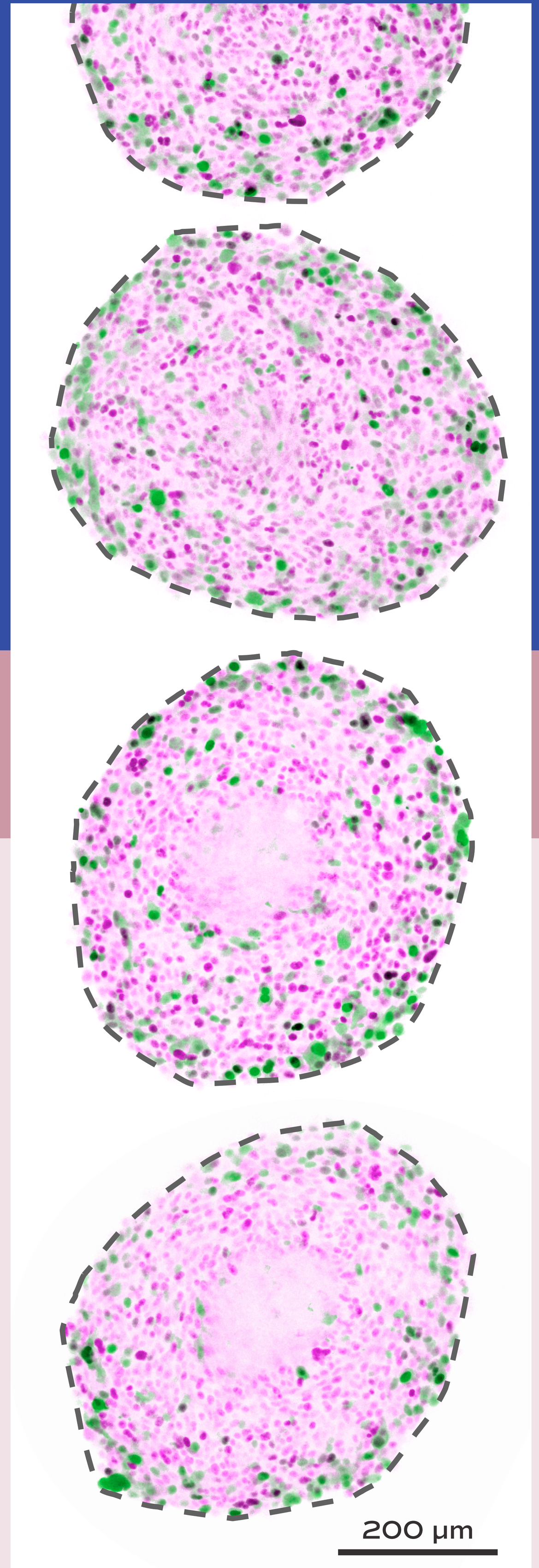
Modelling heterogenity with random parameters

Alexander P Browning[1,2,3], Christopher Drovandi[1,2], Ian W Turner[1], Adrianne L Jenner[1,2], Christopher Drovandi[1,2], and Matthew J Simpson[1,2]

[1] School of Mathematical Sciences, Queensland University of Technology
[2] Centre for Data Science, Queensland University of Technology
[3] Mathematical Institute, University of Oxford

Heterogeneity in biology



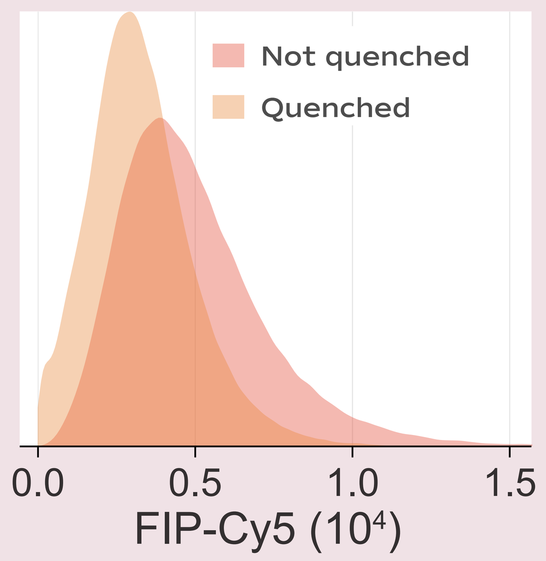


Figure 1
Flourescent measurements relating to amount of internalised molecules in ~10 000 C1R cells.

Figure 2
Identically prepared tumour spheroids from an *in vitro* experiment. Melanoma cells are induced with fluorescent cell cycle indicators (green indicates proliferation, magenta indicates cells in gap 1).

- **Biological variability** is seen even in **tightly controlled experiments** (figs. 1 and 2).
- Such variability is routinely captured in quantitative analysis by coupling a **deterministic mathematical model** to an independent **probabilistic observation process**.
- For example, consider coupling observations of tumour spheroid radius, $r^{\text{obs}}(t)$, to the logistic growth model using normal noise:

$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = \lambda r(t)\left(1 - \frac{r(t)}{R}\right), \qquad r(0) = r_0.$$

$$r^{\text{obs}}(t) = r(t) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2).$$

- The model is capable of recapturing experimental observations (fig. 3).

But what about variability in the initial condition and the dynamical parameters?

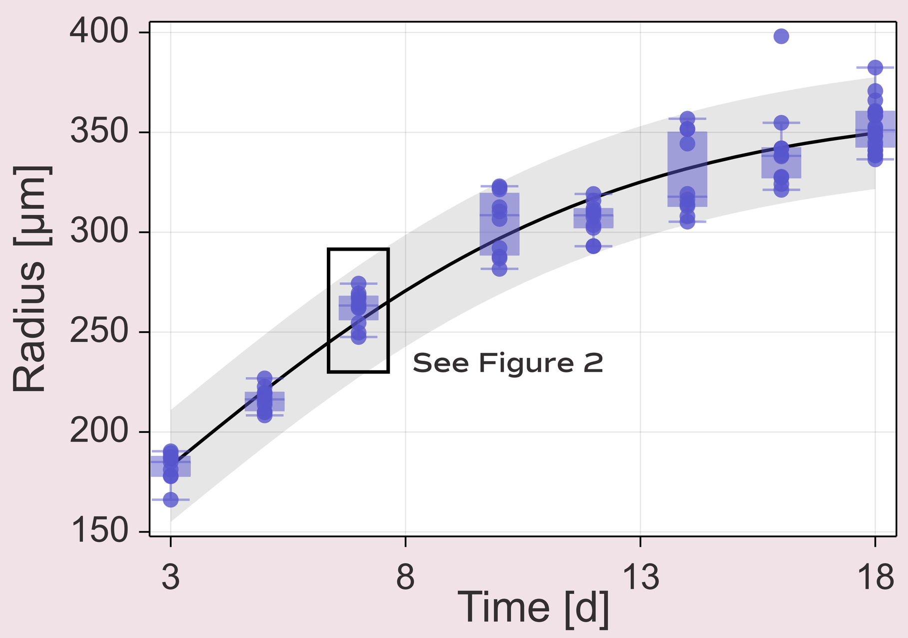


Figure 3
Experimental measurements of temporal tumour spheroid radius described using the logistic growth model with additive normal noise. Shown is experimental data (blue) and model prediction (mean and 95% prediction interval).

Dealing with parameter variability

- We aim to better capture potential sources of variability by describing **parameters and measurements** as **random variables**.
- We take a fast, novel **direct approach** and construct an approximate output distribution for a given model and set of input (parameter) distributions.
- Consider that the mathematical model maps inputs (i.e., parameters), $\boldsymbol{\theta}$, to a set of outputs, $\boldsymbol{f}(\boldsymbol{\theta})$. Then we wish to approximate the distribution of $\boldsymbol{f}(\boldsymbol{\theta})$.

One-dimensional example

- Consider that both input and output are one-dimensional, then **expectation of the Taylor expansion** of $f(\theta)$ about $\hat{\theta} = \langle\theta\rangle$ yields

$$\langle f(\theta)\rangle \approx f(\hat{\theta}) + \frac{\mathrm{d}f(\hat{\theta})}{\mathrm{d}\theta}\underbrace{\langle\theta - \hat{\theta}\rangle}_{=0} + \frac{1}{2}\frac{\mathrm{d}^2 f(\hat{\theta})}{\mathrm{d}\theta^2}\underbrace{\langle\theta - \hat{\theta}^2\rangle}_{=\mathrm{Var}(\theta)}.$$

- We can do this for inputs and outputs of **any dimension** yielding **analytical approximations for the mean, covariance, and skewnesses** of $\boldsymbol{f}(\boldsymbol{\theta})$ in terms of the moments of $\boldsymbol{\theta}$.
- We then approximate the distribution of $\boldsymbol{f}(\boldsymbol{\theta})$ using normal and gamma distributions with the same moments (fig. 4).

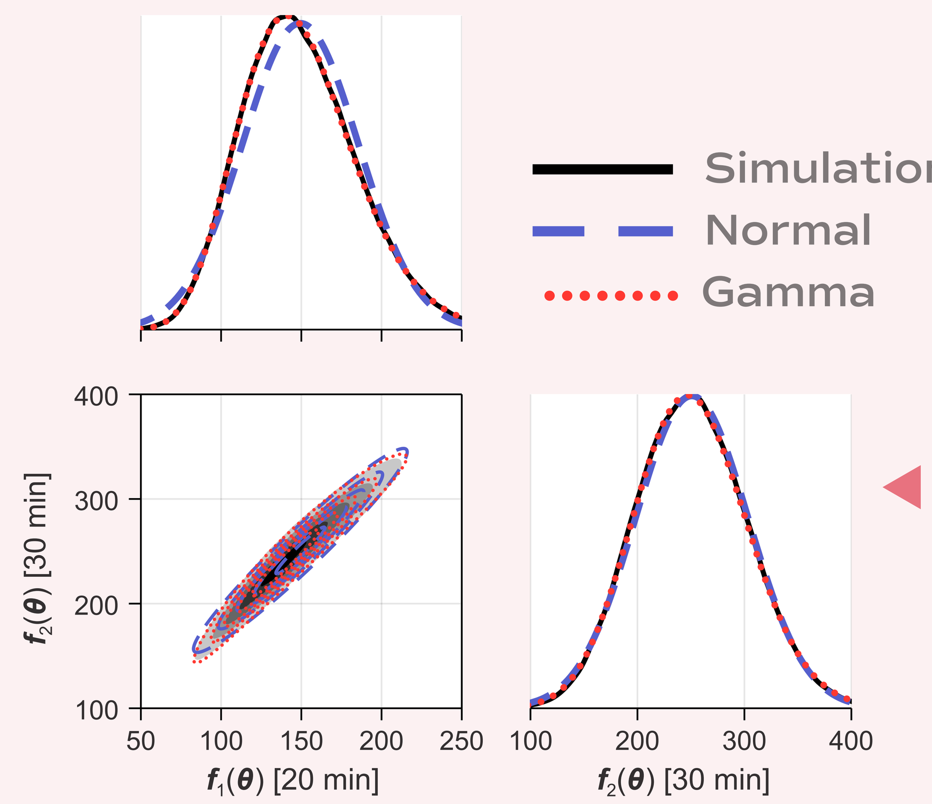


Figure 4
Approximate solution to the logistic model with random parameters using a moment-matched normal distribution (blue; two moments) and gamma distribution (red; three moments).

Parameter inference

- We can **directly construct a likelihood** using our approximate output distribution, which describes the data distribution for a given set of input distributions and allows application of any likelihood-based inference tool (here, we use profile likelihood).

Random parameter logistic model

- Consider the logistic model with parameter distribution

$$r_0 \sim \mathcal{N}\left(\mu_{r_0}, \sigma_{r_0}^2\right), \quad \lambda \sim \mathcal{N}\left(\mu_\lambda, \sigma_\lambda^2\right),$$
$$R \sim \mathcal{N}\left(\mu_R, \sigma_R^2\right), \quad \varepsilon \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right),$$

where $\boldsymbol{\xi} = \left[\mu_{r_0}, \mu_\lambda, \mu_R, \ln\sigma_{r_0}, \ln\sigma_\lambda, \ln\sigma_R, \ln\sigma_\varepsilon\right]^\intercal$ is a vector of unknown "hyperparameters" that we wish to infer.

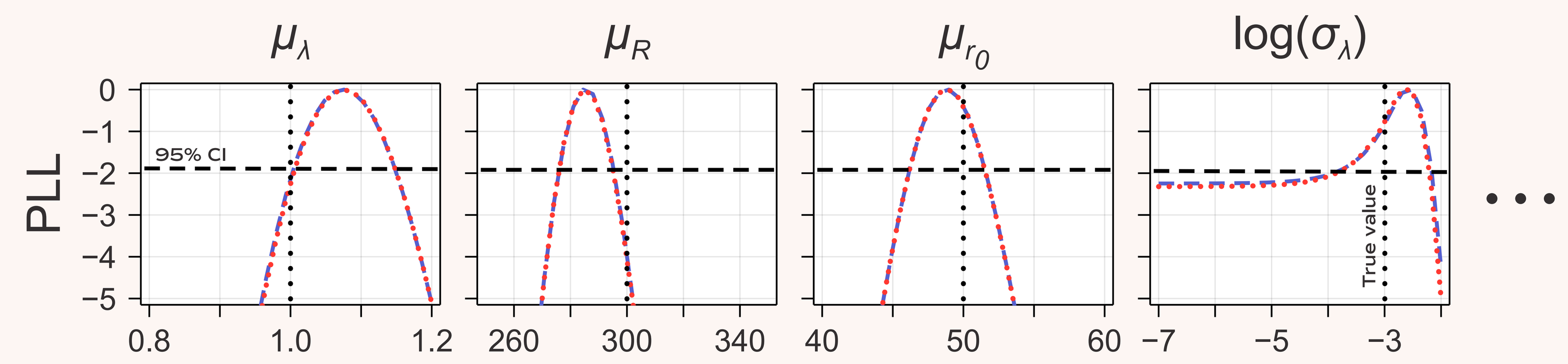


Figure 5
Profile likelihoods for random parameter logistic model using synthetic data (fig. 6). Horizontal line shows the threshold for a 95% confidence interval.

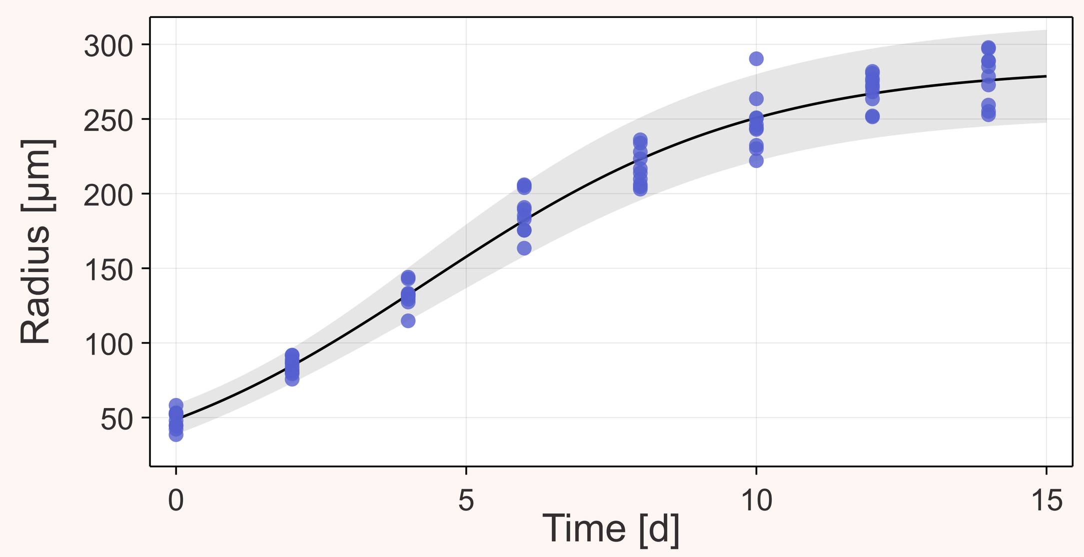


Figure 6
Synthetic data and predictive distribution (mean and 95% prediction interval) for the random parameter logistic model.

- In addition to inferring parameters and quantifying uncertainty, we much more accurately capture the variability in the data (fig. 6).
- **Julia module** for implementation can be applied to any differential-equation model, for a wide variety of input distributions (correlated, bimodal, skewed, etc).