approximations, such as centered differences, can be used to calculate $D_{\theta,\phi}^2 G(\Psi(\phi), \phi)$.

Calculating the Hessian $D_\theta^2 G(\theta, \phi)$ via finite differences is simple to implement but computationally expensive due to the number of objective function evaluations. In the following examples, we use a finite difference scheme to calculate $D_\theta^2 G(\theta, \phi)$. We calculate the diagonal elements of $D_\theta^2 G(\theta, \phi)$ using forward second order differences and the off-diagonal terms by

$$\frac{\partial G(\theta, \phi)}{\partial \theta_i \partial \theta_j} = \left(\frac{1}{4(\Delta \theta_i)(\Delta \theta_j)} \right) \big[G(\theta + \Delta \theta_i + \Delta \theta_j, \phi) - G(\theta + \Delta \theta_i - \Delta \theta_j, \phi)$$
$$+ G(\theta - \Delta \theta_i + \Delta \theta_j, \phi) + G(\theta - \Delta \theta_i - \Delta \theta_j, \phi) \big] + \mathcal{O}\left((\Delta \theta_i)^2, (\Delta \theta_j)^2 \right).$$

Our computation of the Hessian requires $2p(p + 1)$ objective function evaluations.

We note that the Hessian, or the observed Fisher Information, is commonly used throughout parameter optimization algorithms. The Hessian is also used in other techniques such as profile likelihood calculations, estimates of the likelihood function, and classical sensitivity anaylsis. Consequently, computationally efficient techniques to calculate $D_\theta^2 G(\theta, \phi)$ have recently been developed (Stapor et al. 2018).

In fact, many gradient-based optimization techniques approximate the Hessian $D_{\theta, \theta}^2 G(\theta, \phi)$ at each iteration of solving an optimization problem (MATLAB 2017). For example, both `fmincon` and `fminunc` in MATLAB (2017) calculate $D_{\theta, \theta}^2 G(\theta, \phi)$ at each optimization step and print the pre-computed Hessian as an output of the optimizer. It is therefore possible, and efficient, to recycle this calculation when calculating an update to θ_0^* using (8).

In total, this numerical implementation of (8) requires $2p(p+2)$ model simulations. Finally, when evaluating (8), it is computationally more appropriate to solve the linear system of equations

$$D_\theta^2 G(\theta, \phi) D\Psi = -D_{\theta, \phi}^2 G(\Psi(\phi), \phi)$$

for the unknown $D\Psi$.

Code to implement this continuation technique is available at `https://github.com/ttcassid/MLE_Continuation`.

3 Results

3.1 Informing Experimental Design Through the Continuation Method

There are a number of existing techniques to study the relationship between model parameters and data. While our continuation technique focuses on the relationship between the MLE and the calibration data, it has many ties to other existing techniques. Here, we focus on using the explicit relationship between data and the MLE to suggest additional experimental measurements and thus leveraging the continuation method for experimental design. In Appendix A, we discuss how the continuation method relates to parameter identifiability as assessed by the profile likelihood and local sensitivity analysis.

In our derivation of $D\Psi$, we assumed that the Hessian matrix $D_\theta^2 G(\theta, \phi)$ was invertible. The Hessian gives the curvature of the loglikelihood at the MLE and is known as the observed Fisher information matrix $\mathcal{I}_{obs}$. The observed Fisher information is a local measurement in data space. Conversely, the expected Fisher information considers the entirety of data space for fixed model parameters θ. The expected Fisher information is obtained by taking the expectation of $D_\theta^2 G(\theta, \phi)$ over all possible experimental

measurements ϕ and is defined via

$$\mathcal{I} = \mathbb{E}\left[D_\theta^2 G(\theta, \phi) \right].$$

Many existing experimental design methods leverage the expected Fisher information matrix to minimize the covariance in model parameter estimates via the Cramér-Rao inequality. These experimental design techniques typically maximize some aspect, often the determinant, of the Fisher information matrix as a function of possible data to select the most informative calibration data set (Kreutz and Timmer 2009). From a geometric perspective, maximizing the determinant of the Fisher information matrix corresponds to minimizing the volume of the confidence ellipsoid engendered from the covariance matrix (Braniff et al. 2019).

In particular, Braniff et al. (2019) considered the case of bistable gene regulatory networks where the fold bifurcation and unstable manifold between stable equilibria complicates experimental design and parameter estimation. Sharp et al. (2022) considered an information-geometry perspective to propose the expected Fisher information matrix and resulting Riemannian manifold as a guide for data collection. As is often the case, both Sharp et al. (2022) and Braniff et al. (2019) used the expected Fisher information, which considers all possible calibration data via the expectation over ϕ. Here, we show how our approach complements the classical Fisher information approach to experimental design, albeit through a local measurement, in (θ, ϕ) space. We recall that

$$D\Psi \Delta\phi = -[\mathcal{I}_{obs}]^{-1} D_{\theta,\phi}^2 G(\Psi(\phi), \phi)\Delta\phi.$$

Now, if $D_{\theta,\phi}^2 G(\Psi(\phi), \phi)$ were the identity, then $D\Psi$ would correspond to the observed Fisher information approach to measuring uncertainty in MLE.

In the calculation of $D\Psi \Delta\phi$, the matrix $D_{\theta,\phi}^2 G(\Psi(\phi), \phi)$ maps perturbations in the calibration data $\Delta\phi$ through the curvature of the loglikelihood to changes in the MLE. Consequently, $D_{\theta,\phi}^2 G(\Psi(\phi), \phi)$ acts as a change of basis matrix from the space of calibration data to parameter space. Simply, $D_{\theta,\phi}^2 G(\Psi(\phi), \phi)\Delta\phi$ scales changes in the calibration data to the confidence ellipsoid in parameter space obtained from $[\mathcal{I}_{obs}]^{-1}$. Geometrically, if $D_\theta^2 G$ has eigenvalues λ_i with corresponding eigenvectors v_i, then choosing $\Delta\phi$ such that

$$v_i = D_{\theta,\phi}^2 G \Delta\phi$$

translates perturbations in calibration data to the corresponding eigenspace of the covariance matrix.

For example, the i-th column of $D\Psi$ maps perturbations of the i-th data point to changes in the MLE. Specifically, the sum

$$\frac{\Delta\theta^*}{\Delta\phi_k} = \sum_{k=1}^{p} |D\Psi_{k,j}|$$

measures the sensitivity of the MLE θ^* to perturbations in the k-th data point. Thus,

$$\|\mathrm{D}\Psi\|_1 = \max_{k=1,2,\ldots,p} \frac{\Delta\theta^*}{\Delta\phi_k}$$

and the most informative data point satisfies

$$l = \mathrm{argmax}_{k=1,2,\ldots,p} \frac{\Delta\theta^*}{\Delta\phi_k},$$

where informative is understood as the data point inducing the largest sensitivity in the MLE. As an extreme example, if

$$\frac{\Delta\theta^*}{\Delta\phi_n} = 0,$$

then perturbations in ϕ_n do not impact the MLE estimate, which implies complete insensitivity of the model fit to ϕ_n. This example corresponds to $\Delta\phi$ belonging to the kernel of the matrix $\mathrm{D}^2_{\theta,\phi}G$ since we have assumed that $\mathrm{D}^2_\theta G$ is invertible.

We can therefore utilize our analysis to identify which additional experimental measurements could increase confidence in model parameterization. Consider k additional measurements $\{\phi_{s,i} = y_{s,i}(\theta^*)\}_{i=1}^k$ taken directly from the model simulation at times $\{t_{s,i}\}_{i=1}^k$ where the subscript s indicates simulated data. Including $\{\phi_{s,i}\}$ in the objective function (4) does not change the MLE or objective value function as these simulated data exactly match the model values. However, $\|\mathrm{D}\Psi(\phi + \Delta\phi_{s,i})\|$ quantifies the sensitivity of the MLE to variability in the k simulated measurements. Accordingly, the measurement that maximizes $\|\mathrm{D}\Psi(\phi + \Delta\phi_{s,i})\|$ for a fixed perturbation size Δ is a good candidate for an additional experimental measurement to decrease parameter uncertainty.

3.2 Examples

The continuation framework derived earlier is applicable to a large variety of models throughout in the mathematical biology literature. To demonstrate the utility of the continuation method, we consider two examples from distinct fields and model formulations. Further, we show how the continuation framework can be leveraged to evaluate the robustness of model parameterizations or identify additional experimental measurements.

First, we consider a classical model of HIV-1 viral dynamics. This model has been used extensively to understand viral dynamics data (Perelson 2002) and the identifiability of model parameters was considered by Wu et al. (2008). In that work, Wu et al. (2008) used simulated data to validate their identifiability results; we follow Wu et al. (2008) and use simulated data to illustrate our approach to predicting the MLE to updated calibration data. We also show how the expression for $\mathrm{D}\Psi$ developed for our continuation method can be used to evaluate the robustness of model parameterizations. Quantifying the robustness of model parameterizations is particularly important

for practically unidentifiable models, such as the viral dynamics model considered here.

Next, we consider a mathematical model of phenotypic heterogeneity in non-small cell lung cancer (NSCLC) (Cassidy et al. 2021). This model is given by a system of two non-local, structured PDEs representing the density of drug-sensitive and drug-tolerant NSCLC cells. The PDE model is equivalent to a system of ODEs (see (Cassidy et al. 2021) and Appendix C for details). The parameters of the model were fit to in vitro NSCLC data taken from growth experiments in treated and untreated media (Craig et al. 2019; Cassidy et al. 2021). We use this example to demonstrate the effectiveness and computational efficiency of our method to predict the MLE. In addition, we also use the continuation method to suggest additional experimental measurements to increase confidence in model parameterization.

Parameter continuation in a viral dynamics model

The standard viral dynamics model has been extensively used to understand the dynamics of viral infection in HIV-1 (Perelson 2002; Hill et al. 2018). The model tracks the concentration of uninfected target cells, $T(t)$, infected cells $I(t)$, and free infectious virus $V(t)$. Here, we follow Wu et al. (2008) and consider a model of HIV-1 dynamics where the target cells are CD4$^+$ T-cells. These cells are produced at a constant rate λ and cleared linearly at rate d. Infection occurs at a rate β following contact between a target cell and infectious viral particle and infected cells are cleared at rate δ. Upon lysis, infected cells release N viral particles into the circulation and free virus is cleared at a constant rate c. The viral dynamics model is given by

$$
\left.
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dt}}T(t) &= \lambda - \beta T(t)V(t) - dT(t) \\
\frac{\mathrm{d}}{\mathrm{dt}}I(t) &= \beta T(t)V(t) - \delta I(t) \\
\frac{\mathrm{d}}{\mathrm{dt}}V(t) &= \delta N I(t) - cV(t).
\end{aligned}
\right\}
\tag{10}
$$

It is common to set $p = \delta N$ so the final equation for $V(t)$ becomes

$$
\frac{\mathrm{d}}{\mathrm{dt}}V(t) = pI(t) - cV(t),
$$

and the system (10) is equipped with initial conditions $T(0) = T_0$, $I(0) = I_0$, and $V(0) = V_0$. In typical clinical studies, temporal data is only collected for circulating free virus so the model output corresponding to the calibration measurements is

$$
y_i(\theta) = \log_{10}(V(t_i, \theta)),
$$

where using log_{10} measurements of viral load is standard in HIV studies.

During antiretroviral therapy (ART), the viral load may fall below the limit of detection of standard assays. While there are a number of techniques to account for

this censored data, we do not consider data collected during ART, so the objective function is given by the sum of squares error

$$G_{HIV}(\theta, \phi) = \sqrt{\sum_{i=1}^{n} \left(\log_{10}(V(t_i, \theta) - \log_{10}(\phi_i))\right)^2}. \tag{11}$$

Wu et al. (2008) characterized the identifiability of this model using a higher order derivative method. They found that, if the initial conditions of the model T_0, I_0, and V_0 are known, then all six model parameters $\theta = \{\beta, d, \delta, c, N, \lambda, \}$ are identifiable. To illustrate their results, they fixed $\theta = \{(2 \times 10^{-5}, 0.15, 0.55, 5.5, 900, 80\}$ and simulated the ODE model (10). They sampled the simulated viral load at 37 distinct time points and added noise ε_i sampled from a Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$ (Wu et al. 2008).

Here, we illustrate how model dynamics evolve during MLE continuation. We follow Wu et al. (2008) but consider a smaller subset of calibration data collected at time $t_i = \{0.4, 1, 8, 14, 20, 36, 46, 58\}$. We add noise ε_i^0 sampled from a Gaussian distribution with $\mu = 0$ and $\sigma^2 = 0.15$ so the initial calibration data is

$$\phi_i^0 = \log_{10}(V(t_i, \theta)) + \varepsilon_i^0.$$

We first fit the model to the simulated data ϕ_i^0 to obtain an initial MLE. We then generate 4 additional viral load time courses $\{\phi_i^j\}_{j=1}^4$ by

$$\phi_i^j = \phi_i^0 + h_{step}|\varepsilon_i^j|$$

for ε_i^j sampled from a Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$ and $h_{step} = \pm 0.1, \pm 0.2$. This collection of 4 data sets could feasibly represent experimental data measured from an increasingly large sample drawn from a population of HIV-1 positive individuals with population viral dynamic parameters given by $\theta = \{2 \times 10^{-5}, 0.15, 0.55, 5.5, 900, 80\}$. Here, we test the ability of our continuation technique to predict reasonable viral dynamic curves *without* refitting the data.

In Fig. 1A), we compute the predicted $\Psi(\phi^j)$ and plot the predicted model dynamics obtained from $\Psi(\phi^j)$ alongside the perturbed data ϕ^j for comparison. In Fig. 1B), we show the fit model predictions, corresponding to the calculated MLE, to the perturbed data. In each case, the viral dynamics show comparable model predictions for the fit and predicted model parameters demonstrating that our continuation method can successfully predict reasonable model simulations. In fact, the Bayesian Information Criteria (Kass and Raftery 1995) indicates no significant differences between the predicted and true MLE for all 4 data sets, which, as we are comparing fits of the same model, corresponds to no significant difference in the objective value function between the predicted and true fits. However, Fig. 1C) shows the significant computational improvement obtained by only calculating the continuation step rather than fitting all model parameters at each step. The predicted model dynamics track the true viral load trajectory.
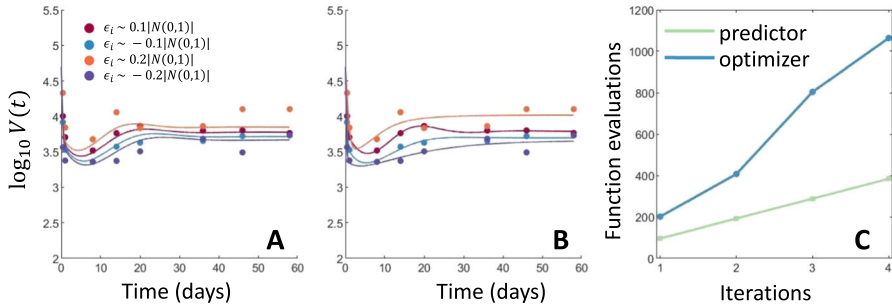
Fig. 1 Comparison of predicted model fits to randomly perturbed data. **A** and **B** show model trajectories obtained using predicted and fit model parameters to the simulated experimental data perturbed by $\phi_i^j = \phi_i^0 + h_{step}|\varepsilon_i^j|$. Panel **A** shows the predicted model fits to the experimental data while **B** shows the model fits to data resulting from the true MLE. **C** shows the number of objective value evaluations required to predict the MLE using this continuation technique or fit the model parameters to the perturbed data using the known parameters as a starting guess

It is common to find numerous local minima of (11) when fitting (10) to simulated data as the model is practically unidentifiable without precisely knowing the initial conditions. As measured by the value of the log-likelihood function or information criteria, these local minima can produce comparable fits to a given data set despite different dynamics. We perturbed the initial data set ϕ_0 by

$$\log(\phi_i^1) = \log(\phi_i^0) + 0.8\varepsilon_i$$

for ε_i sampled from a Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$. We fit this perturbed data from 10 distinct initial guesses using `fmincon` (MATLAB 2017). These 10 starting initial guesses converged to two local minima. We denote the corresponding parameter estimates by $\hat{\theta}_1$ and $\hat{\theta}_2$ and plot the resulting model trajectories in Fig. 2. These fits both accurately describe the viral load data and are indistinguishable by BIC. As we are comparing fits of the same model, the BIC corresponds to similar objective function values $G(\hat{\theta}_1, \phi)$ and $G(\hat{\theta}_2, \phi)$. Consequently, it is not obvious which of $\hat{\theta}_1$ and $\hat{\theta}_2$ best describe the data.

However, it is reasonable to expect that the MLE should be robust to small perturbations of the calibration data. We measure the robustness of each of these minima by calculating $\|D\Psi(\phi^1)\|$ at $\hat{\theta}_1$ and $\hat{\theta}_2$. While calculating $\|D\Psi(\phi^1)\|$ is not, strictly speaking, a continuation step, we note that

$$D\Psi\Delta\phi = \Psi(\phi + \Delta\phi) - \Psi(\phi) + \mathcal{O}\left(\Delta\phi^2\right).$$

Consequently, $\|D\Psi\|$ measures how robust the potential MLEs θ_i are to perturbations in the calibration data. As robustness to small perturbations in calibration data is a desirable attribute of the MLE, the continuation framework developed here can evaluate the robustness of potential MLEs, particularly in practically unidentifiable models.
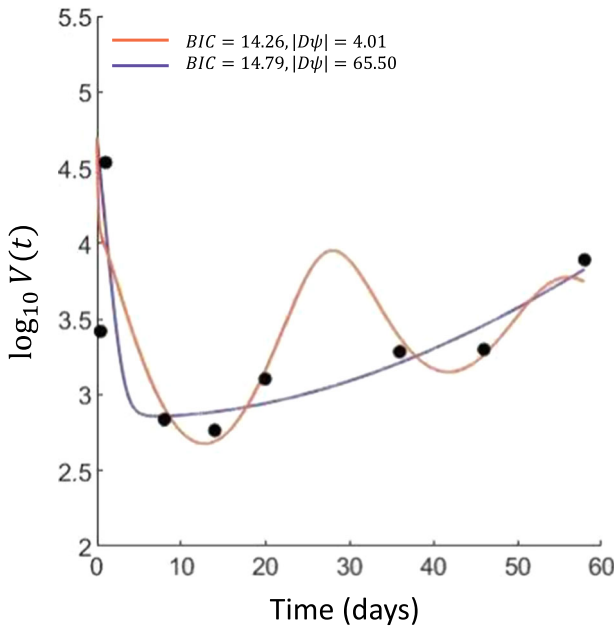
Fig. 2 Comparison of two potential fits to randomly perturbed viral dynamics models. Model trajectories obtained from two local minima from fitting 10 initial guesses to viral load data shown in black. Both trajectories accurately describe the viral load dynamics as evidenced by a small difference in BIC. However, the parameter estimate corresponding to the oscillatory trajectory is much more robust as measured by $\|D\Psi\|$ and is closer to the true underlying parameter set

Specifically, a smaller value of $\|D\Psi\|$ implies less sensitivity of the MLE to perturbations of the calibration data. For the example shown in Fig. 2, there is a 16 fold difference in sensitivity to calibration data. The oscillatory behavior present in Fig. 2 is somewhat surprising as there is no clear indication of oscilliations in the viral load data. However, the sparsity of sampling does not allow for either simulation to be excluded based on the current calibration data alone. Nevertheless, $D\Psi$ can be used to distinguish between these otherwise similar fits. We suggest that, when choosing between multiple fits with similar BIC values, the parameter estimate with the smaller sensitivity to the data is a more robust, and thus preferential, fit.

A PDE model of phenotypic switching in mathematical oncology

Non-genetic phenotypic heterogeneity has been increasingly studied as a driver of treatment resistance in solid cancers (Goldman et al. 2015) and a number of mathematical models have recently been developed (Gunnarsson et al. 2020; Jolly et al. 2018; Sahoo et al. 2021; Craig et al. 2019). We consider the Cassidy et al. (2021) model that tracks the density of NSCLC cells with a drug-sensitive ($A(t, a)$) or drug-tolerant ($B(t, a)$) phenotype at time t and age a through an age structured PDE. As mentioned, this PDE model can be reduced to the system of ODEs given in Appendix C. However, we present the simpler biological interpretation of the PDE model here. The total

number of cells of each phenotype is given by

$$\bar{A}(t) = \int_0^\infty A(t,a)\mathrm{d}a \quad \text{and} \quad \bar{B}(t) = \int_0^\infty B(t,a)\mathrm{d}a. \tag{12}$$

The total number of NSCLC cells is given by $N(t) = \bar{A}(t) + \bar{B}(t)$. Cassidy et al. (2021) considered logistic growth with an Allee effect, wherein cooperation between cells of the same phenotype can lead to increased growth rates, given by

$$R_A(\bar{A}(t), \bar{B}(t)) = r_A \left(1 - \frac{\bar{A}(t) + \bar{B}(t)}{K}\right) \quad \text{and}$$

$$R_B(\bar{A}(t), \bar{B}(t)) = r_B \left(1 - \frac{\bar{A}(t) + \bar{B}(t)}{K}\right) f_n(\bar{A}(t), \bar{B}(t)). \tag{13}$$

where r_A and r_B are phenotype specific growth rates, the carrying capacity is K, and the strength of the Allee effect is $f_n(\bar{A}(t), \bar{B}(t))$. We give the full details of $f_n(\bar{A}(t), \bar{B}(t))$ in Appendix C. Drug-tolerant cells have a constant death rate d_B while the death rate of drug-sensitive cells depends on the presence of anti-cancer treatment via

$$d_A = \begin{cases} d_A & \text{If untreated,} \\ d_A^{max} & \text{during treatment.} \end{cases}$$

$A(t,a)$ and $B(t,a)$ satisfy the age structured PDEs

$$\left.\begin{aligned} \partial_t A(t,a) + \partial_a A(t,a) &= -[d_A + R_A(\bar{A}(t), \bar{B}(t))]A(t,a) \\ \partial_t B(t,a) + \partial_a B(t,a) &= -[d_B + R_B(\bar{A}(t), \bar{B}(t))]B(t,a). \end{aligned}\right\} \tag{14}$$

We detail the corresponding boundary and initial conditions in Appendix C.

The model (14) was fit to in vitro experimental data collected in Craig et al. (2019) using the equivalent ODE formulation for $\bar{A}$ and $\bar{B}$. In Craig et al. (2019), NSCLC cell population growth was measured in untreated and treated environments. Anti-cancer drugs are applied from day 3 onwards during the treated experiment.

The model is fit to six total calibration data points taken from both the untreated, or control experiment, and the treated experiment. These six measurements correspond to 4 measurements taken at time $t_i = 0, 2, 4, 6$ days during the control experiment, denoted by $\{\phi_i\}_{i=1}^4$, and 2 measurements taken at time $t_i = 4, 6$ days during the treated experiment, denoted by $\{\phi_i\}_{i=5}^6$. We note that ϕ_3 and ϕ_5 were both collected on day 4 of the control and treated experiments, respectively, while ϕ_4 and ϕ_6 were both collected on day 6 of the control and treated experiments, respectively. As anti-cancer treatment is applied from day 3 on-wards of the treated experiment and decreases the cancer cell population, we necessarily have $\phi_5 \leqslant \phi_3$ and $\phi_6 \leqslant \phi_4$. We denote the experimental data used to parameterize the model by $\{\phi_i^0\}_{i=1}^6$. The model output corresponding to the experimental measurements is thus

$$y_i(\theta) = N(t_i, \theta),$$

and the objective function is the standard sum of squares error given by

$$G_{pheno}(\theta, \phi) = \sqrt{\sum_{i=1}^{6} \left(\log_{10}(N(t_i, \theta) - \log_{10}(\phi_i))\right)^2}.$$

Cassidy et al. (2021) fit model parameters $[r_A, r_B, d_A = d_B, d_A^{max}]$ to treated and untreated experimental data simultaneously for a number of cell lines. The MLE found by Cassidy et al. (2021) corresponds to $\theta^*(\phi^0) = [0.4827, 0.3498, 0.7025, 0.4198]$.

We perturbed the experimental data collected by Craig et al. (2019) with increasing amounts of Gaussian noise. We created 10 perturbed data sets $\{\phi_i^j\}_{i=1}^6$ where the index $j = 1, 2, ..., 10$, denotes the j-th perturbed data set and the normally distributed noise with $\mu = 0, \sigma^2 = 1$, and scaled such that

$$\| \log_{10}(\phi_i^j) - \log_{10}(\phi_i^*) \| = \left(0.05 + jh_{step}\right) \| \log_{10}(\phi_i^0) \|$$

where $h_{step} = 0.65/55$ was chosen such that $\| \log_{10}(\phi_i^{10}) - \log_{10}(\phi_i^0) \| = 0.75 \| \log_{10}(\phi_i^0) \|$.

We enforce that this randomly perturbed data satisfies $\phi_5 \leqslant \phi_3$ and $\phi_6 \leqslant \phi_4$. For each perturbed data set $\{\phi_i^j\}$, we use the continuation method to calculate

$$\Psi(\phi^j) = \theta^*(\phi^{j-1}) + D\Psi(\phi^{j-1})\Delta\phi + \mathcal{O}(\Delta\phi^2). \tag{15}$$

The naive approach to calculate the MLE $\theta^*(\phi^j)$ for updated data ϕ^j would be to use the MLE from the previous data, $\theta^*(\phi^{j-1})$, as an initial starting guess for the parameter fitting step. Hence, to illustrate the utility of our continuation technique, we calculated $\Psi(\phi^j)$ using (15) and then calculated $G_{pheno}(\Psi(\phi^j), \phi^j)$. We also calculated the true MLE $\theta^*(\phi^j)$ using the Matlab algorithm `fmincon` from both starting guesses $\Psi(\phi^j)$ and $\theta^*(\phi^{j-1})$. In Fig. 3A), we show the objective function value evaluated at the updated data ϕ^j and three parameter sets: the naive starting point, $\theta^*(\phi^{j-1})$; the predicted MLE, $\Psi(\phi^j)$; and the true MLE, $\theta^*(\phi^j)$.

The non-monotonic profile of the objective function G_{pheno} in Fig. 3A) is to be expected as we are adding noise to experimental data. This noise may perturb the existing data away from dynamics that can be well-described by the mathematical model. Accordingly, the important information from Fig. 3A) is the comparison

$$G_{pheno}(\theta^*(\phi^i), \phi^i) \leqslant G_{pheno}(\Psi(\phi^i), \phi^i) < G_{pheno}(\theta^*(\phi^{i-1}), \phi^i),$$

which demonstrates the accuracy of the continuation step (7) in driving a relative decrease in G_{pheno}.

Further, in Fig. 3B), we show the cumulative number of objective function evaluations when calculating $\theta^*(\phi^j)$ for $j = 1, 2, ..., 10$ when starting the optimization from $\theta^*(\phi^{j-1})$ and $\Psi(\phi^j)$. The total number of function evaluations used is lower when starting the optimization from the predicted MLE $\Psi(\phi^j)$ than when starting from $\theta^*(\phi^{j-1})$. More strikingly, the predicted MLE $G(\Psi(\phi^j), \phi^j)$ is comparable against
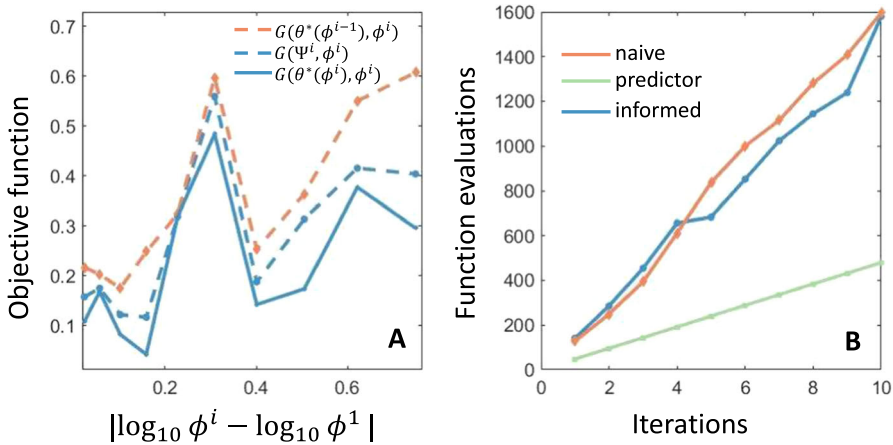
Fig. 3 Comparison between MLE estimates obtained using the naive and continuation approaches. **A** shows a comparison of the objective function value for the naive and continuation guesses as well as the true minimal objective function value as a function of the perturbation of the experimental data from the initial data. Here, the naive approach is shown in dashed orange, the predicted approach is shown in dashed blue, and the true minimal objective is in solid blue. **B** shows a comparison of the number of objective value evaluations required to obtain the minimal value when starting from the naive or predicted MLE with the number of function evaluations required to calculate $\Psi(\theta^i)$

$G(\theta^*(\phi^j, \phi^j))$ in Fig. 3A) and there is notable computational benefit to only calculating the predicted MLE $\Psi(\phi^j)$ rather than re-fitting the parameters. Taken together, the results shown in Fig. 3 demonstrate the accuracy and computation efficiency gained by calculating $\Psi(\phi^j)$.

We next utilize the continuation framework to identify additional time points to increase confidence in model parameters. We focus on the treated environment and consider additional time points $t_{s,i} = 3.1, 3.2, 3.3, 3.4, 3.5, 5, 7$ days with corresponding simulated measurements $\{\phi_{i,s}\}_{i=1}^{7} = N(t_{s,i})$. We perturb each of these simulated measurements by a fixed amount, $\Delta\phi = \pm 0.3 N(3.1)$, to give 14 additional, perturbed measurements. We appended each of these 14 measurements to the experimental data and predicted the MLE to these appended data sets.

We calculated the relative change in the MLE for each model parameter and each of the 14 appended data sets. Each of the simulated data point occurs following the beginning of therapy, although it would be simple to test other additional experimental measurements. The immediate decrease observed in $N(t)$ following the beginning of treatment is due to the death of sensitive cells following treatment administration and controlled by the parameter d_A^{max}. From the biological interpretation of the parameters, we expect d_A^{max} to be highly sensitive to perturbations in these data points.

As expected, d_A^{max} was the most sensitive model parameter to perturbations of the simulated data. We show the percent relative change in d_A^{max} from the unperturbed data in Fig. 4B. We plot the relative change of the other model parameters for the same perturbations in Appendix C. As expected, the maximal death rate of sensitive cells increased when the simulated data point was decreased from the true value and decreased when the simulated data point was increased.
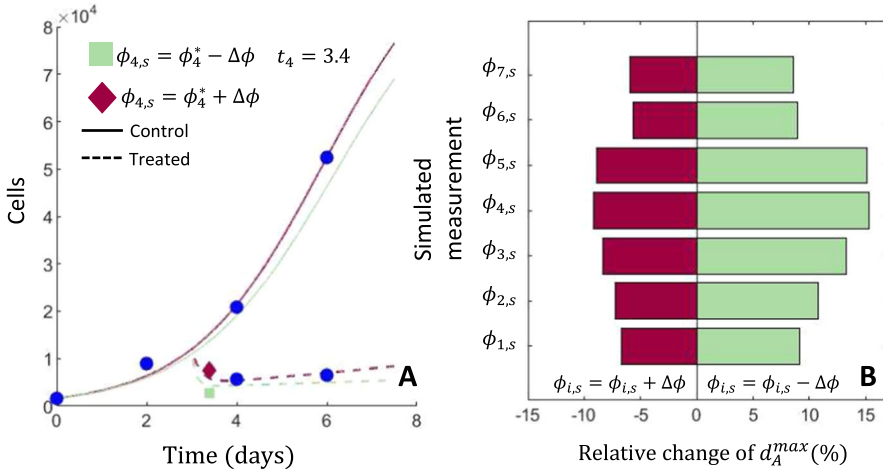
Fig. 4 Evaluating additional time points to identify d_A^{max} in an in vitro model of NSCLC. **A** shows the a selection of predicted model dynamics when fit to experimental data with a single additional time point $\phi_{i,s}^*$ that is perturbed by a $\Delta\phi$ from the true simulated value. For figure clarity, the model trajectories corresponding to the perturbation of $\{\phi_{4,s}\}$ are shown. Here, the solid line corresponds to the control experiment while the results of the treated experiment are plotted with a dashed line. **B** shows a tornado plot of the predicted relative change in the best-fit parameter d_A^{max} for each additional simulated data point $\{\phi_{i,s}\}_{i=1}^7$ where the perturbation $\Delta\phi > 0$. The left side of the tornado plot, in red, shows the relative change when the perturbed value $\phi_{i,s} = \phi_{i,s}^* + \Delta\phi$ is larger than the simulated value $\phi_{i,s}^*$. The right-hand side, in green, shows the relative change in d_a^{max} when $\phi_{i,s} = \phi_{i,s}^* - \Delta\phi$ is smaller than the simulated value $\phi_{i,s}^*$

The treatment sensitive population rapidly shrinks during therapy. The stabilization and rebound of the population during therapy is due to the expansion of the drug resistant population. This stabilization occurs once the drug sensitive population has been maximally suppressed which due to the drug effect. The most informative simulated data point, as measured by the magnitude of the relative change in the parameter d_A^{max}, was at time $t_{i,s} = 3.4$. At $t = 3.4$, drug sensitive cells are no longer dominant due to drug pressure. The depth of the population response to treatment, as measured by $N(3.4)$, is thus highly sensitive to death rate of these drug sensitive cells under treatment. However, expecting an additional experimental measurement to be made at precisely 3.4 days is unrealistic due to experimental constraints. However, Fig. 4B shows that measurements at $t_{s,3} = 3.3$ and $t_{s,3} = 3.5$ would also strongly inform d_A^{max}. Consequently, our conclusion that $t_{s,4} = 3.4$ is the most informative time for an additional experimental measurement is robust to the fact that an experimental measurement cannot be made at precisely 3.4 days. Our results indicate that including an additional experimental measurement in the 4.8 h window between 3.3−3.5 days will strongly inform d_A^{max}, which is experimentally feasible.

In Fig. 4A, we show the simulated experimental measurements and predicted model dynamics for the most informative time point. The predicted model simulations capture the perturbed data point while retaining good fits to the true experimental data.

4 Discussion

Parameter fitting is crucial step when using mathematical models to predict novel treatment strategies, extrapolate from clinical trials, identify new drug targets or schedules, or propose non-pharmaceutical interventions (Brady and Enderling 2019; Cassidy et al. 2020; Cassidy and Craig 2019). However, parameter fitting can be difficult and computationally expensive. A large variety of fitting techniques have therefore been developed to calibrate model predictions against data (Toni et al. 2009; Horbelt et al. 2002; Kreutz et al. 2013; Lauß et al. 2018). Moreover, mathematical modeling is increasingly applied to understand emerging data and make real-time predictions. In this case, as new data emerges, the model parameters must be refit with potential computational cost. Here, we developed a continuation type technique to quantify how updates to experimental data will impact the MLE and predict the evolution of the MLE as a function of the experimental data used to calibrate the model.

We used the IFT to calculate the trajectory of the MLE through parameter space. As the IFT only guarantees the existence of a differentiable trajectory Ψ through calibration data–parameter space, we utilized the first order Taylor expansion Ψ to extrapolate the evolution of the MLE due to changes in experimental data. We showed how this calculation is intrinsically linked to local sensitivity analysis and the curvature of the objective function. In two examples drawn from mathematical biology, we showed how this continuation technique can predict acceptable model fits to experimental data while significantly reducing computational overhead. In fact, in most applications, our continuation technique requires no dedicated computational overhead as the Hessian of the objective function is calculated at each step when using common optimization algorithms, such as `fmincon` (MATLAB 2017), and local sensitivity analysis is a standard step in model fitting.

Perhaps more importantly that gains in computational efficiency, our approach explicitly identifies relationships between individual experimental measurements and parameter estimates. Our approach addresses similar questions to local sensitivity analysis from a distinct perspective. Rather than using simulations to understand how small perturbations in model parameters from the best-fit parameters change model outputs as in standard sensitivity analysis, we quantify how changes in the training data impact the best-fit parameters and measure the sensitivity of the best-fit parameters to variations in this calibration data. As we showed in Sect. 2, this perspective can be used to suggest additional experimental measurements to increase confidence in model parameterization. Further, we showed how to use $D\Psi$ to understand which experimental measurements are most informative for model parameterizations and identify redundant measurements that do not provide additional information for parameter estimation.

Our technique is a type of local analysis that explores the functional dependence of the MLE on experimental data starting from a pre-identified MLE. Specifically, we assume that the Hessian of the objective function is invertible at the MLE and our results are necessarily local in parameter space as we are extrapolating from a pre-identified MLE. Nevertheless, our examples show the utility of our continuation approach for even large perturbations of the experimental data.

Despite these limitations, we developed a continuation-type technique to predict the functional dependence of a MLE on the experimental data used to train a mathematical model. While we have focused on applications in mathematical biology, our approach is immediately portable to other domains. As our method is independent of the number of data points, our approach could be particularly useful in big-data applications. Ultimately, our results offer a unified approach to quantify the relationship between training data and best-fit model parameters and to leverage this understanding to suggest additional experiments to increase confidence in model parameterization.

Data access statement The code and data underlying the results in this manuscript are available at `https://github.com/ttcassid/MLE_Continuation`.

Appendix A: Relationships Between MLE Continuation and Exisiting Techniques

A.1 Parameter Identifiability

In the main text, we have explicitly written the MLE estimator as a function of the experimental data used to fit a model. This approach is intrinsically related to parameter identifability analysis. Identifiability analysis attempts to determine if available experimental observations are capable to uniquely determine model parameters. Accordingly, the practical identifiability of a mathematical model depends on available experimental data. The *profile likelihood*, given by

$$PLE_{\theta_i}(c) = \min_{\theta_i = c, \theta \in \mathbb{R}^p} G(\theta, \phi),$$

and introduced by Raue et al. (2009), is a projection of the likelihood function onto the model parameter $\theta_i = c$. The profile likelihood illustrates the behaviour of the likelihood function as the parameter θ_i is fixed away from the optimal value θ_i^*. The shape of $PLE_{\theta_i}(c)$ illustrates the confidence interval of the parameter estimate θ_i^* for given experimental data. Formally, Raue et al. (2009) define these confidence intervals by

$$\text{C.I.}(\theta_i, \alpha) = \{c | PLE_{\theta_i}(c) - PLE_{\theta_i}(\theta_i^*) < \Delta_\alpha\}$$

where $\Delta_\alpha = \chi^2(\alpha, df)$ is the χ^2 distribution at significance level α and df degrees of freedom (Raue et al. 2009). A parameter is practically identifiable in the sense of Raue et al. (2009) with confidence level $1 - \alpha$ if C.I.(θ_i, α) is bounded in parameter space

for given experimental data. Conversely, a non-identifiable parameter has a profile likelihood that does not increase past the threshold Δ_α.

The profile likelihood is intrinsically linked to the available experimental data ϕ_i. We view the PLE as a function of both the parameter θ_i and the experimental data ϕ

$$PLE_{\theta_i}(c, \phi) = \min_{\theta_i=c, \theta \in \mathbb{R}^p} G(\theta, \phi).$$

For practically unidentifiable models, it is natural to ask what perturbations to the experimental data could render the model practically identifiable. Raue et al. (2009) use the profile likelihood of a model parameter to suggest additional experiments to resolve practical non-identifiability. They simulate the model for parameter values along PLE_{θ_i} to suggest additional experimental measurements at times $t_{s,i}$, where $t_{s,i}$ represents the i-th *simulated* measurement time. In our framework, we define

$$\theta^*|_{\theta_i=c}(\phi) = \mathrm{argmin}_{\theta_i=c, \theta \in \mathbb{R}^p} G(\theta, \phi),$$

so that

$$PLE_{\theta_i}(c, \phi) = G(\theta^*|_{\theta_i=c}(\phi), \phi).$$

We note that the definition of $\theta^*|_{\theta_i=c}(\phi)$ is precisely that of $\theta^*(\phi)$ with the added constraint that $\theta_i = c$. We can calculate $\mathrm{D}_\phi \theta^*|_{\theta_i=c}$ as a function of the experimental data ϕ in precisely the same manner as described previously. Consequently, our continuation approach can complement the experimental design approach suggested by Raue et al. (2009) by incorporating the sensitivity of the MLE to perturbations in the (simulated or experimental) calibration data.

A.2 Sensitivity analysis

Local sensitivity analysis quantifies how small perturbations of the best-fit parameters impact model output (Zi 2011). A standard approach to local sensitivity analysis is using the finite difference approximation of

$$S_n(t_i) = \frac{\partial y(\theta)}{\partial \theta_n} = \frac{h(t_i, \theta + \Delta \theta_n) - h(t_i, \theta - \Delta \theta_n)}{2\Delta \theta_n} + \mathcal{O}(\Delta \theta_n)$$

to identify which parameter values strongly impact model projections. When $|S_n|$ is small, the model output is considered to be insensitive to θ_n. The n-th row of $\mathrm{D}^2_{\theta, \phi} G(\Psi(\phi), \phi)$ is precisely $S_n(t_i)$ for t_i corresponding to calibration data measurements. When implementing (7), the magnitude of the continuation step $\mathrm{D}\Psi(\phi)\Delta\phi$ in the direction of θ_n is scaled by S_n. This scaling encodes the local sensitivity of model predictions to variations in parameters in the prediction of $\Psi(\phi)$. Consequently, our continuation method naturally includes the information gained from local sensitivity analysis.
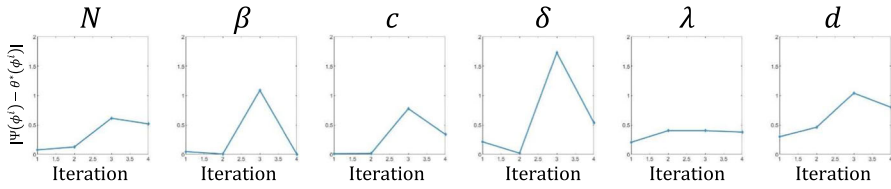
Fig. 5 Absolute error between the predicted model parameter obtained using the continuation method and best fit parameter obtained by fitting the data for the viral dynamics shown example in Fig. 1. Each panel shows the absolute error for each model parameter plotted against the iteration number

Appendix B: Viral Dynamics Model

In Fig. 5, we plot the absolute error between the model parameters obtained by our continuation method alongside the best fit parameters represented in Fig. 5. Here, we fit the natural log of the viral dynamics parameters. The predicted parameter value is identical to the MLE for some, although not all, iterations.

Appendix C: Structured PDE Model of Tumour Dynamics

Here, we give additional details regarding the phenotype PDE in (14). The proofs of these results and more information can be found in Cassidy et al. (2021).

C.1 Model Description

In Sect. 2, we considered a structured PDE model of cancer phenotypic plasticity developed in Cassidy et al. (2021). The PDE model tracks the age density of drug sensitive and drug tolerant cells through (14) where the variable a in (14) corresponds to chronological cellular age. The left-hand side of (14) is strictly negative, as is typical in age-structured populations models that distinguish between the loss of mother cells due to mitosis and the appearance of daughter cells with age $a = 0$ Perthame (2007); Cassidy et al. (2019). Specifically, cellular reproduction is included in the Cassidy et al. (2021) model via the boundary condition for $A(t, 0)$ and $B(t, 0)$. Cassidy et al. (2021) assumed that cells can change phenotype at birth and that older mother cells are more likely to produce daughter cells that change phenotype Arora et al. (2017); Uetake and Sluder (2010). Then, the boundary condition of (14) is given by

$$
\begin{aligned}
A(t, 0) &= 2\int_0^\infty \Big[R_A(\bar{A}(t), \bar{B}(t))\beta_{AA}(a)A(t, a) \\
&\quad + R_B(\bar{A}(t), \bar{B}(t))\beta_{BA}(a)B(t, a)\Big]\,\mathrm{d}a \\
B(t, 0) &= 2\int_0^\infty \Big[R_A(\bar{A}(t), \bar{B}(t))\beta_{AB}(a)A(t, a) \\
&\quad + R_B(\bar{A}(t), \bar{B}(t))\beta_{BB}(a)B(t, a)\Big]\,\mathrm{d}a.
\end{aligned}
$$

where

$$\beta_{ii}(a) = P_{ii}^* + (P_{ii}^{max} - P_{ii}^*) \exp[-\sigma_i a]$$

and $\beta_{ij}(a) = 1 - \beta_{ii}(a)$. These terms give the probability of a parent cell of phenotype i and age a will produce a daughter cell with phenotype j. We refer to Cassidy et al. (2021) for a discussion of the functional forms and parameterization of β_{ij}.

As mentioned in the Main Text, Cassidy et al. (2021) considered an Allee effect in the growth rate of drug tolerant cells. The function $f_n(\bar{A}(t), \bar{B}(t))$ is given by

$$
\begin{aligned}
f_n(\bar{A}(t), \bar{B}(t)) &= 1 + \left(\frac{r_A - r_B}{r_B}\right)\left(\frac{\bar{B}(t)^n}{\bar{A}(t)^n + \bar{B}(t)^n}\right) \\
&= 1 + \left(\frac{r_A - r_B}{r_B}\right)\left(\frac{\theta(t)^n}{1 + \theta(t)^n}\right) = f_n(\theta(t)) \quad \text{for} \quad \theta(t) = \bar{B}(t)/\bar{A}(t).
\end{aligned}
$$

When fitting (14) to data, we are primarily interested in the total number of cells, rather than their age density. Consequently, we focus on $\bar{A}(t)$ and $\bar{B}(t)$. Cassidy et al. (2021) derived the following system of ODEs for these quantities

$$
\left.
\begin{aligned}
\frac{d}{dt}\bar{A}(t) &= -[R_A(\bar{A}(t), \bar{B}(t)) + d_A]\bar{A}(t) \\
&\quad + 2R_A(\bar{A}(t), \bar{B}(t))N_{AA}(t) + 2R_B(\bar{A}(t), \bar{B}(t))\left[\bar{B}(t) - N_{BB}(t)\right] \\
\frac{d}{dt}\bar{B}(t) &= -[R_B(\bar{A}(t), \bar{B}(t) + d_B]\bar{B}(t) + 2R_A(\bar{A}(t), \bar{B}(t))\left(\bar{A}(t) - N_{AA}(t)\right) \\
&\quad + 2R_B(\bar{A}(t), \bar{B}(t))N_{BB}(t) \\
\frac{d}{dt}N_{AA}(t) &= P_{AA}^{max}\left[2R_A(\bar{A}(t), \bar{B}(t))N_{AA}(t) + 2R_B(\bar{A}(t), \bar{B}(t))\left(\bar{B}(t) - N_{BB}(t)\right)\right] \\
&\quad - \left(R_A(\bar{A}(t), \bar{B}(t)) + d_A\right)N_{AA}(t) + \sigma_A\left(P_A^*\bar{A}(t) - N_{AA}(t)\right) \\
\frac{d}{dt}N_{BB}(t) &= P_{BB}^{max}\left[2R_A(\bar{A}(t), \bar{B}(t))\left(\bar{A}(t) - N_{AA}(t)\right) + 2R_B(\bar{A}(t), \bar{B}(t))N_{BB}(t)\right] \\
&\quad - \left(R_B(\bar{A}(t), \bar{B}(t)) + d_B\right)N_{BB}(t) - \sigma_B N_{BB}(t) + \sigma_B P_B^*\bar{B}(t).
\end{aligned}
\right\}
$$

$$(16)$$

We use the system ODEs in (16) to simulate the model when performing MLE continuation. We use the initial conditions derived by Cassidy et al. (2021) to equip the initial value problem. For a population in stable exponential growth with growth rate λ_p, we have

$$\bar{A}(0) = \frac{A_0}{r_A + d_A + \lambda_p} \quad \text{and} \quad \bar{B}(0) = \frac{B_0}{r_B + d_B + \lambda_p}$$
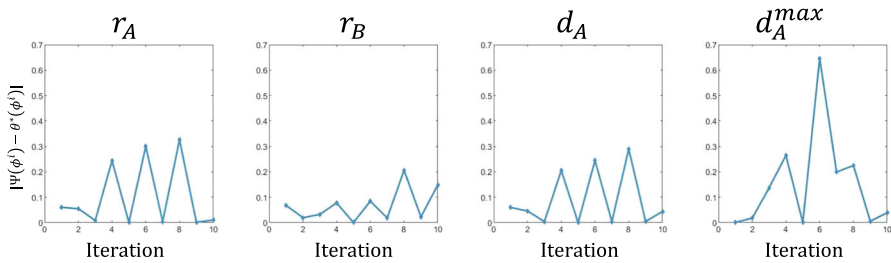
Fig. 6 Absolute error between the predicted model parameter obtained using the continuation method and best fit parameter obtained by fitting the data for the phenotype model example shown in Fig. 3. Each panel shows the absolute error for each model parameter plotted against the iteration number

where A_0 and B_0 are linked to the initial age distribution of drug-sensitive and drug-tolerant cells. It follows that

$$N_{AA}(0) = \int_0^\infty \left(P^*_{AA} + (P^{max}_{AA} - P^*_{AA})e^{-\sigma_A a} \right) A_0 \exp(-(r_A + d_A + \lambda_p)a)\mathrm{d}a$$
$$= A_0 \left(\frac{P^*_{AA}}{r_A + d_A + \lambda_p} + \frac{P^{max}_{AA} - P^*_{AA}}{r_A + d_A + \lambda_p + \sigma_A} \right),$$

and

$$N_{BB}(0) = B_0 \left(\frac{P^*_{BB}}{r_B + d_B + \lambda_p} + \frac{P^{max}_{BB} - P^*_{BB}}{r_B + d_B + \lambda_p + \sigma_B} \right).$$

C.2 Comparison Between Predicted and True MLE

In Fig. 3, we compared the objective value function corresponding to the naive, predicted, and fit MLE. Those results demonstrate that the predicted MLE produces similar model trajectories to the fit MLE. Here, we show a comparison between parameter estimates of the true and predicted MLE in Fig. 6. We note that, for some iterations, the predicted parameter value was identical to the parameter value found through optimization.

C.3 Experimental Design Effects on Other Model Parameters

As mentioned in the main text, the death rate of drug sensitive cells was the most sensitive parameter to perturbations of the simulated data. However, the other parameters were also predicted to change in response to perturbations of the simulated data. We show the relative changes for the other model parameters at each of the 7 simulated data points $\{\phi\}_{i=1}^7$ in Fig. 7. Unsurprisingly from the biological interpretation of the parameters, d_A^{max} is the most sensitive to the perturbed simulated data. However, there is a small increase in the growth rate of drug-tolerant cells, r_B, when the simulated data is perturbed by a negative amount. This can be understood by a fitness-increase of the drug-tolerant cells and thus stronger competition between the drug-tolerant and
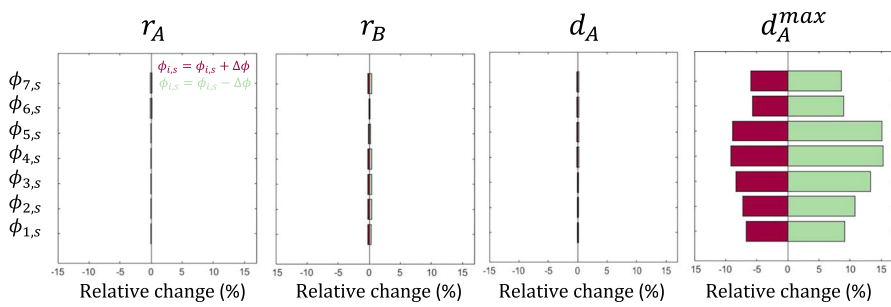
Fig. 7 Tornado plots of the predicted relative change in the best-fit parameters for each additional simulated data point. Each panel shows the predicted change to the model parameters r_a, r_b, d_a and d_a^{max} for the perturbed data points $\{\phi_{i,s}\}_{i=1}^{7}$ where the perturbation $\Delta\phi_{i,s} > 0$. In each tornado plot, the red bar shows the relative change in each model parameter when the perturbed value $\phi_{i,s} = \phi_{i,s}^* + \Delta\phi$ is larger than the simulated value $\phi_{i,s}^*$. The green bar shows the relative change in each model parameter when $\phi_{i,s} = \phi_{i,s}^* - \Delta\phi$ is smaller than the simulated value $\phi_{i,s}^*$

drug-sensitive cells. This competition decreases the population size prior to treatment which allows for a better fit to the perturbed data point post-treatment despite a larger resistant population.

References

Allen RJ, Rieger TR, Musante CJ (2016) Efficient generation and selection of virtual populations in quantitative systems pharmacology models. CPT Pharmacomet Syst Pharmacol 5(3):140–146. https://doi.org/10.1002/psp4.12063

Altrock PM, Liu LL, Michor F (2015) The mathematics of cancer: integrating quantitative models. Nat Rev Cancer 15(12):730–745. https://doi.org/10.1038/nrc4029

Arora M, Moser J, Phadke H, Basha AA, Spencer SL (2017) Endogenous replication stress in mother cells leads to quiescence of daughter cells. Cell Rep 19(7):1351–1364. https://doi.org/10.1016/j.celrep.2017.04.055

Brady R, Enderling H (2019) Mathematical models of cancer: when to predict novel therapies, and when not to. Bull Math Biol 81(10):3722–3731. https://doi.org/10.1007/s11538-019-00640-x

Braniff N, Scott M, Ingalls B (2019) Component characterization in a growth-dependent physiological context: optimal experimental design. Processes 7(1):52. https://doi.org/10.3390/pr7010052

Braniff N, Richards A, Ingalls B (2019) Optimal experimental design for a bistable gene regulatory network. IFAC-PapersOnLine 52(26):255–261. https://doi.org/10.1016/j.ifacol.2019.12.267

Cárdenas SD, Reznik CJ, Ranaweera R, Song F, Chung CH, Fertig EJ, Gevertz JL (2022) Model-informed experimental design recommendations for distinguishing intrinsic and acquired targeted therapeutic resistance in head and neck cancer. NPJ Syst Biol Appl 8(1):32. https://doi.org/10.1038/s41540-022-00244-7

Cassidy T, Craig M (2019) Determinants of combination GM-CSF immunotherapy and oncolytic virotherapy success identified through in silico treatment personalization. PLoS Comput Biol 15(11):1007495. https://doi.org/10.1371/journal.pcbi.1007495

Cassidy T, Craig M, Humphries AR (2019) Equivalences between age structured models and state dependent distributed delay differential equations. Math Biosci Eng 16(5):5419–5450. https://doi.org/10.3934/mbe.2019270

Cassidy T, Humphries AR, Craig M, Mackey MC (2020) Characterizing chemotherapy-induced neutropenia and monocytopenia through mathematical modelling. Bull Math Biol 82(8):104. https://doi.org/10.1007/s11538-020-00777-0

Cassidy T, Nichol D, Robertson-Tessi M, Craig M, Anderson ARA (2021) The role of memory in non-genetic inheritance and its impact on cancer treatment resistance. PLoS Comput Biol 17(8):1009348. https://doi.org/10.1371/journal.pcbi.1009348

Castro M, de Boer RJ (2020) Testing structural identifiability by a simple scaling method. PLoS Comput Biol 16(11):1008248. https://doi.org/10.1371/journal.pcbi.1008248

Craig M, Kaveh K, Woosley A, Brown AS, Goldman D, Eton E, Mehta RM, Dhawan A, Arai K, Rahman MM, Chen S, Nowak MA, Goldman A (2019) Cooperative adaptation to therapy (CAT) confers resistance in heterogeneous non-small cell lung cancer. PLoS Comput Biol 15(8):1007278. https://doi.org/10.1371/journal.pcbi.1007278

Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, O'Reilly K, Endo A, Hellewell J, Nightingale ES, Quilty BJ, Jarvis CI, Russell TW, Klepac P, Bosse NI, Funk S, Abbott S, Medley GF, Gibbs H, Pearson CAB, Flasche S, Jit M, Clifford S, Prem K, Diamond C, Emery J, Deol AK, Procter SR, Zandvoort K, Sun YF, Munday JD, Rosello A, Auzenbergs M, Knight G, Houben RMGJ, Liu Y (2020) Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Public Health 5(7):375–385. https://doi.org/10.1016/S2468-2667(20)30133-X

De Souza DC, Humphries AR (2019) Dynamics of a mathematical hematopoietic stem-cell population model. SIAM J Appl Dyn Syst 18(2):808–852. https://doi.org/10.1137/18M1165086

Dhooge A, Govaerts W, Kuznetsov YA, Meijer HGE, Sautois B (2008) New features of the software MatCont for bifurcation analysis of dynamical systems. Math Comput Model Dyn Syst 14(2):147–175. https://doi.org/10.1080/13873950701742754

Goldman A, Majumder B, Dhawan A, Ravi S, Goldman D, Kohandel M, Majumder PK, Sengupta S (2015) Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition. Nat Commun 6(1):6139. https://doi.org/10.1038/ncomms7139

Gunnarsson EB, De S, Leder K, Foo J (2020) Understanding the role of phenotypic switching in cancer drug resistance. J Theor Biol 490:110162. https://doi.org/10.1016/j.jtbi.2020.110162

Hill AL, Rosenbloom DIS, Nowak MA, Siliciano RF (2018) Insight into treatment of HIV infection from viral dynamics models. Immunol Rev 285(1):9–25. https://doi.org/10.1111/imr.12698

Holmdahl I, Buckee C (2020) Wrong but useful - what covid-19 epidemiologic models can and cannot tell us. N Engl J Med 383(4):303–305. https://doi.org/10.1056/NEJMp2016822

Horbelt W, Timmer J, Voss HU (2002) Parameter estimation in nonlinear delayed feedback systems from noisy data. Phys Lett A 299(5–6):513–521. https://doi.org/10.1016/S0375-9601(02)00748-X

Jenner AL, Cassidy T, Belaid K, Bourgeois-Daigneault M-C, Craig M (2021) In silico trials predict that combination strategies for enhancing vesicular stomatitis oncolytic virus are determined by tumor aggressivity. J Immunother Cancer 9(2):001387. https://doi.org/10.1136/jitc-2020-001387

Jolly MK, Kulkarni P, Weninger K, Orban J, Levine H (2018) Phenotypic plasticity, bet-hedging, and androgen independence in prostate cancer: role of non-genetic heterogeneity. Front Oncol 8(MAR):1–12. https://doi.org/10.3389/fonc.2018.00050

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):773–795. https://doi.org/10.1080/01621459.1995.10476572

Kreutz C, Timmer J (2009) Systems biology: experimental design. FEBS J 276(4):923–942. https://doi.org/10.1111/j.1742-4658.2008.06843.x

Kreutz C, Raue A, Timmer J (2012) Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Syst Biol. https://doi.org/10.1186/1752-0509-6-120

Kreutz C, Raue A, Kaschek D, Timmer J (2013) Profile likelihood in systems biology. FEBS J 280(11):2564–2571. https://doi.org/10.1111/febs.12276

Lauß T, Oberpeilsteiner S, Steiner W, Nachbagauer K (2018) The discrete adjoint method for parameter identification in multibody system dynamics. Multibody SysDyn 42(4):397–410. https://doi.org/10.1007/s11044-017-9600-9

Li P, Vu QD (2013) Identification of parameter correlations for parameter estimation in dynamic biological models. BMC Syst Biol 7(1):91. https://doi.org/10.1186/1752-0509-7-91

Li P, Vu QD (2015) A simple method for identifying parameter correlations in partially observed linear dynamic models. BMC Syst Biol 9(1):92. https://doi.org/10.1186/s12918-015-0234-3

Luo MC, Nikolopoulou E, Gevertz JL (2022) From fitting the average to fitting the individual: a cautionary tale for mathematical modelers. Front Oncol 12(April):8–11. https://doi.org/10.3389/fonc.2022.793908

Maiwald T, Hass H, Steiert B, Vanlier J, Engesser R, Raue A, Kipkeew F, Bock HH, Kaschek D, Kreutz C, Timmer J (2016) Driving the model to its limit: profile likelihood based model reduction. PLoS ONE 11(9):1–18. https://doi.org/10.1371/journal.pone.0162366

Marino S, Hogue IB, Ray CJ, Kirschner DE (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. J Theor Biol 254:178–196. https://doi.org/10.1016/j.jtbi.2008.04.011

MATLAB (2017) R2017a. The MathWorks Inc., Natick, Massachusetts

Meijer H, Dercole F, Oldeman B (2012) Numerical bifurcation analysis. In: Mathematics of complexity and dynamical systems, Meyers, r edn. Springer, New York, pp 6329–6352. https://doi.org/10.1007/978-1-4614-1806-1_71

Perelson AS (2002) Modelling viral and immune system dynamics. Nat Rev Immunol 2(1):28–36. https://doi.org/10.1038/nri700

Perthame B (2007) Transport equations in biology. Front Math. https://doi.org/10.1007/978-3-7643-7842-4

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25(15):1923–1929. https://doi.org/10.1093/bioinformatics/btp358

Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J (2014) Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30(10):1440–1448. https://doi.org/10.1093/bioinformatics/btu006

Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, Kumar Jolly M (2021) A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. BioRxiv. https://doi.org/10.1101/2021.03.14.435359

Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R (2020) High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 26(7):1470–1477. https://doi.org/10.3201/eid2607.200282

Sanche S, Cassidy T, Chu P, Perelson AS, Ribeiro RM, Ke R (2022) A simple model of COVID-19 explains disease severity and the effect of treatments. Sci Rep 12(1):14210. https://doi.org/10.1038/s41598-022-18244-2

Sharp JA, Browning AP, Burrage K, Simpson MJ (2022) Parameter estimation and uncertainty quantification using information geometry. J R Soc Interface. https://doi.org/10.1098/rsif.2021.0940

Silk D, Kirk PDW, Barnes CP, Toni T, Stumpf MPH (2014) Model selection in systems biology depends on experimental design. PLoS Comput Biol 10(6):1003650. https://doi.org/10.1371/journal.pcbi.1003650

Stapor P, Fröhlich F, Hasenauer J (2018) Optimization and profile calculation of ODE models using second order adjoint sensitivity analysis. Bioinformatics 34(13):151–159. https://doi.org/10.1093/bioinformatics/bty230

Thompson RN (2020) Epidemiological models are important tools for guiding COVID-19 interventions. BMC Med 18(1):152. https://doi.org/10.1186/s12916-020-01628-4

Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface 6(31):187–202. https://doi.org/10.1098/rsif.2008.0172. arXiv:0901.1925

Uetake Y, Sluder G (2010) Prolonged prometaphase blocks daughter cell proliferation despite normal completion of mitosis. Curr Biol 20(18):1666–1671. https://doi.org/10.1016/j.cub.2010.08.018

Wu H, Zhu H, Miao H, Perelson AS (2008) Parameter identifiability and estimation of HIV/AIDS dynamic models. Bull Math Biol 70(3):785–799. https://doi.org/10.1007/s11538-007-9279-9

Zhang J, Cunningham J, Brown J, Gatenby R (2022) Evolution-based mathematical models significantly prolong response to abiraterone in metastatic castrate-resistant prostate cancer and identify strategies to further improve outcomes. eLife 11:1–105. https://doi.org/10.7554/eLife.76284

Zi Z (2011) Sensitivity analysis approaches applied to systems biology models. IET Syst Biol 5(6):336–346. https://doi.org/10.1049/iet-syb.2011.0015