INTERFACE

Research

THE ROYAL SOCIETY
PUBLISHING

Computationally efficient framework for diagnosing, understanding and predicting biphasic population growth

Ryan J. Murphy[1], Oliver J. Maclaren[2], Alivia R. Calabrese[3], Patrick B. Thomas[3], David J. Warne[1], Elizabeth D. Williams[3] and Matthew J. Simpson[1]

[1]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia
[2]Department of Engineering Science, University of Auckland, Auckland, New Zealand
[3]Queensland Bladder Cancer Initiative and School of Biomedical Sciences, Faculty of Health, Queensland University of Technology at Translational Research Institute, Brisbane, Australia

RJM, 0000-0002-9844-6734; ARC, 0000-0003-4837-379X; PBT, 0000-0002-3649-7923; DJW, 0000-0002-9225-175X; EDW, 0000-0002-3364-6655; MJS, 0000-0001-6254-313X

Throughout the life sciences, biological populations undergo multiple phases of growth, often referred to as *biphasic growth* for the commonly encountered situation involving two phases. Biphasic population growth occurs over a massive range of spatial and temporal scales, ranging from microscopic growth of tumours over several days, to decades-long regrowth of corals in coral reefs that can extend for hundreds of kilometres. Different mathematical models and statistical methods are used to diagnose, understand and predict biphasic growth. Common approaches can lead to inaccurate predictions of future growth that may result in inappropriate management and intervention strategies being implemented. Here, we develop a very general computationally efficient framework, based on profile likelihood analysis, for diagnosing, understanding and predicting biphasic population growth. The two key components of the framework are as follows: (i) an efficient method to form approximate confidence intervals for the change point of the growth dynamics and model parameters and (ii) parameter-wise profile predictions that systematically reveal the influence of individual model parameters on predictions. To illustrate our framework we explore real-world case studies across the life sciences.

1. Introduction

Quantifying population growth, whether it be the total number of individuals in a group or the total area covered by a species, has motivated the development of a range of mathematical models [1–3]. Here, we focus on populations that undergo two phases of growth, often called *biphasic growth*. Biphasic growth is prevalent across a wide range of applications in the life sciences, including ecological applications, for example coral reef growth after a disturbance (figure 1*a*) [6]; two-dimensional cell biology assays, for example proliferation and scratch-wound assays (figure 1*b*) [7,8]; three-dimensional cancer tumour spheroid cell biology experiments (figure 1*c*) [5,9]; decay of pathogens [10]; and bacterial dynamics [11,12]. Given the wide range of applications, a variety of mathematical and statistical methods have been developed in different disciplines to understand specific cases of biphasic population growth. Here, we develop a new computationally efficient general framework for diagnosing, understanding and predicting biphasic population growth that is broadly applicable across the life sciences. The approach, based on profile likelihood analysis in combination with parameter-wise profile predictions, enhances the accuracy and reliability of previous methods. These improvements
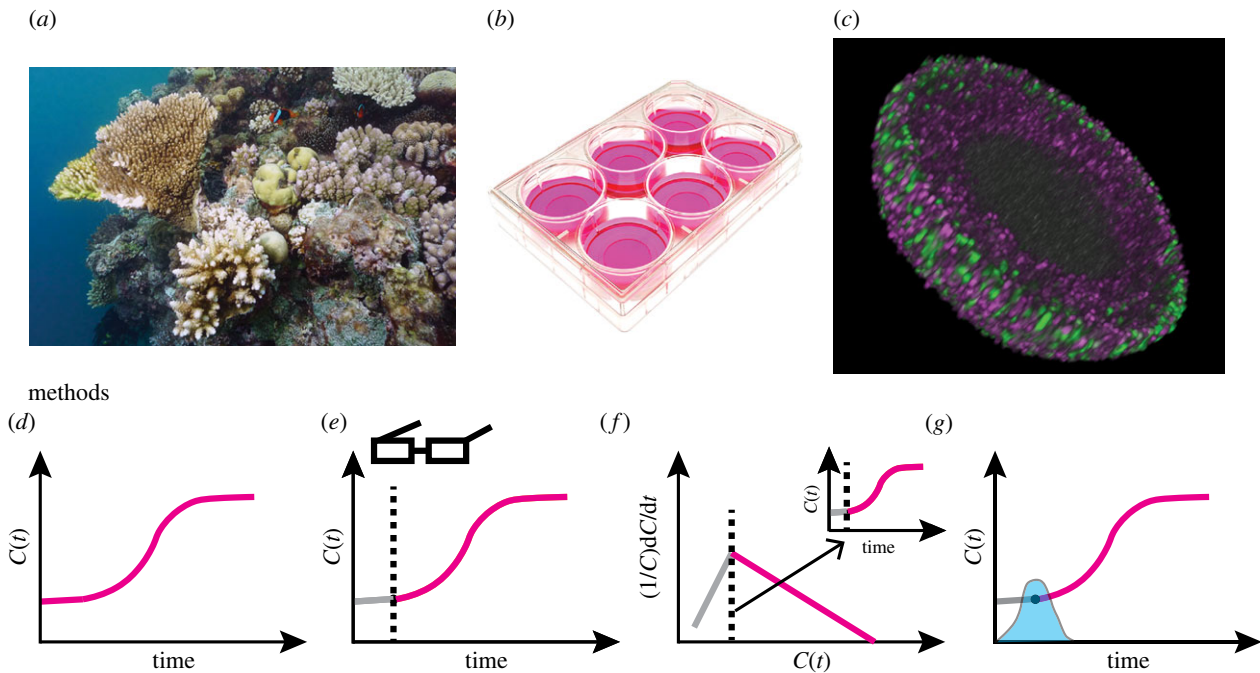
Figure 1. Case studies and methods for diagnosing, understanding and predicting biphasic population growth in the life sciences: (*a*) coral reef growth after a disturbance (reproduced from [4] (CC BY 4.0)); (*b*) two-dimensional cell proliferation assays; and (*c*) a three-dimensional tumour spheroid experiment (reproduced from [5] (CC BY 4.0)). Existing methods to analyse biphasic population growth and the temporal evolution of a population $C(t)$ range between: (*d*) overlooking the first phase (pink); (*e*) manually identifying the change point through visual inspection (first phase (grey), point estimate of change point (black-dashed), second phase (pink)); and (*f*) analysing and identifying a change point in *per capita* growth rate data, $1/C(t) \, dC(t)/dt$ against $C(t)$, and mapping this to a change point in time (first phase (grey), point estimate of change point (black-dashed) and second phase (pink)). (*g*) In this study, we form a profile likelihood for the change point (blue shaded) and each model parameter. Using the profile likelihoods, we estimate approximate confidence intervals. To quantify and visualize how variations in a model parameter influence predictions, we use parameter-wise profile predictions.

enable greater understanding of population growth dynamics and assist decision-making.

As biphasic population growth occurs across a wide range of applications and disciplines, different terminology is used to describe similar phenomena. A key term we refer to is the *change point*, which is the time at which the growth dynamics switches from the first phase to the second phase. In cell biology and ecological applications, the first phase is sometimes referred to as a lag, delay, adaptation or settling phase and the change point is sometimes referred to as the end of those respective phases or the start of the growth phase [6–8]. Change point detection has a long history, with applications in signal analysis and econometrics [13–15], and standard tools have been developed in software such as Matlab [16]. However, such tools and approaches typically do not incorporate a mechanistic model. In contrast, mechanistic model-based approaches usually assume a specific model or do not provide a systematic statistical framework to assess uncertainty in change point estimates. Here, we aim to bridge this gap by developing a general differential equation-based framework that does not rely on a specific model form while also providing systematic statistical uncertainty quantification.

Existing methods to analyse biphasic population growth vary in terms of simplicity, accuracy and reliability. The simplest method to interpret biphasic population growth is to overlook the two phases and analyse the experimental data with a single-phase model (figure 1*d*) [6,17–19]. Other approaches explicitly account for the existence of the two phases of growth and identify the change point manually through visual inspection (figure 1*e*) [5,7,9]. More sophisticated methods involve seeking statistical point estimates of the change point. In econometrics, this is sometimes referred

to as a regression discontinuity study or two-segment regression with change point detection [13,14]. Recent studies explore noisy *per capita* growth rate data to identify a change point in time (figure 1*f*) [6,8]. In another recent study examining biphasic growth of individual fish [20], profile likelihood analysis is used to form an approximate confidence interval for the change point, albeit for a specific mathematical model only (figure 1*g*).

When using differential equations to describe and interpret data, one should consider whether model parameters are identifiable. Many studies focus on the formal question of *structural identifiability*, namely, whether parameters of the mathematical model be uniquely identified given a set of continuous noise-free observations [21–23]. Such analysis can be performed using software tools such as DAISY [24] or GenSSI [25]. However, such tools focus on differential equations that are described by smooth functions and do not apply to biphasic growth models that are defined piecewise. Here, we focus on *practical identifiability*, namely, whether given a finite set of noisy experimental data can we uniquely identify model parameters. Profile likelihood analysis is one approach to assess parameter identifiability [22,26–32]. We choose to base our framework on profile likelihood analysis for two key reasons: (i) computational efficiency in comparison with other standard approaches [33] and (ii) to introduce new parameter-wise profile predictions to quantify and visualize how variations in a model parameter influence predictions of population growth trajectories. Alternative approaches to assess parameter identifiability include Markov chain Monte Carlo techniques [34–36].

To illustrate our framework, we explore four case studies across the life sciences: (i) coral reef regrowth after a disturbance; (ii) two different examples of two-dimensional cell

proliferation assays and (iii) a three-dimensional cancer tumour spheroid experiment. In §2, we describe the various experimental and field-scale datasets. In §§3–5, we detail the mathematical model, techniques for parameter estimation, practical identifiability analysis and prediction intervals, including parameter-wise profile predictions. In §6, we apply our framework, and in §7, we discuss insights that are gained by using this new framework.

2. Data

In this section, we describe the data used in this study. Since we deal with two different proliferation assay experiments, we present one of these cases, based on a bladder cancer cell line, in electronic supplementary material, F.

2.1. Coral reef growth after disturbance

Coral reef data analysed in this study are published in [6,37] and are part of the Australian Institute of Marine Science's Long Term Monitoring Program. The data describe the temporal evolution of the percentage coral cover following a major storm disturbance event (19 November 2008 to 18 September 2018) at Broomfield Island located within the Great Barrier Reef, Australia (figure 2*a*).

2.2. Two-dimensional cell proliferation assay

This dataset is obtained from an *in vitro* cell proliferation assay performed in [7]. A freshly prepared flask is placed in an incubator on a microscopic stage, and the number of cells are observed as they divide to form a confluent monolayer. The experiment is performed on tissue culture plastic with NIH-3T3 fibroblast cells for 120 hours (5 days). Experimental measurements are normalized using the mean maximum cell density such that the normalized cell density ranges from zero to unity.

2.3. Three-dimensional cancer tumour spheroid experiment

This dataset is obtained from tumour spheroid experiments we performed in [5,9]. The experiment is performed for 432 hours (18 days) with human melanoma WM983b spheroids formed with 5000 cells per well in a 96-well plate. Top-down area measurements of the spheroid are obtained using automated brightfield imaging and processing with the IncuCyte S3 live cell imaging system (Sartorius, Goettingen, Germany) (electronic supplementary material, table S1). Images are captured every 2 hours for the first 2 days and every 6 hours for the remainder of the experiment. In the first phase, the cells in the well migrate and adhere to form a shrinking spheroid. In the second phase, the spheroid grows as a solid mass. We quantify both phases by estimating the area enclosed by a projection of the spheroids, A, and, assuming a spherical geometry, convert these estimates into an equivalent radius ($r = \sqrt{A/\pi}$).

3. Mathematical model

3.1. Process model

Let $C(t)$ denote the variable of interest: for coral reef data, this is coral cover percentage [6]; for two-dimensional cell proliferation assays, this is the normalized cell density [7]; and for three-dimensional tumour spheroid experiments, this is the tumour spheroid radius [5,9]. To describe the population dynamics, we prescribe a biphasic mathematical model,

$$\frac{dC(t)}{dt} = \begin{cases} f_1(C), & t \le T, \\ f_2(C), & t > T, \end{cases} \tag{3.1}$$

where $f_1(C)$ and $f_2(C)$ describe the time rate of change of $C(t)$ before and after the change point, $t = T$, respectively. This framework is very general and can be used to describe several phenomena depending on how we specify $f_1(C)$ and $f_2(C)$. For example, if there is no growth or decay before $t = T$ and logistic growth for $t > T$, we set $f_1(C) = 0$ and $f_2(C) = r\,C\,(1 - C/K)$, where $r > 0$ is the growth rate and $K > 0$ is the long-time carrying capacity. For this application, we have four unknown parameters, i.e. a vector (r, K, $C(0)$, T), that we will estimate from data. For this particular choice of $f_1(C)$ and $f_2(C)$, we can solve the model exactly to give $C(t) = C(0)$ for $t \le T$ and $C(t) = KC(0)/[C(0) + (K - C(0))\exp(-r(t - T))]$ for $t > T$. Although, in principle, we can solve for $C(t)$ exactly for certain choices of $f_1(C)$ and $f_2(C)$, all results presented in this work involve solving the mathematical model numerically using a second-order explicit Runge–Kutta method that means that we do not have to rely on integrating equation (3.1) to obtain a closed-form solution.

3.2. Observation model

We assume that observed data y_i^o are measured at I discrete times, t_i, for $i = 1, 2, 3, \ldots, I$. We use a superscript 'o' to distinguish the noisy observed data from the model predictions. The model predictions are denoted by $y_i(r, K, C(0), T) = C(t_i \mid r, K, C(0), T)$. We collect the (noisy) data into a vector denoted by $y_{1:I}^o$. Similarly the process model solution is denoted by $y_{1:I}(r, K, C(0), T)$ for the vector of grid point values and by $y(r, K, C(0), T)$ for the full model trajectory over the time interval of interest. We estimate the process model parameter vector ($r, K, C(0), T$) by assuming that the observed data are noisy versions of the model solutions of the form $y_i^o \mid r, K, C(0), T, \sigma^2 \sim \mathcal{N}(y_i(r, K, C(0), T), \sigma^2)$. This means we assume that the observation errors are independent, identically distributed, additive and normally distributed with zero mean and constant variance σ^2. Different error models could be used within our likelihood-based framework if that data suggested that the normal error model was inappropriate [26]. Here, the constant variance will be estimated along with the process model parameters.

4. Parameter estimation

We hence combine both the process model parameter vector ($r, K, C(0), T$) and the observation parameter σ^2 into an overall vector parameter $\theta = (r, K, C(0), T, \sigma^2)$. We can then consider scalar or vector sub-parameters as interest parameters defined as functions of the full vector parameter, e.g. $\sigma^2 = \sigma^2(\theta)$, where we use the same symbol for the function and its value. The process model solution is itself an interest parameter in this sense and does not depend on the variance, i.e. $y_i(\theta) = y_i(r, K, C(0), T, \sigma^2) = y_i(r, K, C(0), T)$. Putting these elements together, we hence write our model for the data given the full parameter compactly as follows:

$$y_i^o \mid \theta \sim \mathcal{N}(y_i(\theta), \sigma^2(\theta)). \tag{4.1}$$

Taking a likelihood-based approach to parameter inference and uncertainty quantification, given a time series of observations together with our assumptions about the process and noise models, the log-likelihood function is given as follows:

$$\ell(\theta \mid y_{1:I}^{\mathrm{o}}) = \sum_{i=1}^{I} \log[\phi(y_i^{\mathrm{o}}; y_i(\theta), \sigma^2(\theta))], \qquad (4.2)$$

where $\phi(x; \mu, \sigma^2)$ denotes a Gaussian probability density function with mean μ and variance σ^2. Maximum likelihood estimation (MLE) provides an estimate of θ that gives the best match (in the sense of highest likelihood) to the data. The MLE is given by

$$\hat{\theta} = \sup_{\theta}[\ell(\theta \mid y_{1:I}^{\mathrm{o}})], \qquad (4.3)$$

subject to bound constraints. The procedure for estimating $\hat{\theta}$ involves numerical maximization of the log-likelihood, which can be achieved using many different algorithms. In this work, we find that a local optimization routine from the open-source `NLopt` optimization package in Julia performs well [38]. In particular, we use the Nelder–Mead optimization routine within the `NLopt` with the default stopping criteria.

5. Practical identifiability analysis and profile predictions

We use a profile likelihood-based approach to explore practical identifiability by working with a normalized log-likelihood function

$$\hat{\ell}(\theta \mid y_{1:I}^{\mathrm{o}}) = \ell(\theta \mid y_{1:I}^{\mathrm{o}}) - \sup_{\theta} \ell(\theta \mid y_{1:I}^{\mathrm{o}}), \qquad (5.1)$$

which we consider as a function of θ for a fixed set of data $y_{1:I}^{\mathrm{o}}$. Note that normalizing the log-likelihood means that we have $\hat{\ell}(\hat{\theta} \mid y_{1:I}^{\mathrm{o}}) = 0$.

5.1. Profile likelihood for interest parameters

Assuming the full parameter θ can be partitioned into an interest parameter ψ and nuisance parameter λ, where one or both of these may be vector valued in general, we write $\theta = (\psi, \lambda)$. More generally we can consider an interest parameter as any well-defined function of the full parameter, $\psi = \psi(\theta)$, and leave the implied nuisance parameter implicit (that this always exists in the appropriate sense is implied by the results in [39]). For a set of data, $y_{1:I}^{\mathrm{o}}$, the profile log-likelihood for the interest parameter ψ given a partition (ψ, λ) is defined as follows [26,40]:

$$\hat{\ell}_p(\psi \mid y_{1:I}^{\mathrm{o}}) = \sup_{\lambda \mid \psi} \hat{\ell}(\psi, \lambda \mid y_{1:I}^{\mathrm{o}}), \qquad (5.2)$$

which indicates that λ is optimized out for each fixed value of ψ. This implicitly defines a function $\lambda^*(\psi)$ of optimal values of λ for each value of ψ. In the case of an interest parameter given as a general function of the full parameter, the profile (or induced) log-likelihood is defined in terms of the constrained optimization problem [41,42],

$$\hat{\ell}_p(\psi \mid y_{1:I}^{\mathrm{o}}) = \sup_{\theta \mid \psi(\theta) = \psi} \hat{\ell}(\theta \mid y_{1:I}^{\mathrm{o}}), \qquad (5.3)$$

in which the 'nuisance degrees of freedom' in θ, after fixing ψ, are optimized out. As a concrete demonstration, consider the example in §§3.1–3.2, where we had $f_1(C) = 0$ and $f_2(C) = r\,C(1 - C/K)$, and the full parameter vector is $\theta = (r, K, C(0), T, \sigma^2)$. If we wish to profile the change point T, then we have $\psi(\theta) = T$ and $\lambda(\theta) = (r, K, C(0), \sigma^2)$ so that

$$\begin{aligned}
\hat{\ell}_p(T = t \mid y_{1:I}^{\mathrm{o}}) &= \sup_{(r,K,C(0),T,\sigma^2) \mid T = t} \hat{\ell}(r, K, C(0), T, \sigma^2 \mid y_{1:I}^{\mathrm{o}}) \\
&= \sup_{(r,K,C(0),\sigma^2) \mid T = t} \hat{\ell}(r, K, C(0), T, \sigma^2 \mid y_{1:I}^{\mathrm{o}}).
\end{aligned} \qquad (5.4)$$

In all cases, we implement this numerical optimization using the same Nelder–Mead routine in `NLopt` that we use to estimate the MLE, $\hat{\theta}$ [38]. We define two uniformly spaced meshes either side of the MLE in the interest parameter: (i) starting at the MLE to the lower bound of the interest parameter and (ii) starting from the MLE to the upper bound of the interest parameter. For all results in this work, each mesh is formed by 40 points resulting in a total of 80 mesh points for each profile. For each mesh point to run the numerical optimization, we provide a starting estimate of the parameters. For the first mesh point closest to the MLE, we set the starting estimate of r, K, $C(0)$ and σ^2 equal to their respective values in the MLE. We then seek the values of r, K, $C(0)$ and σ^2 that maximize $\hat{\ell}(r, K, C(0), T, \sigma^2 \mid y_{1:I}^{\mathrm{o}})$. For the second mesh point closest to the MLE, we use the estimate from the previous point as the starting estimate. For the starting estimate for all other mesh points, we make a linear approximation using estimates at the previous two mesh points. The linear approximation holds provided the estimate remains within bounds. If it does not hold, we set the first guess as the previous estimate provided it remains within bounds and as the MLE otherwise. With these profiles, log-likelihood-based confidence intervals can be defined from the profile log-likelihood by an asymptotic approximation in terms of the chi-squared distribution that holds for sufficiently regular problems [26]. For example, 95%, 99% and 99.9% confidence intervals for a univariate (scalar) interest parameter correspond to a threshold profile log-likelihood value of −1.92, −3.32 and −5.41, respectively [43].

5.2. Predictive profile likelihood and parameter-wise profile predictions

Profile likelihoods for predictive quantities that are a (deterministic) function of the full parameter θ are defined in the same way as for any other function of the full parameter, as described in §5.1. For example, the full process model trajectory, $y(\theta)$, has an associated profile likelihood

$$\hat{\ell}_p(y(\theta) = y \mid y_{1:I}^{\mathrm{o}}) = \sup_{\theta \mid y(\theta) = y} \hat{\ell}(\theta \mid y_{1:I}^{\mathrm{o}}), \qquad (5.5)$$

i.e. the profile likelihood value for a prediction is equal to the maximum likelihood value across parameters consistent with that prediction. Here, we give the profile prediction for the full (infinite-dimensional) model trajectory, which here is more straightforward in principle than general functional estimation problems as the variation is driven by a finite-dimensional parameter vector (and the constraint defined by the differential equation). However, this constraint may be more difficult to enforce in practice than the solution at a single time, and the literature typically focuses on a

single-time prediction [27,44]. In the special case that $y(\theta)$ is a one-to-one function, the aforementioned reduces to

$$\hat{\ell}_p(y(\theta)\,|\,y_{1:I}^o) = \hat{\ell}(\theta\,|\,y_{1:I}^o), \qquad (5.6)$$

since the constraint $y(\theta)$ is uniquely invertible for θ. That is, profiling preserves the usual parametrization invariance of the likelihood function under one-to-one transformations [41,42]. However, profile predictions are still well defined even without a one-to-one relationship between the parameters and model solution (i.e. in the absence of structural identifiability) [27] (see also [44,45]).

Here, we are also interested in some measure of the dependence of predictions on given target (interest) parameters. However, given a partition $\theta = (\psi, \lambda)$ and a function $q(\theta)$ of the full parameter, there is not in general a well-defined meaning of $q(\psi)$, unless q is independent of λ. A natural approach then to exploring the dependence of a predictive function $q(\psi, \lambda)$ of the full parameter on an interest parameter ψ is to consider its value along the corresponding profile curve, i.e. $q(\psi, \lambda^*(\psi))$, where $\lambda^*(\psi)$ is the optimal value of the nuisance parameter for a given value of the interest parameter. We call these *parameter-wise profile predictions*, in contrast to the more standard predictive profile likelihood [27]. In the simple case where q is independent of λ (or if λ is known) and is 1–1 in ψ, then this amounts to a re-parametrization of the ψ profile likelihood. Hence, in this case, confidence intervals for ψ are directly transformed into confidence intervals for q (by transformation invariance of likelihood functions). The 1–1 requirement can be relaxed in the same way as for standard interest parameters but, in more complex cases with non-trivial dependence on the nuisance parameters, the transformation of confidence intervals for ψ into confidence intervals for the predictive quantity of interest will only be approximate and the precise statistical properties of these approximate prediction intervals are more difficult to establish (though can always be evaluated by simulation). In particular, if the predictive quantity of interest has weak or no dependence on the interest parameter being profiled and non-trivial dependence on the nuisance parameters, the associated predictive interval would be expected to have poor coverage. However, we can still use these parameter-wise intervals as an intuitive model diagnostic tool revealing the influence of an interest parameter on predictions. In contrast, a standard predictive profile cannot reveal the individual influence of particular parameters. With these caveats in mind, we define the associated profile likelihood for $q(\psi, \lambda^*(\psi))$ analogously to standard profile likelihood for an interest parameter, now starting from the profile likelihood for ψ,

$$\hat{\ell}_p(q(\psi, \lambda^*(\psi)) = q\,|\,y_{1:I}^o) = \sup_{\psi\,|\,q(\psi,\lambda^*(\psi))=q} \hat{\ell}_p(\psi\,|\,y_{1:I}^o). \qquad (5.7)$$

As with the standard profile likelihood, this definition preserves parametrization invariance under 1–1 transformations, i.e. if $q(\psi, \lambda^*(\psi))$ is 1–1 in ψ, then

$$\hat{\ell}_p(q(\psi, \lambda^*(\psi))\,|\,y_{1:I}^o) = \hat{\ell}_p(\psi\,|\,y_{1:I}^o). \qquad (5.8)$$

In addition to parameter-wise intervals, given a collection of individual intervals for the same quantity but based on different interest parameters, more conservative confidence intervals (relative to the individual intervals) for the predictions can be constructed by taking the union over all intervals. For example, given two intervals (or sets)

$\mathcal{I}_q^T = [L_q^T, U_q^T]$ and $\mathcal{I}_q^r = [L_q^r, U_q^r]$ for a quantity q based on the profiles for T and r, respectively, we can form an interval (or set) $\mathcal{I}_q^{T,r} = \mathcal{I}_q^T \cup \mathcal{I}_q^r$, which has coverage at least as great as the individual intervals. In the case where the two intervals overlap, we have $\mathcal{I}_q^{T,r} = [\min\{L_q^T, L_q^r\}, \max\{U_q^T, U_q^r\}]$. Again, the precise coverage properties of these intervals are difficult to establish, but such union intervals can provide an intuitive picture of overall variation in the predictive quantity.

6. Results and discussion

Here, we apply the general modelling framework that we present in §§3–5 to three case studies across the life sciences. We discuss a fourth case study, an additional two-dimensional cell proliferation assay that we perform with a bladder cancer cell line, in electronic supplementary material, F. These case studies cover a range of spatial and temporal scales, from microns and hours to kilometres and years, respectively.

6.1. Coral reef growth after disturbance

Recent modelling studies that examine the regrowth of coral reefs after some kind of disturbance (e.g. cyclone) have begun to explore the possibility that the regrowth involves biphasic growth [6], whereas earlier studies have simply ignored this possibility [6,17,18]. Here, we explore measurements of the coral cover percentage, $C(t)$ (%), of the reef at Broomfield Island, Great Barrier Reef, Australia (figure 2a) [6]. In the first phase of growth, $C(t)$ remains approximately constant. In the second phase of growth, $C(t)$ is sigmoidal. To describe the second phase, we take the simplest approach and use the logistic growth model [18]. Therefore, we set $f_1(C) = 0$ and $f_2(C) = r\,C\,(1 - C/K)$ in equation (3.1) and seek estimates of five parameters, $\theta = (T, C(0), r, K, \sigma)$.

Comparing the experimental data with the mathematical model simulated with the MLE, we observe very good agreement with small residuals that appear to be independent and identically distributed (figure 2b). In terms of practical identifiability, the profile likelihood for T is wide, with approximate 99.9% confidence interval $0 \leq T \leq 1220$ (34% of entire dataset duration). Our interpretation of this result is that these coral data are insufficient to obtain a precise point estimate of T. Therefore, while it is unclear from these data whether there is a delay we have quantified the uncertainty in T. This is useful since understanding whether the coral reef growth involves a delay is important for management and intervention strategies [6]. The profile likelihoods for $C(0)$, r, K and σ are relatively narrow and each well formed around a single central peak, suggesting that these parameters are practically identifiable (figure 2d–g). We validate that the framework accurately estimates model parameters by repeating this analysis with synthetic data based on the coral reef data (electronic supplementary material, B). Reducing the variance of the synthetic data suggests that the model parameters are structurally identifiable.

To improve our understanding of how each parameter influences mathematical modelling predictions, we use parameter-wise profile predictions. We generate parameter-wise profile predictions for each of the five parameters and their union. A great advantage of using parameter-wise profile predictions is that we can identify the contribution of each of the parameters to predictions. Firstly, we present the parameter-wise profile prediction for T (figure 3a) and the difference between the parameter-wise profile prediction for T, $\mathcal{I}_q^T$, and the mathematical
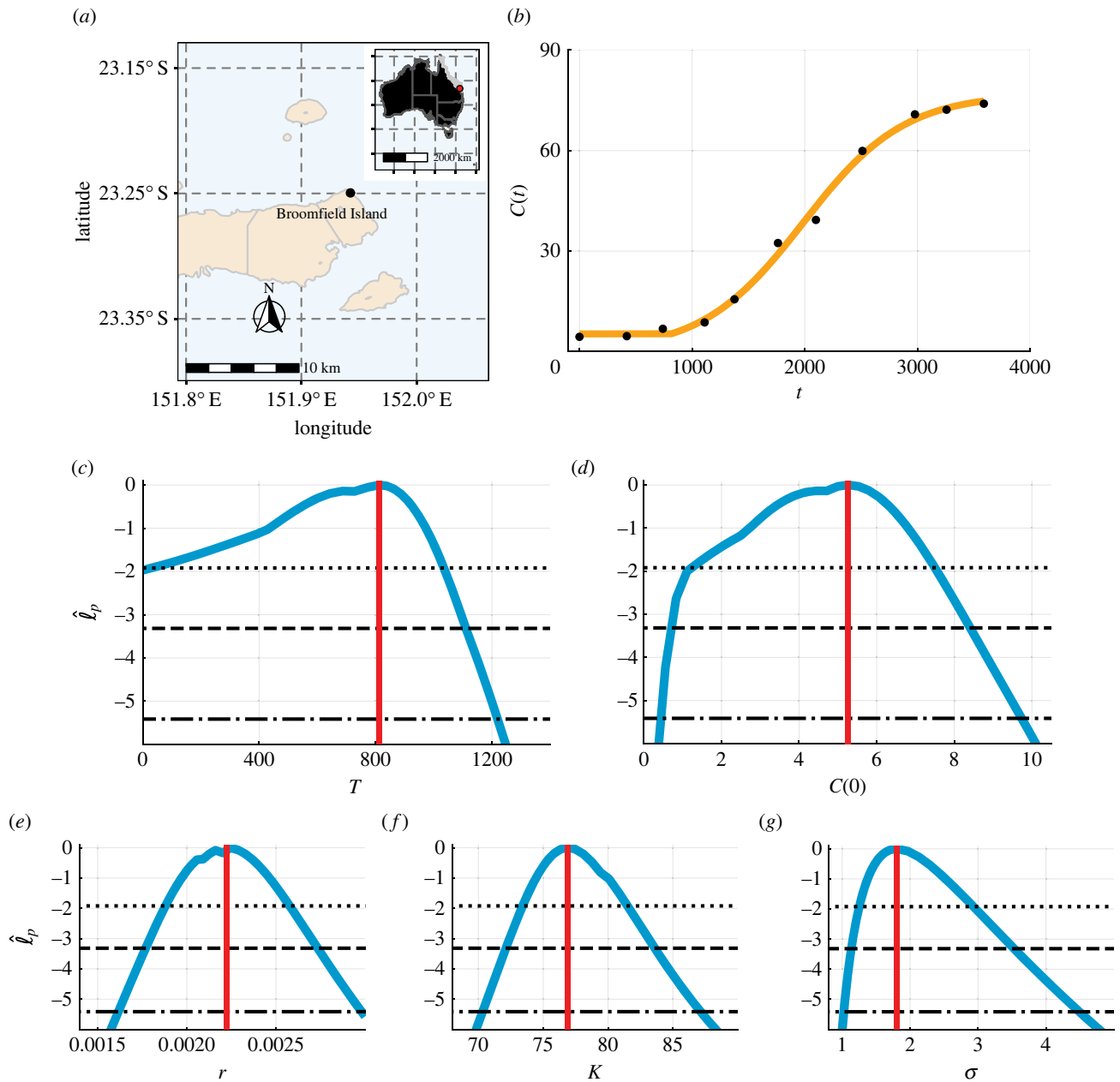
Figure 2. Biphasic coral reef growth after a disturbance. The coral cover percentage, $C(t)$ (%), at (a) Broomfield Island, Great Barrier Reef, Australia. (b) Comparison of the mathematical model simulated with the MLE (orange line) and field data (black circles) for the coral cover percentage, $C(t)$ (%). (c–g) Profile likelihoods for (c) T (days), (d) $C(0)$ (%), (e) r (days^{-1}), (f) K (%) and (g) σ (%) (blue) together with the maximum likelihood estimation (MLE) (vertical red line) and approximate 95% (dotted), 99% (dashed) and 99.9% (dash-dotted) confidence interval thresholds. The approximate 99.9% confidence intervals are as follows: (c) $T \in (0, 1220)$ (days), (d) $C(0) \in (0.42, 9.74)$ (%), (e) $r \in (0.016, 0.0030)$ (days^{-1}), (f) $K \in (70.3, 87.1)$ (%) and (g) $\sigma \in (1.02, 4.50)$ (%).

model simulated with the MLE, $y(\hat{\theta})$, denoted $\mathcal{I}_q^T - y(\hat{\theta})$ (figure 3b). These results are very insightful since they show how uncertainty in each parameter affects different aspects of the predictions made using the model. For example, uncertainty in T leads to a relatively wide prediction interval at early time, but has very little impact upon the late-time prediction interval (figure 3a,b) which is intuitively reasonable since the late-time behaviour of the model is dictated by K rather than T. Similarly, we see that uncertainty in $C(0)$ leads to a wide prediction interval at early time, but a smaller prediction interval at late time, which is also consistent with our understanding that $C(0)$ plays in this model (figure 3c,d). In contrast, uncertainty in K leads to a relatively wide prediction interval at late time, as expected, but a narrow prediction interval at early time (figure 3g,h). As expected, σ provides zero contribution to the prediction of the mean due to the form of the error model.

Given these parameter-wise prediction intervals, we can then take the union of the parameter-wise profile predictions and understand how it is formed and the contribution of each parameter (figure 3i,j).

Many early studies of coral reef regrowth often ignore the possibility of biphasic growth (i.e. fixing $T = 0$) and do not allow for the possibility that $C(0)$ can be estimated from the data (i.e. fixing $C(0)$ equal to the first measurement) [6,17,18]. To demonstrate the impact of these more standard choices, we repeat the analysis of this data under these assumptions (figure 4). The mathematical model simulated with the MLE is fixed to capture the first data point (figure 4a), but agreement to the other data points is considerably poorer in comparison with the biphasic model (figure 2a). Furthermore, the residuals in this case are visually correlated, with systematic underestimation at early times
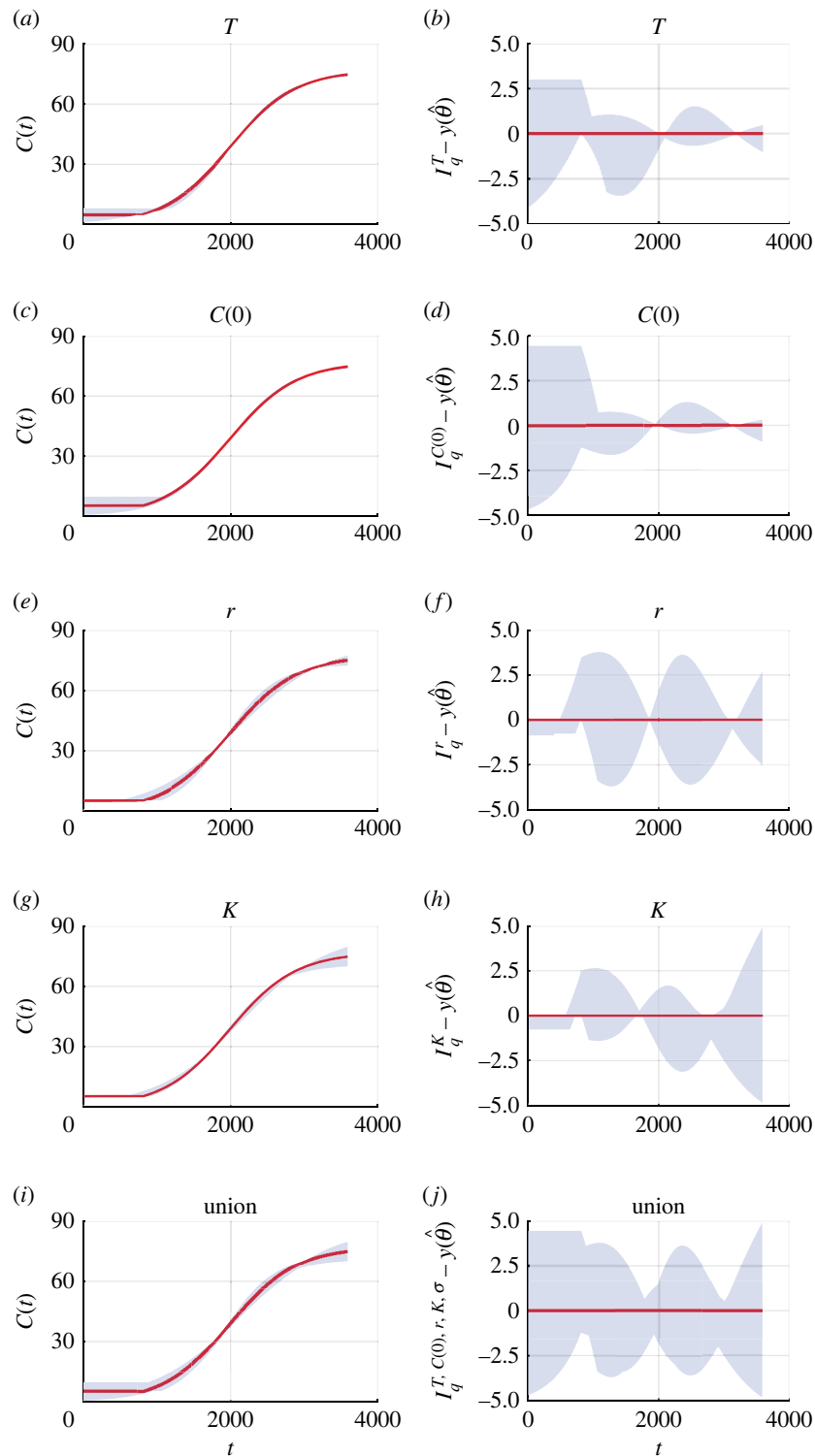
Figure 3. Parameter-wise profile predictions for coral reef growth. (a,c,e,g,i) Parameter-wise profile predictions for the mean (shaded) and the mathematical model simulated with the MLE (red). (b,d,f,h,j) Difference between parameter-wise profile predictions for the mean and the mathematical model simulated with the MLE. Results shown for (a,b) T, (c,d) $C(0)$, (e,f) r, (g,h) K and (i,j) the union of the parameter-wise profile predictions.

and some overestimation at later times, violating statistical assumptions that the residuals are independent and identically distributed [18]. To compare the model where all parameters are estimated (approach 1) with the model where we fix $T = 0$ and set $C(0)$ equal to the first experimental measurement (approach 2), we use the Akaike information criteria (AIC) [46]. The AIC is a standard tool for model selection studies and defined as AIC $= -2 \sup_\theta \ell(\theta | y_{1:I}^o) + 2k$, where k is the dimensionality of θ [47]. When k is the same for different models, the AIC is a comparison of the

maximum likelihood estimates, and when k is different, the model with more parameters is given a larger penalty. The AIC is smaller for approach 1 (54.5) than approach 2 (69.5), and this suggests that approach 1 is more appropriate.

The AIC provides a single numerical value to compare mathematical models. Here, we provide further insights by comparing the profile likelihoods [18]. Profile likelihoods and approximate confidence intervals for parameters of the single-phase model are different to the corresponding profile likelihoods of the biphasic model. Specifically, estimates of r are
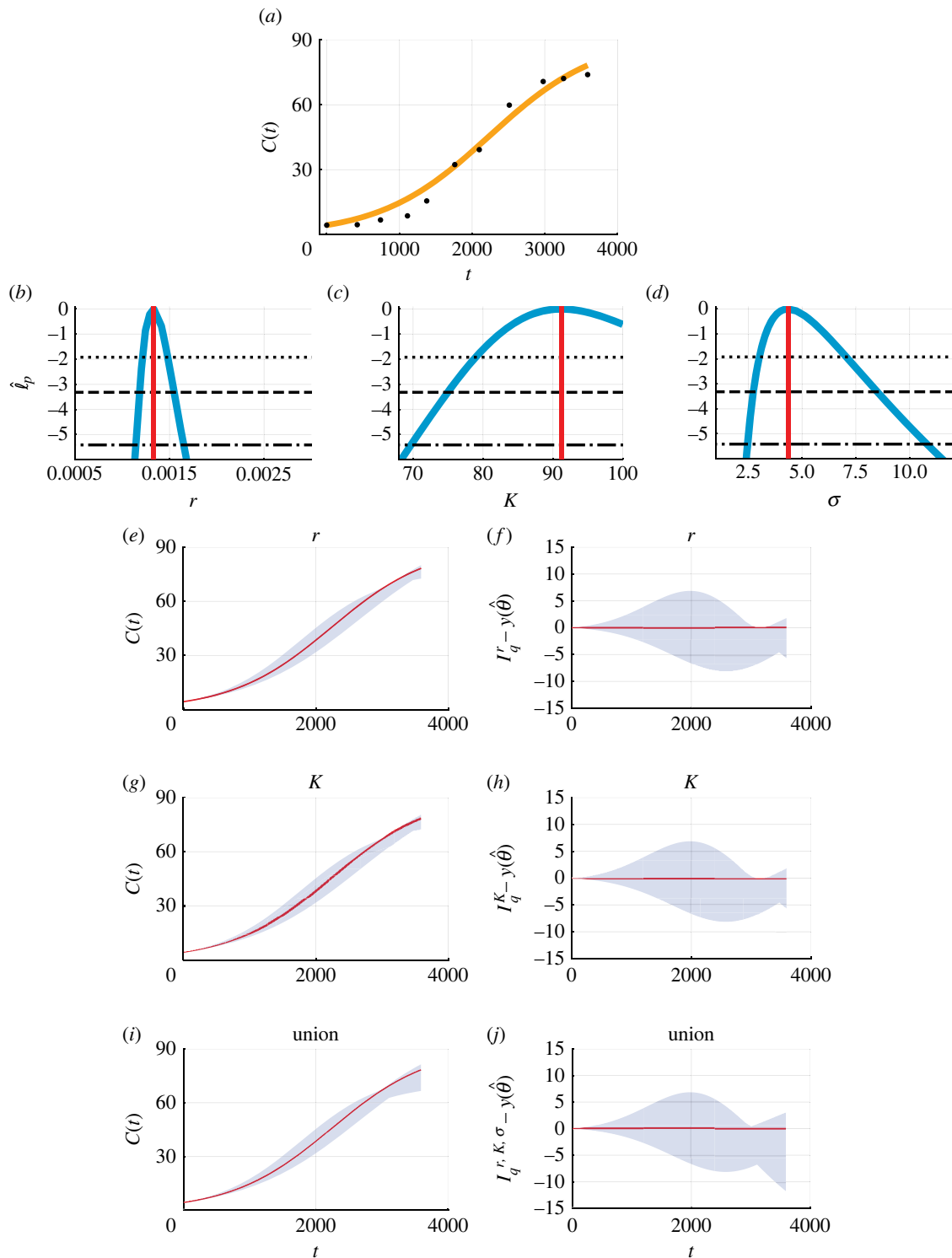
Figure 4. Biphasic coral reef growth after a disturbance: fixing $T = 0$ and setting $C(0)$ equal to the first measurement. (a) Comparison of the mathematical model simulated with the MLE (orange line) and field data (black circles) for the coral cover percentage, $C(t)$ (%), measured at Broomfield Island, Great Barrier Reef, Australia. (b–d) Profile likelihoods for (b) r (days^{-1}), (c) K (%) and (d) σ (%) (blue) together with the MLE (vertical red line) and approximate 95% (dotted), 99% (dashed) and 99.9% (dash-dotted) confidence interval thresholds. The approximate 99.9% confidence intervals are as follows: (b) $r \in (0.0011, 0.0017)$ (days^{-1}), (c) $K \in (69.7, 100.00)$ (%) and (d) $\sigma \in (2.46, 10.81)$ (%). (e, g, i) Parameter-wise profile predictions for the mean (shaded) and the mathematical model simulated with the MLE (red). (f, h, j) Difference between parameter-wise profile predictions for the mean and the mathematical model simulated with the MLE. Results shown for (e, f) r, (g, h) K and (i, j) the union of the parameter-wise profile predictions.

smaller in the single-phase model than the biphasic model (figures 2e and 4b). Furthermore, the approximate confidence interval for K is much larger for the single-phase model (figures 2f and 4c). Such differences in parameter estimates could have major impacts on intervention and management strategies. For example, the single-phase model suggests that it is likely that coral cover will eventually reach 100% ($K = 100$%),

whereas $K = 100$% is a very unlikely prediction from the biphasic model. In electronic supplementary material, C, we explore approach 3, a single-phase model (i.e. $T = 0$) without fixing $C(0)$. We find that approach 3 does not capture the first data point, cannot be used to quantify uncertainty in T, and results in wider approximate confidence intervals for model parameters in comparison with approach 1.
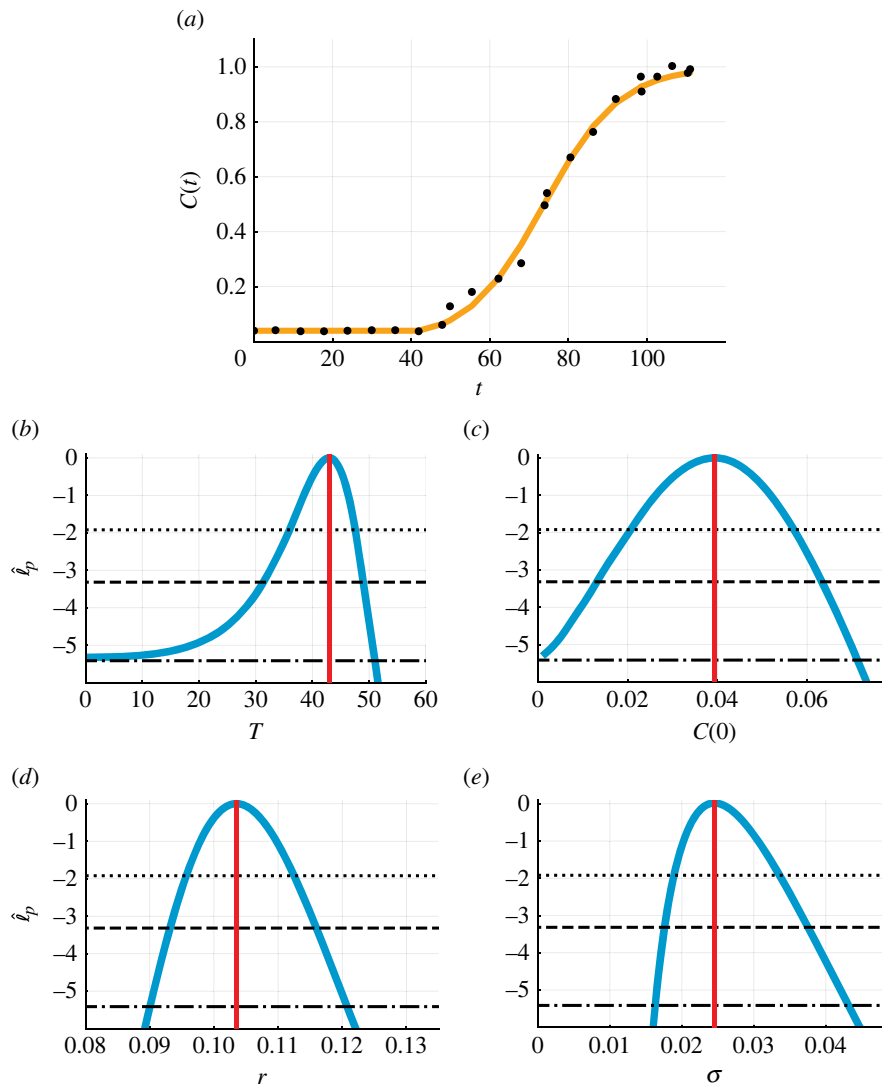
Figure 5. Biphasic population growth in a two-dimensional cell proliferation assay. (a) Comparison of the mathematical model simulated with the MLE (orange line) and experimental data (black circles) for the normalized cell density, $C(t)$ (—). (b–e) Profile likelihoods for (b) T (hours), (c) $C(0)$ (—), (d) r (hours^{-1}) and (e) σ (—) (blue) together with the MLE (vertical red line) and approximate 95% (dotted), 99% (dashed) and 99.9% (dash-dotted) confidence interval thresholds. The approximate 99.9% confidence intervals are as follows: (b) $T \in (0.0, 51.0)$ (hours), (c) $C(0) \in (0.001, 0.071)$ (—), (d) $r \in (0.090, 0.121)$ (hours^{-1}) and (e) $\sigma \in (0.016, 0.043)$ (—).

6.2. Two-dimensional cell proliferation assay

Inspecting the time evolution of the normalized cell density, $C(t) \in [0, 1]$ (—), in two-dimensional cell proliferation assays, we observe biphasic population growth (figure 5a). In the first phase of growth, $C(t)$ remains approximately constant. In the second phase of growth, $C(t)$ is sigmoidal. As mentioned earlier, we take the simplest approach and describe the second phase using the logistic growth model. Therefore, we set $f_1(C) = 0$ and $f_2(C) = r\, C\, (1 - C)$ in equation (3.1). We now seek estimates of four parameters, $\theta = (T, C(0), r, \sigma)$.

Comparing the experimental data with the mathematical model simulated with the MLE, we observe very good agreement with small visually uncorrelated residuals (figure 5a). The profile likelihood for T is well-formed around a single central peak, suggesting that T is practically identifiable to a 99% approximate confidence interval threshold (figure 5b). However, the approximate 99.9% confidence interval is wider, $0 < T < 51$ (hours), and the MLE is 43 (hours). The previous analysis of this dataset used visual inspection to estimate $T = 40$ (hours). The approach we use here is more objective and reproducible and consequently more reliable

and accurate than the previous method. Further, our approach provides an approximate confidence interval rather than a point estimate. Profile likelihoods for the three other parameters, $C(0)$, r and σ suggest that they are practically identifiable (figure 5c–e). Parameter-wise profile predictions reveal the influence of individual model parameters on predictions (figure 6). Similar results are obtained for the fourth case study, a different cell proliferation assay experiment that we perform with a bladder cancer cell line and larger initial density (electronic supplementary material, F).

6.3. Three-dimensional cancer tumour spheroid experiment

We now consider a growing population of cancer cells in a three-dimensional tumour spheroid experiment reported in [5,9]. The overall process of spheroid formation and growth involves two phases: in phase (i), cells placed in the well migrate and adhere to form a shrinking spheroid; and in phase (ii), the newly formed spheroid grows as compact solid mass increases (figures 1b and 7a–e). Over the entire experimental duration
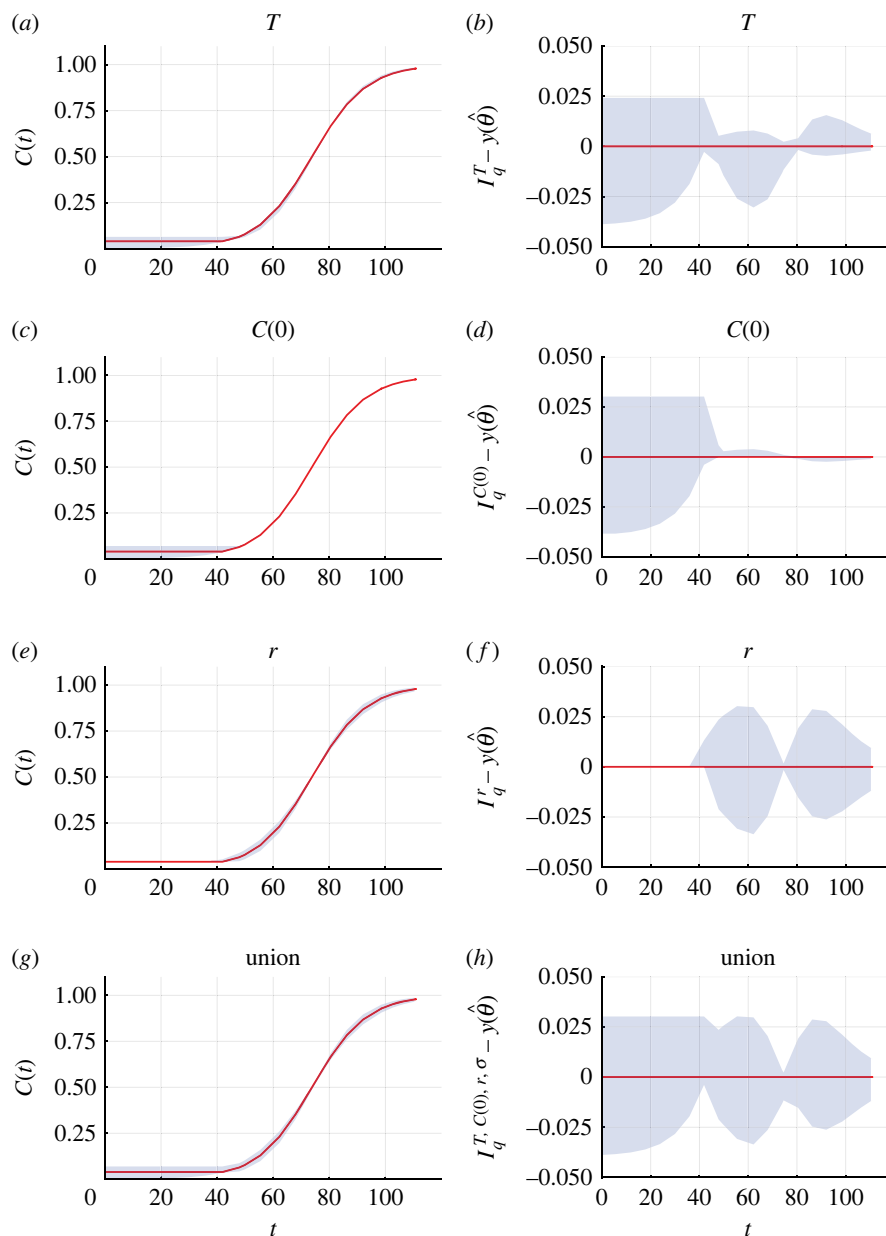
Figure 6. Parameter-wise profile predictions for two-dimensional cell proliferation assay. (a,c,e,g) Parameter-wise profile predictions for the mean (shaded) and the mathematical model simulated with the MLE (red). (b,d,f,h) Difference between parameter-wise profile predictions for the mean and the mathematical model simulated with the MLE. Results shown for (a,b) T, (c,d) $C(0)$, (e,f) r and (g,h) the union of the parameter-wise profile predictions.

$0 < t < 432$ (hours), the spheroid radius, $R(t)$, increases to a long-time maximum radius, $\mathcal{R}_2$ (μm) (electronic supplementary material, G). Here, to illustrate the early-time biphasic behaviour, we focus on $0 < t < 120$ (hours).

Many models could be chosen to describe and analyse how $R(t)$ evolves in time. Model selection has been well studied for the second phase of growth [48,49]; however, the first phase where the spheroid forms is rarely studied. Here, we take a minimal approach and assume both phases can be described by distinct logistic growth models, giving

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \begin{cases} r_1 R(t)\left(1 - \frac{R(t)}{\mathcal{R}_1}\right), & 0 \leq t \leq T, \\ r_2 R(t)\left(1 - \frac{R(t)}{\mathcal{R}_2}\right), & t > T, \end{cases} \quad (6.1)$$

where r_1 and r_2 are the growth rates in the first and second phase, respectively, and $\mathcal{R}_1$ and $\mathcal{R}_2$ are the associated limiting radii in each phase. Overall, we have seven parameters to estimate $\theta = (T, R(0), r_1, r_2, \mathcal{R}_1, \mathcal{R}_2, \sigma)$. Using the logistic

growth model to simulate the growth of cell populations where the density is less than the long-time carrying capacity density is extremely common [3,19,48,49]. In contrast, using logistic growth where the dependent variable is greater than the long-time carrying capacity, as we do here to describe the first phase of spheroid formation, is quite unusual [50]. However, we find that this approach provides a good description of our experimental observations using a very familiar mathematical model.

Comparing the experimental data with the mathematical model simulated with the MLE, we observe excellent agreement (figure 7e). The profile likelihood for T is well formed around a single central peak, suggesting that T is practically identifiable to the 99.9% approximate confidence interval threshold (figure 7b). Profile likelihoods suggest that five of the six other parameters, $R(0)$, $C(0)$, r_1, r_2, $\mathcal{R}_1$ and σ, are practically identifiable (figure 7g–j and l). The profile likelihood for $\mathcal{R}_2$ is well formed around a single central peak and practically identifiable to a 95% approximate confidence interval threshold (figure 7k).
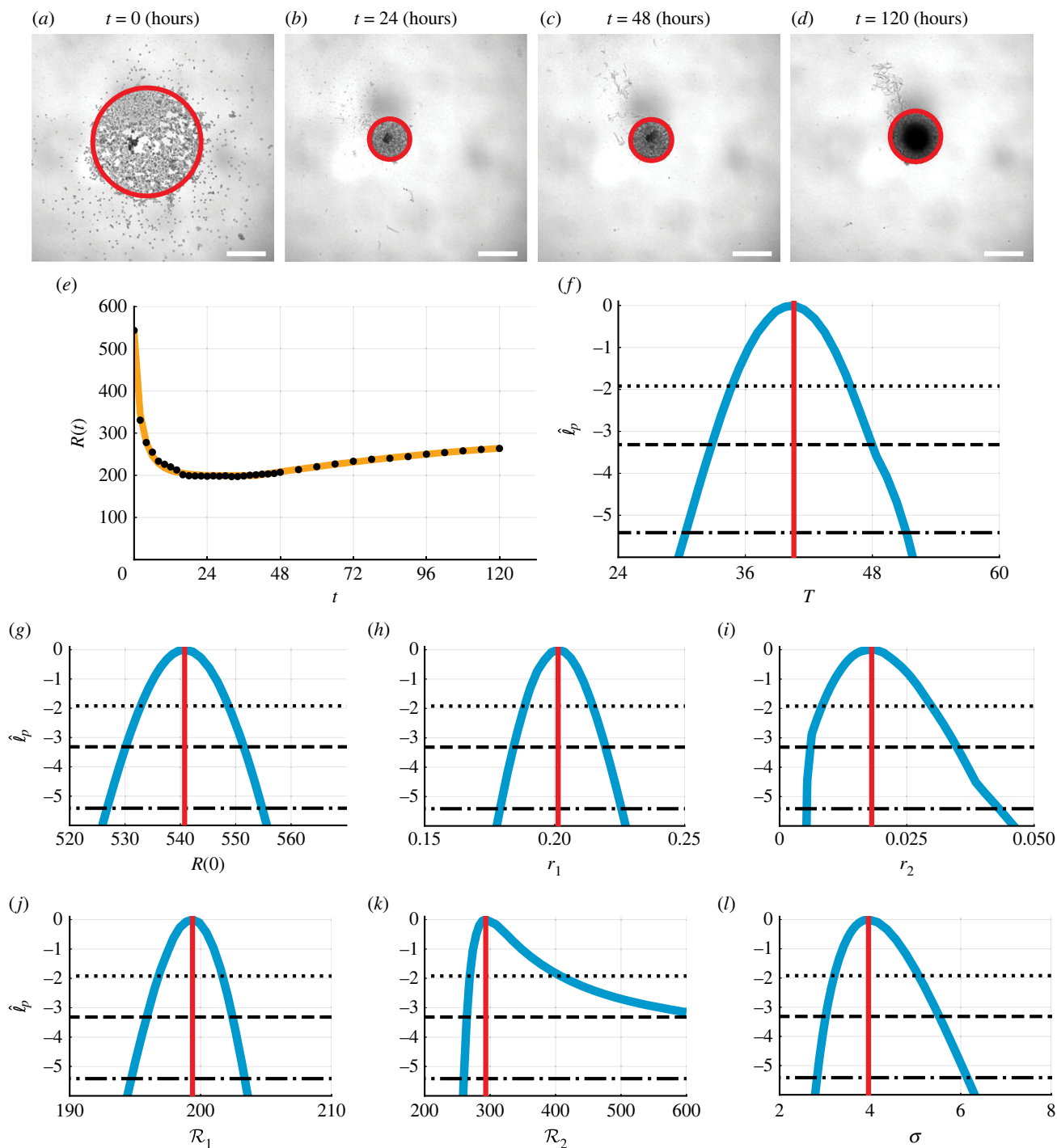
Figure 7. Biphasic population growth in three-dimensional tumour spheroid experiments. (*a–d*) Experimental images of spheroid experiments at (*a*) $t = 0$, (*b*) $t = 24$, (*c*) $t = 48$ and (*d*) $t = 120$ (hours). Scale bar is 400 μm, and red circle shows approximate equivalent area. At early times, cells migrate and adhere to form spheroid. At later times, the spheroid grows as a compact mass. (*e*) Comparison of the mathematical model simulated with the MLE (orange line) and experimental data (black circles) for the equivalent radius, $R(t)$ (μm). (*f–l*) Profile likelihoods for (*f*) T (hours), (*g*) $R(0)$ (μm), (*h*) r_1 (hours^{-1}), (*i*) r_2 (hours^{-1}), (*j*) $\mathcal{R}_1$ (μm), (*k*) $\mathcal{R}_2$ (μm) and (*l*) σ (μm) (blue) together with the MLE (vertical red line) and approximate 95% (dotted), 99% (dashed) and 99.9% (dash-dotted) confidence interval thresholds. The approximate 99.9% confidence intervals are as follows: (*f*) $T \in (30.4,\ 51.2)$ (hours), (*g*) $R(0) \in (526.8,\ 554.8)$ (μm), (*h*) $r_1 \in (0.179,\ 0.226)$ (hours^{-1}), (*i*) $r_2 \in (0.005,\ 0.043)$ (hours^{-1}), (*j*) $\mathcal{R}_1 \in (194.7,\ 203.4)$ (μm), (*k*) $\mathcal{R}_2 \in (259.6,\ 600.0)$ (μm) and (*l*) $\sigma \in (2.82,\ 6.14)$ (μm).

However the approximate 99% and 99.9% confidence intervals are wider, suggesting that the parameter $\mathcal{R}_2$ is practically non-identifiable using this dataset (figure 2*b*). Increasing the experimental duration narrows the confidence intervals for r_2 and $\mathcal{R}_2$, suggesting that they are practically identifiable with appropriate additional data (electronic supplementary material, figures S12 and S13). Parameter-wise profile predictions reveal the influence of individual model parameters on predictions (figure 8). Here, our framework improves on previous methods that use visual inspection to identify the start of the second phase of growth for analysis [5,9].

7. Conclusion and outlook

In this study, we present a computationally efficient framework for diagnosing, understanding and predicting biphasic population growth. Our framework involves two key components:
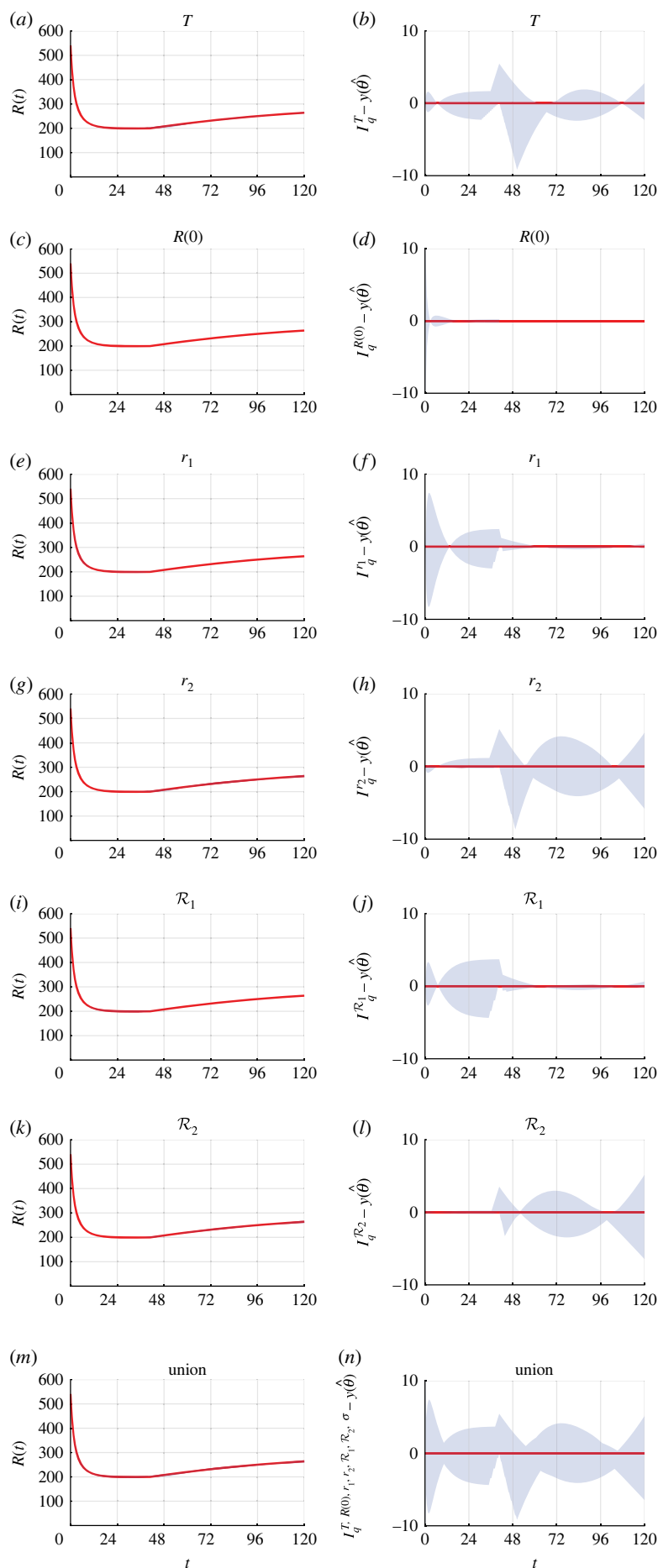
Figure 8. Parameter-wise profile predictions for three-dimensional cancer tumour spheroid experiment. (*a,c,e,g,i,k,m*) Parameter-wise profile predictions for the mean (shaded) and the mathematical model simulated with the MLE (red). (*b,d,f,h,j,l,n*) Difference between parameter-wise profile predictions for the mean and the mathematical model simulated with the MLE. Results shown for (*a,b*) T, (*c,d*) $R(0)$, (*e,f*) r_1, (*g,h*) r_2, (*i,j*) $\mathcal{R}_1$, (*k,l*) $\mathcal{R}_2$ and (*m,n*) the union of the parameter-wise profile predictions.

(i) an efficient method to form approximate confidence intervals for the change point of the growth dynamics and model parameters and (ii) parameter-wise profile predictions that systematically reveal the influence of uncertainty for individual model parameters upon the model predictions. To demonstrate our framework, we explore real-world case studies across the life sciences. This work builds on previous studies that focus on single-phase models to describe biphasic growth, point estimates of biphasic model parameters and specific mathematical models and applications.

The ability to estimate the change point and model parameters in combination with parameter-wise profile predictions is powerful. For experimental design, parameter-wise profile predictions can inform when additional measurements should be taken to improve estimates of individual parameters. For the cell biology case studies, we provide accurate estimates of growth rates that can assist decision-making in experiments, for example when to apply drug treatments [51]. For the coral reef growth, understanding whether growth involves a delay is important for management and intervention strategies [6]. Here, by using our biphasic modelling framework rather than a single-phase model, one can account for the existence of a delay phase and quantify the associated uncertainty. These case studies vary in terms of application and data quality, from sparse noisy data in coral reef studies to dense data collected in controlled experimental conditions in cell biology experiments. For all case studies, the framework provides accurate parameter estimates and parameter-wise prediction intervals that lead to valuable insights.

Our work introduces parameter-wise prediction intervals in terms of the intuitive picture of variation in the predictive quantity. An open question from a theoretical point of view is how union intervals are compared with standard profile predictive intervals for the same quantity. Because such parameter-wise (and union of parameter-wise) intervals are based on direct propagation of parameter uncertainties, these are typically easier to compute than standard profile prediction intervals as the latter require enforcing constraints on the model outputs rather than inputs. On the other hand, standard profile prediction intervals are more well-established theoretically.

This framework can be extended in many theoretical directions and to many applications. We take the simplest approach and use the well-known logistic model to describe population dynamics. However, the framework is general and well suited to explore other models, for example Gompertz, generalized logistic and Richard's [18,48,49]. Our biphasic modelling framework can also be applied to the growth of individuals within a population [20,52] and extended to explore growth dynamics that exhibit three or more growth phases [53]. Throughout, we assume a normal error model as it is the simplest and most common approach. Exploring different error models, such as lognormal and exponential, within our likelihood-based framework may be of interest in different biological contexts. Since we simultaneously estimate parameters from the mathematical model as well as parameters in the statistical model, our framework is also well suited to analyse different ecological systems with more noise. Exploring process stochasticity is also of interest [54]. We use a profile likelihood-based framework rather than a Markov chain Monte Carlo approach for computational efficiency [33]. In future work, one could compare the computational efficiency of the two approaches specifically for biphasic growth models. One could also explore spatial effects by extending spatio-temporal single-phase partial-differential equation growth models [33,55] to spatio-temporal biphasic growth models. Overall, this work lays the foundation for studies in biphasic population growth using differential equations, efficient change point and model parameter estimation, and parameter-wise prediction intervals.

References

1. Thieme HR. 2003 *Mathematics in population biology*. Princeton, NJ: Princeton University Press.

2. Hastings A. 2013 *Population biology: concepts and models*. New York, NY: Springer Science & Business Media.

3. Murray JD. 2002 *Mathematical biology I: an introduction*. Heidelberg, Germany: Springer.

4. Beeden R, Maynard J, Puotinen M, Marshall P, Dryden J, Goldberg J, Williams G. 2015 Impacts and recovery from severe tropical cyclone Yasi on the Great Barrier Reef. *PLoS ONE* **10**, e0121272. (doi:10.1371/journal.pone.0121272)

5. Murphy RJ, Browning AP, Gunasingh G, Haass NK, Simpson MJ. 2022 Designing and interpreting 4D tumour spheroid experiments. *Commun. Biol.* **5**, 91. (doi:10.1038/s42003-022-03018-3)

6. Warne DJ, Crossman KA, Jin W, Mengersen K, Osborne K, Simpson MJ, Thompson AA, Wu P, Ortiz J-C. 2021 Identification of two-phase recovery for interpretation of coral reef monitoring data. *J. Appl. Ecol.* **59**, 153–164. (doi:10.1111/1365-2664.14039)

7. Tremel A, Cai A, Tirtaatmadja N, Hughes BD, Stevens GW, Landman KA, O'Connor AJ. 2009 Cell migration and proliferation during monolayer formation and wound healing. *Chem. Eng. Sci.* **64**, 247–253. (doi:10.1016/j.ces.2008.10.008)

8. Jin W, Shah ET, Penington CJ, McCue SW, Maini PK, Simpson MJ. 2017 Logistic proliferation of cells in scratch assays is delayed. *Bull. Math. Biol.* **79**, 1028–1050. (doi:10.1007/s11538-017-0267-4)

9. Browning AP, Sharp JA, Murphy RJ, Gunasingh G, Lawson B, Burrage K, Haass NK, Simpson MJ. 2021 Quantitative analysis of tumour spheroid structure. *eLife* **10**, e73020. (doi:10.7554/eLife.73020)

10. Brouwer AF, Eisenberg MC, Remais JV, Collender PA, Meza R, Eisenberg JNS. 2017 Modeling biphasic environmental decay of pathogens and implications for risk analysis. *Environ. Sci. Technol.* **51**, 2186–2196. (doi:10.1021/acs.est.6b04030)

11. Phaiboun A, Zhang Y, Park B, Kim M. 2015 Survival kinetics of starving bacteria is biphasic and density-dependent. *PLoS Comput. Biol.* **11**, e1004198. (doi:10.1371/journal.pcbi.1004198)

12. Møller COA, Christensen BB, Rattray FP. 2021 Modelling the biphasic growth of non-starter lactic acid bacteria on starter-lysate as a substrate. *Int. J. Food Microbiol.* **337**, 108937. (doi:10.1016/j.ijfoodmicro.2020.108937)

13. Aminikhanghahi S, Cook DJ. 2017 A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **51**, 339–367. (doi:10.1007/s10115-016-0987-z)

14. Hinkley DV. 1970 Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–17. (doi:10.2307/2334766)

15. Pettitt AN. 1979 A non-parametric approach to the change-point problem. *J. R. Stat. Soc. C-App.* **28**, 126–135. (doi:10.2307/2346729)

16. MathWorks `findchangepts`. See https://au.mathworks.com/help/signal/ref/findchangepts.html (accessed 7 July 2022).

17. Thompson A, Martin K, Logan M. 2020 Development of the coral index, a summary of coral reef resilience as a guide for management. *J. Environ. Manage.* **271**, 111038. (doi:10.1016/j.jenvman.2020.111038)

18. Simpson MJ, Browning AP, Warne DJ, Maclaren OJ, Baker RE. 2022 Parameter identifiability and model selection for sigmoid population growth models. *J. Theor. Biol.* **535**, 110998. (doi:10.1016/j.jtbi.2021.110998)

19. Warne DJ, Baker RE, Simpson MJ. 2017 Optimal quantification of contact inhibition in cell populations. *Biophys. J.* **113**, 1920-1924. (doi:10.1016/j.bpj.2017.09.016)

20. Honsey AE, Staples DF, Venturelli PA. 2017 Accurate estimates of age at maturity from the growth trajectories of fishes and other ectotherms. *Ecol. Appl.* **27**, 182–192. (doi:10.1002/eap.1421)

21. Chis O-T, Banga JR, Balsa-Canto E. 2011 Structural identifiability of systems biology models: a critical comparison of methods. *PLoS ONE* **6**, e27755. (doi:10.1371/journal.pone.0027755)

22. 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**, 1923–1929. (doi:10.1093/bioinformatics/btp358)

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

24. Bellu G, Saccomani MP, Audoly S, D'Angió L. 2007 DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput. Meth. Prog. Bio.* **88**, 52–61. (doi:10.1016/j.cmpb.2007.07.002)

25. Ligon TS, Fröhlich F, Chiş OT, Banga JR, Balsa-Canto E, Hasenauer J. 2017 GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models. *Bioinformatics* **34**, 1421–1423. (doi:10.1093/bioinformatics/btx735)

26. Pawitan Y. 2001 *In all likelihood: statistical modelling and inference using likelihood*. Oxford, UK: Oxford University Press.

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

28. Campbell DA, Chkrebtii O. 2013 Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimate. *Math. Biosci.* **246**, 283–292. (doi:10.1016/j.mbs.2013.03.011)

29. Eisenberg MC, Hayashi MAL. 2014 Determining identifiable parameter combinations using subset profiling. *Math. Biosci.* **256**, 115–126. (doi:10.1016/j.mbs.2014.08.008)

30. Fröhlich F, Theis FJ, Hasenauer J. 2014 Uncertainty analysis for non-identifiable dynamical systems: profile likelihoods, bootstrapping and more. In *Int. Conf. on Computational Methods in Systems Biology, 17–19 November*, pp. 61–72. Cham, Switzerland: Springer. (doi:10.1007/978-3-319-12982-2_5)

31. He D, Ionides EL, King AA. 2010 Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J. R. Soc. Interface* **7**, 271–283. (doi:10.1098/rsif.2009.0151)

32. Wieland F-G, Hauber AL, Rosenblatt M, Tönsing C, Timmer J. 2021 On structural and practical identifiability. *Curr. Opin. Syst. Biol.* **25**, 60–69. (doi:10.1016/j.coisb.2021.03.005)

33. Simpson MJ, Baker RE, Vittadello ST, Maclaren OJ. 2020 Practical parameter identifiability for spatio-temporal models of cell invasion. *J. R. Soc. Interface* **17**, 20200055. (doi:10.1098/rsif.2020.0055)

34. Siekmann I, Sneyd J, Crampin EJ. 2012 MCMC can detect nonidentifiable models. *Biophys. J.* **103**, 2275–2286. (doi:10.1016/j.bpj.2012.10.024)

35. Hines KE, Middendorf TR, Aldrich RW. 2014 Determination of parameter identifiability in nonlinear biophysical models: a Bayesian approach. *J. Gen. Physiol.* **143**, 401–406. (doi:10.1085/jgp.201311116)

36. Raue A, Kreutz C, Theis FJ, Timmer J. 2013 Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Philos. Trans. R. Soc. A.* **371**, 20110544. (doi:10.1098/rsta.2011.0544)

37. Australian Institute of Marine Science (AIMS) and Queensland University of Technology (QUT). 2021 Identification of two-phase coral reef recovery patterns. See https://doi.org/10.25845/ad6j-zm19 (accessed 7 July 2022).

38. Johnson SG. 2022 The NLopt module for Julia. See https://github.com/JuliaOpt/NLopt.jl (accessed 7 July 2022).

39. Maclaren OJ, Nicholson R. 2020 What can be estimated? Identifibiliy, estimability, casual inference and ill-posed inverse problems. (http://arxiv.org/abs/1904.02826)

40. Cox DR. 2006 *Principles of statistical inference*. Cambridge, UK: Cambridge University Press.

41. Casella G, Berger R. 2001 *Statistical inference*. Belmont, CA: Duxbury.

42. Pace L, Salvan A. 1997 *Principles of statistical inference from a Neo-Fisherian perspective*. Singapore: World Scientific.

43. Royston P. 2007 Profile likelihood for estimation and confidence intervals. *Stata J.* **7**, 376–387. (doi:10.1177/1536867X0700700305)

44. Wu D, Petousis-Harris H, Paynter J, Suresh V, Maclaren OJ. 2022 Likelihood-based estimation and prediction for a measles outbreak in Samoa. (http://arxiv.org/abs/2103.16058)

45. Bjornstad JF. 1990 Predictive likelihood: a review. *Stat. Sci.* **5**, 242–254. (doi:10.1214/ss/1177012175)

46. Akaike H. 1974 A new look at the statistical model identification. *IEEE T. Automat. Cont.* **19**, 716–723. (doi:10.1109/TAC.1974.1100705)

47. Warne DJ, Baker RE, Simpson MJ. 2018 Using experimental data and information criteria to guide model selection for reaction-diffusion problems in mathematical biology. *Bull. Math. Biol.* **81**, 1760–1804. (doi:10.1007/s11538-019-00589-x)

48. Gerlee P. 2013 The model muddle: in search of tumor growth laws. *Cancer Res.* **73**, 2407–2411. (doi:10.1158/0008-5472.CAN-12-4355)

49. Sarapata EA, de Pillis LG. 2014 A comparison and catalog of intrinsic tumor growth models. *Bull. Math. Biol.* **76**, 2010–2024. (doi:10.1007/s11538-014-9986-y)

50. Simpson MJ, Landman KA, Bhaganagarapu K. 2007 Coalescence of interacting cell populations. *J. Theor. Biol.* **247**, 525–543. (doi:10.1016/j.jtbi.2007.02.020)

51. Hafner M, Niepel M, Chung M, Sorger PK. 2016 Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods.* **13**, 521–527. (doi:10.1038/nmeth.3853)

52. Wilson KL, Honsey AE, Moe B, Venturelli P. 2017 Growing the biphasic framework: techniques and recommendations for fitting emerging growth models. *Methods Ecol. Evol.* **9**, 822–833. (doi:10.1111/2041-210X.12931)

53. Monod J. 1949 The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394. (doi:10.1146/annurev.mi.03.100149.002103)

54. King AA, Nguyen D, Ionides EL. 2016 Statistical inference for partially observed Markov processes via the R package pomp. *J. Stat. Softw.* **69**, 1–43. (doi:10.18637/jss.v069.i12)

55. Renardy M, Kirschner D, Eisenberg M. 2022 Structural identifiability analysis of age-structured PDE epidemic models. *J. Math. Biol.* **84**, 9. (doi:10.1007/s00285-021-01711-1)

56. Murphy RJ, Maclaren OJ, Calabrese AR, Thomas PB, Warne DJ, Williams ED, Simpson MJ. 2022 Computationally efficient framework for diagnosing, understanding, and predicting biphasic population growth. Figshare. (doi:10.6084/m9.figshare.c.6315651)