



**Cite this article:** Simpson MJ, Baker RE, Vittadello ST, Maclare OJ. 2020 Practical parameter identifiability for spatio-temporal models of cell invasion. *J. R. Soc. Interface* **17**: 20200055.  
<http://dx.doi.org/10.1098/rsif.2020.0055>

Received: 22 January 2020

Accepted: 12 February 2020

### Subject Category:

Life Sciences—Mathematics interface

### Subject Areas:

biophysics

### Keywords:

identifiability analysis, Bayesian inference, profile likelihood, reaction–diffusion, cell cycle

### Author for correspondence:

Matthew J. Simpson

e-mail: matthew.simpson@qut.edu.au

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4862265>.

# Practical parameter identifiability for spatio-temporal models of cell invasion

Matthew J. Simpson<sup>1</sup>, Ruth E. Baker<sup>2</sup>, Sean T. Vittadello<sup>1</sup>  
and Oliver J. Maclare<sup>3</sup>

<sup>1</sup>School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

<sup>2</sup>Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

<sup>3</sup>Department of Engineering Science, University of Auckland, Auckland 1142, New Zealand

MJS, 0000-0001-6254-313X; REB, 0000-0002-6304-9333; STV, 0000-0002-4476-6732

We examine the practical identifiability of parameters in a spatio-temporal reaction–diffusion model of a scratch assay. Experimental data involve fluorescent cell cycle labels, providing spatial information about cell position and temporal information about the cell cycle phase. Cell cycle labelling is incorporated into the reaction–diffusion model by treating the total population as two interacting subpopulations. Practical identifiability is examined using a Bayesian Markov chain Monte Carlo (MCMC) framework, confirming that the parameters are identifiable when we assume the diffusivities of the subpopulations are identical, but that the parameters are practically non-identifiable when we allow the diffusivities to be distinct. We also assess practical identifiability using a profile likelihood approach, providing similar results to MCMC with the advantage of being an order of magnitude faster to compute. Therefore, we suggest that the profile likelihood ought to be adopted as a screening tool to assess practical identifiability before MCMC computations are performed.

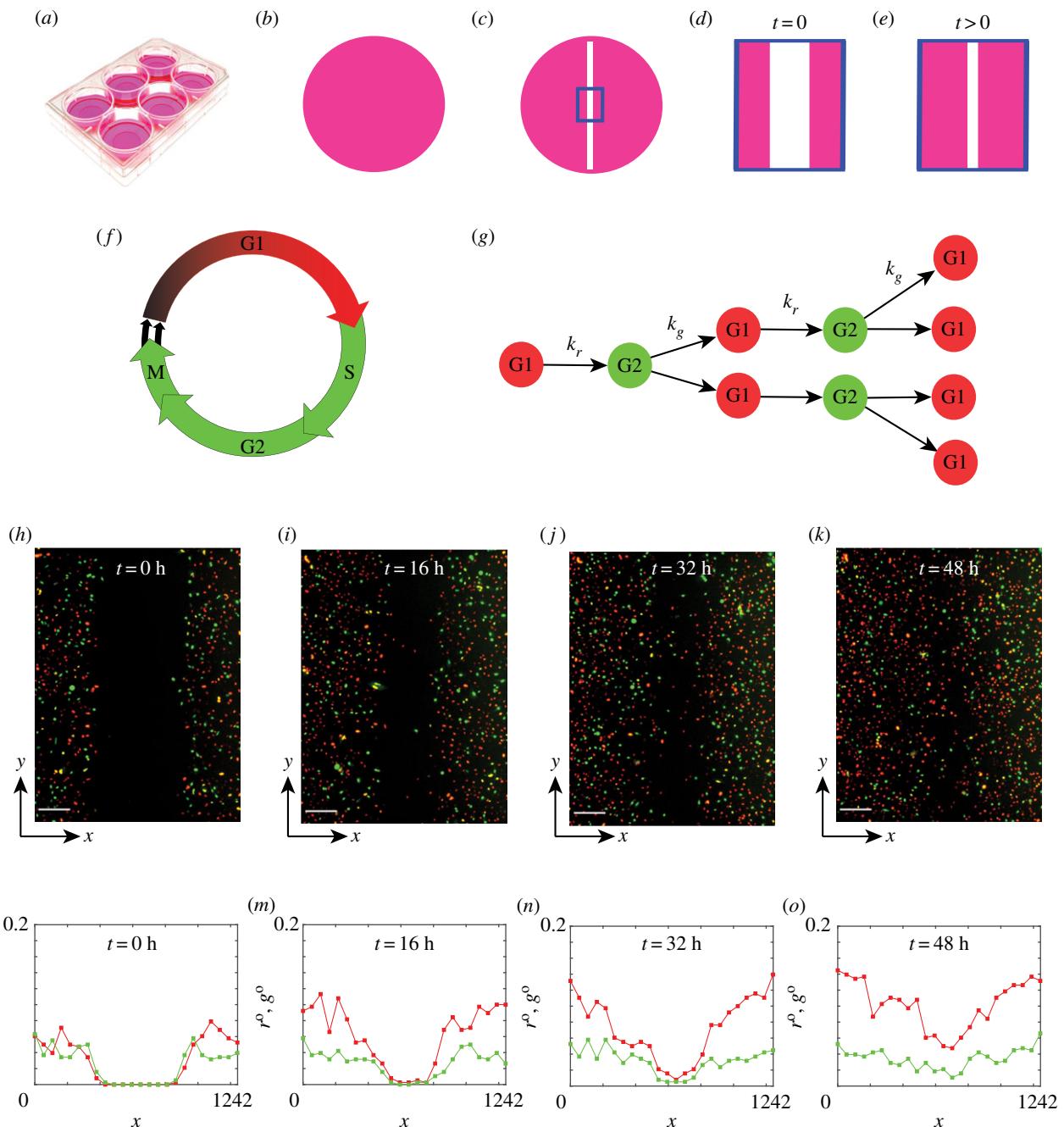
## 1. Introduction

Combined cell migration and cell proliferation leads to moving fronts of cells, often called *cell invasion* [1], which is essential for tissue repair and wound healing [2]. While sophisticated experimental techniques are continually developed to interrogate cell invasion, traditional *scratch assays* remain widely used because they are simple, fast and inexpensive [3]. Scratch assays involve growing (figure 1*a,b*) and then scratching (figure 1*c*) a cell monolayer, and imaging a small region around the scratch as the wounded area closes (figure 1*d,e*).

The first mathematical models to interpret experimental observations of cell invasion were reaction–diffusion models [5]. Sherratt & Murray [6] modelled a wound-healing-type assay using the Fisher–Kolmogorov model, and many subsequent studies have related solutions of mathematical models to experimental observations, including Maini *et al.* [2], Sengers *et al.* [7] and Nardini *et al.* [8]. By matching the solution of a mathematical model with certain experimental observations, these studies provide both parameter estimates and mechanistic insight.

One challenge of using scratch assays is that there is no standard, widely accepted experimental protocol. There are many differences in: (i) the initial monolayer density [9]; (ii) the width and shape of the scratch [1,10]; and (iii) the experimental time scale [3]. These differences make it difficult to compare different experiments, and it is unclear whether different mathematical models of varying complexity (i.e. varying numbers of unknown parameters or varying the model structure) are required under different experimental protocols.

It is reasonable to argue that simple experimental data ought to be modelled using simple mathematical models with a small number of unknown parameters, and that more sophisticated mathematical models with additional parameters should be used only when greater amounts of experimental data are available. However, it is difficult to use this kind of qualitative reasoning



**Figure 1.** Experiments in a six-well plate (*a*) are initiated with a uniform cell monolayer (*b*). A scratch is made (*c*), and the  $1296 \mu\text{m} \times 1745 \mu\text{m}$  field of view is observed between  $t=0$  h (*d*) and  $t=48$  h (*e*). The cell cycle is labelled using fluorescent ubiquitination-based cell cycle indicator (FUCCI) [4] so that cells in G1 phase fluoresce red and cells in S/G2/M phase fluoresce green (*f*). A freely cycling cell in G1 phase (red) will transition into S/G2/M phase (green) at rate  $k_r > 0$ . A freely cycling cell in S/G2/M phase (green) will divide into two cells in G1 phase (red) at rate  $k_g > 0$ , provided there is sufficient space to accommodate division (*g*). Images of the field of view at  $t=0$ , 16, 32 and 48 h (*h*–*k*), with the non-dimensional densities of G1 (red) and S/G2/M (green) subpopulations shown as a function of space and time (*l*–*o*). Scale bars in (*h*–*k*) show 200  $\mu\text{m}$ .

to justify particular decisions about how complicated a mathematical model ought to be. Methods of *identifiability analysis* [11–21] provide more systematic means to determine the appropriate model complexity relative to available data. A model is *identifiable* when distinct parameter values imply distinct distributions of observations, and hence when it is possible to uniquely determine the model parameters using an infinite amount of ideal data [22–25]. In the systems biology literature that focuses on models formulated as ordinary differential equations, *identifiability* has also been referred to as *structural identifiability* [19,20]. Working with ordinary differential equations that describe temporal processes, it is possible to formally analyse the structural identifiability [19,20,26,27]. By contrast, here we work with

reaction–diffusion partial differential equations since we are interested in spatio-temporal processes. In this case, formal analysis of structural identifiability is more difficult and so we explore structural identifiability numerically (electronic supplementary material).

In contrast to predictive performance measures, such as Akaike's information criterion (AIC) [28], identifiability analysis focuses on distinguishing parameter values and, as a result, understanding underlying mechanisms. Shmueli [29] characterizes this general difference as the choice of whether to 'explain or predict'. The formal definition of identifiability requires an infinite amount of ideal data. Thus, in addition to the strict technical concept of (structural) identifiability, the terms *practical identifiability* [16] and *estimability* [22,25]

are used to describe whether it is possible to provide reasonably precise estimates using finite, non-ideal data.

Since practical identifiability is a finite, non-ideal data concept, it depends more strongly on the inferential framework adopted and is less clearly defined in the literature than structural identifiability. We consider practical identifiability analysis to encompass a set of *precision-based* pre-data planning and post-data analysis tools similar to those advocated by Cox [30], Bland [31] and Rothman & Greenland [32]. This is consistent with the systems biology literature [11–18,21], where it is typically characterized in terms of the ability to provide finite, sufficiently precise interval estimates at various levels of confidence or credibility given finite amounts of non-ideal, real data. This is the approach we take here; we consider both the typical precision achieved under repeated simulation of ideal synthetic data (electronic supplementary material) and the post-data, realized precision given real data. It is important to note that, if a model is found to be practically non-identifiable, this does not mean that the model is incorrect. Instead it simply indicates that additional model complexity is not required, and that such a model would be difficult to falsify in realistic experiments.

In systems biology, experimental data often take the form of time series describing temporal variations of molecular species in a chemical reaction network, and these data are modelled using ordinary or stochastic differential equations [33]. Here, we focus on applying methods of practical identifiability analysis to spatio-temporal partial differential equation models, which represent both spatial and temporal variations in quantities of interest. It is instructive to consider both a Bayesian inference framework [14,15] and profile-likelihood analysis [16,17].

Bayesian inference is widely adopted within the mathematical biology community [34–39]. A drawback of the Bayesian approach is that it can be computationally expensive, particularly when Markov chain Monte Carlo (MCMC) methods are used to sample the distributions of interest. By contrast, profile likelihood analysis is a more standard tool of statistical inference [30,40], but is far less familiar within the mathematical biology community. Being optimization-based, profile likelihood can be computationally inexpensive.

Here, we consider data from a scratch assay where cells are labelled using two fluorescent probes to show real-time progression through the cell cycle [4]. Such labelling allows us to describe the total population of cells as two subpopulations according to the two cell cycle labels [41]. This experimental approach provides far more information than a standard scratch assay without cell cycle labels. We model the experiments using a system of reaction-diffusion equations that are an extension of the Fisher–Kolmogorov model. Using these data and this model, we explore the practical identifiability of the model parameters using both Bayesian inference and profile likelihood.

One of the main outcomes of this work is that, instead of performing potentially computationally expensive Bayesian inference to assess practical identifiability and determine parameter estimates, we show that a profile likelihood analysis can provide a fast and reliable preliminary assessment of practical identifiability. Both analysis methods provide good insight into the identifiability of the spatio-temporal models considered here. In summary, we consider minimal model scratch assay data generated using fluorescent cell cycle labelling and show that it is practically identifiable, whereas a simple extension of that model is not, despite the fact that numerical experiments indicate that it is structurally identifiable in the limit of infinite

data. In all cases, the Bayesian analysis is consistent with profile likelihood results. However, the profile likelihood analysis is an order of magnitude faster to implement.

## 2. Experimental data

We consider a scratch assay performed with the 1205Lu melanoma cell line with fluorescent ubiquitination-based cell cycle indicator (FUCCI) labelling [4]. The scratch assay is initiated by seeding 1205Lu melanoma cells into six-well plates and allowing sufficient time for the cells to attach, to begin to proliferate (figure 1a), and to form an approximately uniform monolayer (figure 1b). A scratch is made in the monolayer (figure 1c) and a small region of the experiment is imaged (figure 1c,d) [41]. The geometry of the experiment means that the distribution of cells in the imaged region is independent of vertical position, and so we treat the cell density as a function of horizontal position,  $x$ , and time,  $t$  [41]. The boundaries around the imaged region are not physical boundaries and cells are free to migrate across these boundaries. Since the density of cells is approximately constant away from the scratch, the net flux of cells locally across each boundary of the imaged region is approximately zero [41].

Cells in G1 phase fluoresce red and cells in S/G2/M phase fluoresce green (figure 1f). We represent FUCCI labelling by treating the total population as two subpopulations: (i) red cells in G1 and (ii) green cells in S/G2/M. Cells in G1 (red) transition into S/G2/M cells (green) at rate  $k_r > 0$ . This red-to-green transition does not involve any change in cell number so we assume this transition is unaffected by the availability of physical space. Cells in S/G2/M (green) undergo mitosis at rate  $k_g > 0$  to produce two daughter cells in G1 (red) (figure 1f). Since the green-to-red transition involves the production of two daughter cells we assume this transition only occurs provided there is sufficient space.

Images of the experiment (figure 1h–k) show that cells move into the initially scratched region while simultaneously progressing through the cell cycle. At the beginning of the experiment, the scratch is approximately 500  $\mu\text{m}$  wide, and by the end of the experiment the initially vacant wound space becomes occupied as cells have migrated into the gap. Visual examination of the density of cells away from the scratched region shows that the cell density increases with time, and this is driven by cell proliferation. To quantify these visual observations we note that each experimental image is 1296  $\mu\text{m}$  wide, and we discretize each image into 24 equally spaced columns of width 54  $\mu\text{m}$ . We count the number of red cells and green cells per column and divide by the area of the column and by the theoretical maximum packing density of 0.004 cells  $\mu\text{m}^{-2}$  [41] to give an estimate of the non-dimensional density of red cells and green cells per column. Assuming that each experimental density estimate represents the centre of each column, these data (figure 1l–o) show the spatial variation in density of G1 (red) and S/G2/M (green) cells as a function of space and time.

We summarize the experimental data using the following notation. Our observations are vectors containing the observed non-dimensional density of red cells,  $r_i^o$ , and the observed non-dimensional density of green cells,  $g_i^o$ , per column at each position  $x_i$  and time  $t_i$ , i.e.

$$(r^o, g^o, x, t)_i = (r_i^o, g_i^o, x_i, t_i). \quad (2.1)$$

We use an explicit superscript ‘o’ to distinguish (noisy) observations of  $r$  and  $g$  from their modelled counterparts introduced below, while we assume that the space and time

coordinates  $x$  and  $t$ , respectively, are noise-free. In particular, we use

$$y_i^o = (r^o, g^o)_i \quad \text{and} \quad s_i = (x_i, t_i) \quad (2.2)$$

to denote the  $i$ th combined vector of observations of  $r$  and  $g$  and the  $i$ th combined vector of spatio-temporal coordinates  $x$  and  $t$ , respectively. Observations are taken at 24 equally spaced spatial locations, where the spacing between the observations is 54  $\mu\text{m}$ , at four equally spaced time points,  $t = 0, 16, 32$  and 4 h. Thus,  $y^o$  consists of  $24 \times 4 = 96$  observations of both of  $r$  and  $g$ , giving  $N = 24 \times 4 \times 2 = 192$  measurements of cell density in total. Experimental data are provided in the electronic supplementary material.

### 3. Mathematical model

We treat our *mathematical model* as having two components: (i) a deterministic spatio-temporal *process* model governing the dynamics of the experiment and (ii) a probabilistic *observation* model. This approach invokes the reasonable assumption that observations are noisy versions of a deterministic latent model [14]. We describe both models and note that the same model components are used in the Bayesian and the profile likelihood analyses.

#### 3.1. Process model

A reasonably simple process model for the FUCCI scratch assay can be written as

$$\begin{aligned} \frac{\partial r(x, t)}{\partial t} &= D_r \frac{\partial^2 r(x, t)}{\partial x^2} - k_r r(x, t) \\ &\quad + 2k_g g(x, t)[1 - r(x, t) - g(x, t)] \end{aligned} \quad (3.1)$$

and

$$\begin{aligned} \frac{\partial g(x, t)}{\partial t} &= D_g \frac{\partial^2 g(x, t)}{\partial x^2} + k_r r(x, t) \\ &\quad - k_g g(x, t)[1 - r(x, t) - g(x, t)], \end{aligned} \quad (3.2)$$

where  $D_r > 0$  is the diffusivity of cells in G1,  $D_g > 0$  is the diffusivity of cells in S/G2/M,  $k_r > 0$  is the rate at which cells in G1 transition into S/G2/M, and  $k_g > 0$  is the rate at which cells in S/G2/M undergo mitosis to produce two cells in G1 [41]. The solution of equations (3.1) and (3.2),  $r(x, t)$  and  $g(x, t)$ , are taken to represent the (continuous) underlying mean densities, while  $r_i^o, g_i^o$  represent noisy observations of these mean densities at the measurement locations  $s_i = (x_i, t_i)$ .

The process model has four parameters:  $\theta = (D_r, D_g, k_r, k_g)$ . Given suitable initial conditions,  $r(x, 0)$  and  $g(x, 0)$ , and parameter values  $\theta$ , equations (3.1) and (3.2) are solved numerically (electronic supplementary material).

Despite the abundance of experimental data in figure 1 and the apparent simplicity of equations (3.1) and (3.2), it is unclear whether these experimental data are sufficient to estimate the four parameters:  $\theta = (D_r, D_g, k_r, k_g)$ . Therefore, we also consider a simpler model whereby we set  $D = D_r = D_g$ , implying that cells in G1 diffuse at the same rate as cells in S/G2/M. This simpler model is characterized by three parameters:  $\theta = (D, k_r, k_g)$ .

#### 3.2. Observation model

We assume that observations  $y_i^o$  are noisy versions of the latent model solutions,  $y(s_i, \theta)$ . A standard approach in the

mathematical biology literature is to further assume that this observation error is independent and identically distributed (iid), and that this noise is additive and normally distributed, with zero mean and common variance  $\sigma^2$  [9,14]. Therefore, we have

$$y_i^o | \theta \sim \mathcal{N}(y(s_i, \theta), \sigma^2 I), \quad (3.3)$$

where  $I$  is the  $2 \times 2$  identity matrix. Since our main interest is in  $\theta = (D_r, D_g, k_r, k_g)$ , we directly specify  $\sigma = 0.05$  [9,14], and we will discuss this choice later.

### 4. Practical identifiability analysis

We consider two approaches for practical identifiability analysis. The first is based on Bayesian inference and MCMC, and follows the approach outlined by Hines *et al.* [14]. The second approach is based on profile likelihood and follows Raue *et al.* [16]. Both approaches have been considered in the context of identifiability of non-spatial models, and, as we will show, the same ideas can deal with models of spatio-temporal processes.

As discussed by Raue *et al.* [42], estimation results from MCMC and profile likelihood are typically similar in the case of identifiable parameters (and in the low-dimensional setting), but can be very different in the presence of non-identifiability. In particular, Raue *et al.* [42] demonstrate that results from MCMC can be potentially misleading in the presence of non-identifiability. On the other hand, Hines *et al.* [14] and Seikmann *et al.* [15] argue that appropriate MCMC diagnostics can indicate the presence of non-identifiability. Hence we use both methods and consider the extent to which they agree.

#### 4.1. Parameter bounds

Before outlining the identifiability methodologies, we note that the parameters in equations (3.1) and (3.2) have a physical interpretation and we can formulate some biologically motivated bounds. Practical identifiability, given observed data, can be evaluated by comparing the width of (realized) interval estimates relative to these simple bounds.

Previous estimates of melanoma cell diffusivity in similar experiments found that typical values are often less than  $1000 \mu\text{m}^2 \text{h}^{-1}$  [43], so we take conservative bounds on  $D_r$  and  $D_g$  to be  $0 < D_r, D_g < 2000 \mu\text{m}^2 \text{h}^{-1}$ . Bounds on  $k_r$  and  $k_g$  can be inferred from previous experimental measurements. The duration of time 1205Lu melanoma cells remain in the G1 phase varies between 8 and 30 h, whereas the time that 1205Lu melanoma cells remain in the S/G2/M phase varies between 8 and 17 h [4]. These measurements imply  $0.033 < k_r < 0.125/\text{h}$  and  $0.059 < k_g < 0.125/\text{h}$ , so we take conservative bounds on  $k_r$  and  $k_g$  to be  $0 < k_r, k_g < 0.2/\text{h}$ .

#### 4.2. Bayesian inference

Following Hines *et al.* [14], we take a Bayesian approach to assess parameter identifiability. Bayesian inference relies on Bayes' theorem, written here as

$$p(\theta | y^o) \propto p(y^o | \theta)p(\theta), \quad (4.1)$$

where  $p(y^o | \theta)$  is the *likelihood function*,  $p(\theta)$  is the *prior* and  $p(\theta | y^o)$  is the *posterior*. The posterior distribution is the inferential target; this summarizes the information about the parameter  $\theta$  in light of the observed data,  $y^o$ , and the prior

information specified by  $p(\theta)$ . The likelihood function represents the information contributed by the data, and corresponds to the probabilistic observation model in (3.3) evaluated at the observed data.

We explore the parameter space  $\theta$  by sampling its posterior distribution using a Metropolis–Hastings MCMC algorithm [14]. The Markov chain starts at position  $\theta_i$ , and a potential move to  $\theta^*$  is accepted with probability  $\alpha$ ,

$$\alpha = \min \left[ 1, \frac{p(\theta^* | y^o)}{p(\theta_i | y^o)} \right]. \quad (4.2)$$

This Markov chain tends to move towards regions of high posterior probability, but it also provides a mechanism to move away from local minima by allowing transitions to regions of lower posterior probability. The Markov chain produced by this algorithm explores the parameter space in proportion to the posterior probability and provides a finite number of independent, identically distributed samples from the posterior distribution. Proposals in the MCMC algorithm are made by sampling a multivariate normal distribution with zero mean and covariance matrix  $\Sigma$ . When  $\theta = (D_r, D_g, k_r, k_g)$  we specify  $\Sigma = \text{diag}(10^2, 10^2, 10^{-6}, 10^{-6})$ , whereas when  $\theta = (D, k_r, k_g)$  we specify  $\Sigma = \text{diag}(10^2, 10^{-6}, 10^{-6})$ .

Poor identifiability in a Bayesian setting using MCMC is typically characterized by poorly converging chains, label-switching, multimodal or overly broad distributions, and similar phenomena that can be diagnosed either graphically or by computing various diagnostic statistics [14,15]. We discuss these diagnostic methods in more detail in the Results and discussion section.

### 4.3. Profile likelihood

Here we describe how a profile likelihood identifiability analysis can be undertaken. This is based on the same likelihood as above,  $p(y^o | \theta)$ , but here we will present results in terms of the normalized likelihood function, denoted

$$\mathcal{L}(\theta; y^o) = \frac{p(y^o | \theta)}{\sup_{\theta} p(y^o | \theta)}. \quad (4.3)$$

This is considered as a function of  $\theta$  for fixed data  $y^o$ . It is also common to take the likelihood as only defined up to a constant, but here we fix the proportionality constant by presenting results in terms of the normalized likelihood function as above.

We assume our full parameter  $\theta$  can be partitioned into an *interest* parameter  $\psi$  and a *nuisance* parameter  $\lambda$ , i.e.  $\theta = (\psi, \lambda)$ . These can also be considered functions  $\psi(\theta)$  and  $\lambda(\theta)$  of the full parameter  $\theta$ , specifying the partition components. Here we will restrict attention to a scalar interest parameter and a vector nuisance parameter, i.e. multiple scalar nuisance parameters. Then the profile likelihood for the interest parameter  $\psi$  can be written as [30,40]

$$\mathcal{L}_p(\psi; y^o) = \sup_{\lambda} \mathcal{L}(\psi, \lambda; y^o), \quad (4.4)$$

where  $\lambda$  is ‘optimized out’ for each value of  $\psi$ , and this implicitly defines a function  $\lambda^*(\psi)$  of optimal  $\lambda$ -values for each value of  $\psi$ . For example, given the full parameter  $\theta = (D_r, D_g, k_r, k_g)$ , we may consider the diffusion constant of red cells as the interest parameter and the other parameters as nuisance parameters, i.e.  $\psi(\theta) = D_r$  and  $\lambda(\theta) = (D_g, k_r, k_g)$ . This would give

$$\mathcal{L}_p(D_r; y^o) = \sup_{(D_g, k_r, k_g)} \mathcal{L}(D_r, D_g, k_r, k_g; y^o). \quad (4.5)$$

Given the assumption of normally distributed errors, and assuming a fixed, common standard deviation  $\sigma$ , profiling reduces to solving a series of nonlinear least-squares problems, one for each value of the parameter of interest, by using log-density functions. Additional details and example calculations are provided in the electronic supplementary material.

We implement this optimization using Matlab’s *lsqnonlin* function with the ‘trust-region-reflective’ algorithm to implement bound constraints [44]. For each value of the interest parameter, taken over a sufficiently fine grid, the nuisance parameter is optimized out and the previous optimal value is used as the starting guess for the next optimization problem. Uniformly spaced grids of 50 points, defined between previously discussed lower and upper bounds, are used. Results are plotted in terms of the normalized profile likelihood functions; cubic interpolations are used to define the profiles for all parameter values. Solutions did not exhibit any notable dependence on initial guesses.

The likelihood function is often characterized as representing the information that the data contain about the parameters, and the relative likelihood for different parameter values as indicating the relative evidence for these parameter values [40,45,46]. As such, a flat profile is indicative of non-identifiability, therefore a lack of information in the data about a parameter [16]. In general, the degree of curvature is related to the inferential precision [16,17,42]. Likelihood-based confidence intervals can be formed by choosing a threshold-relative profile likelihood value, which can be approximately calibrated via the  $\chi^2$ -distribution (or via simulation). We use a threshold of 0.15 as a reference, which corresponds to an approximate 95% confidence interval for sufficiently regular problems [40]. The points of intersection were determined using the interpolated profile likelihood functions.

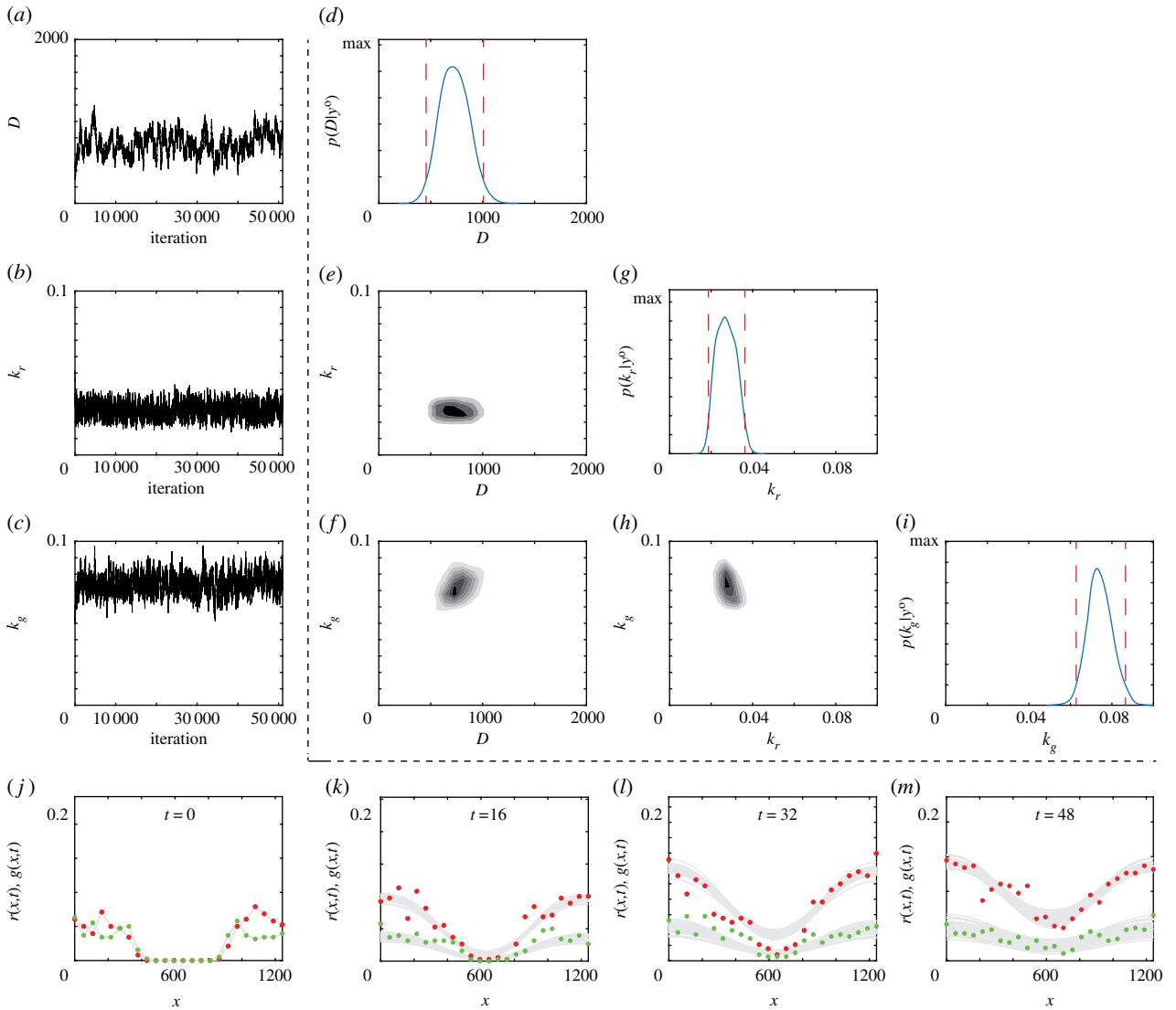
## 5. Results and discussion

Here we consider the results of the above methods of identifiability analysis under the two scenarios introduced previously. In scenario 1, we consider equations (3.1) and (3.2) under the assumption that both subpopulations have the same diffusivity,  $D_r = D_g = D$ , so that  $\theta = (D, k_r, k_g)$ . In scenario 2, we consider the same model without this assumption so that  $\theta = (D_r, D_g, k_r, k_g)$ . We first consider the Bayesian analysis using MCMC, and then the profile likelihood analysis. All units of diffusivities are micrometres squared per hour and all units for rate parameters are per hour.

### 5.1. Bayesian analysis

As we are focused on the question of identifiability, here we use uniform priors on all parameters. This is a natural choice when we are interested in identifiability since we want to focus on the information about the parameter values inherent within the data rather than from some imposed prior [14].

In the first scenario where  $D_r = D_g = D$ , we see rapid convergence of the Markov chain for  $D$ ,  $k_r$  and  $k_g$  (figure 2a–c). Importantly, after the Markov chain moves away from the initial location,  $\theta_0$ , it remains within the previously stated conservative bounds. Additional results in the electronic supplementary material confirm similar results for different  $\theta_0$ . A plot matrix representation of the univariate and bivariate

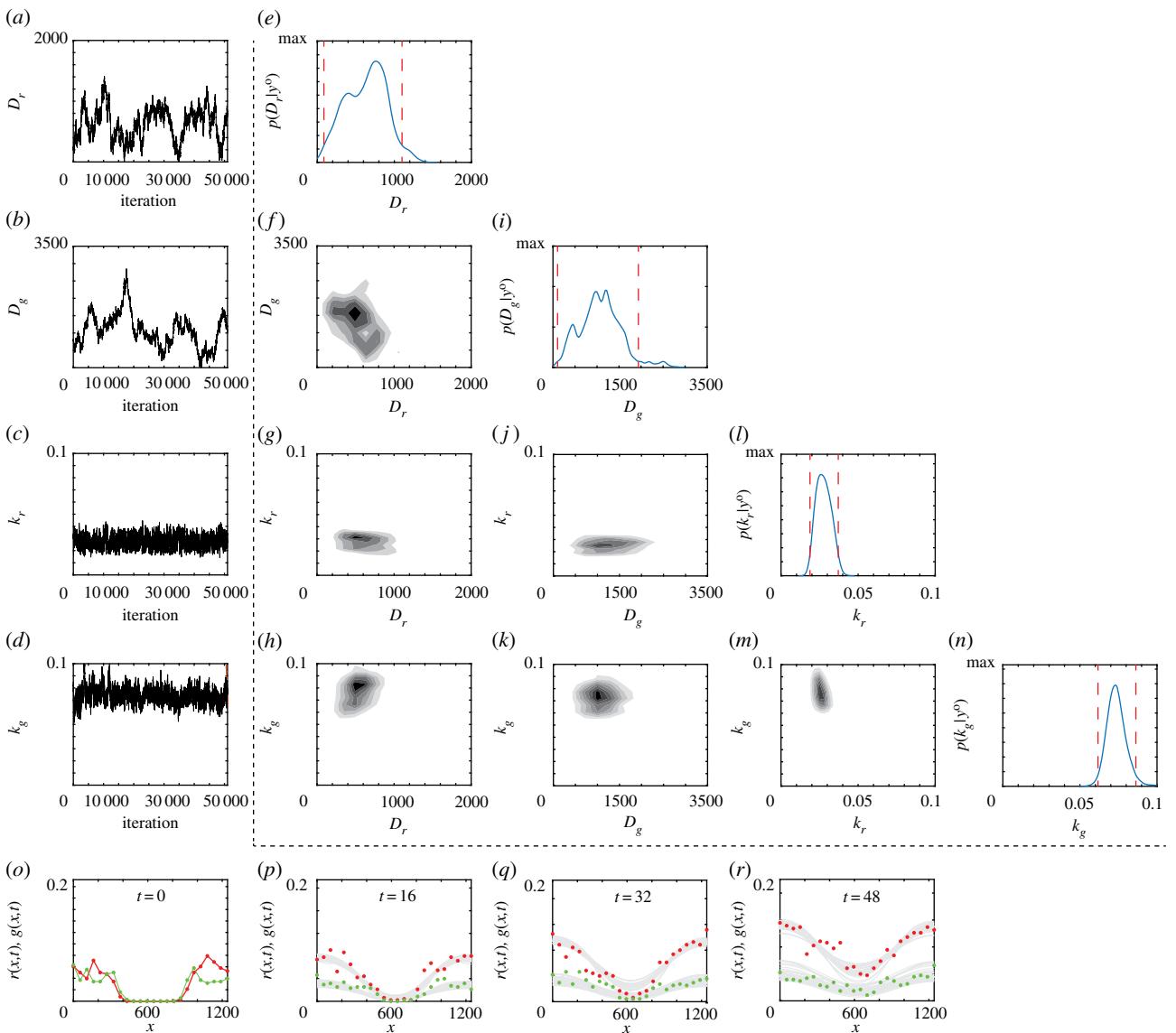


**Figure 2.** Typical Markov chain iterations, of length 51 000, for  $D$ ,  $k_r$  and  $k_g$  in (a)–(c), respectively. In this case, the Markov chain is initiated with  $\theta_0 = (500, 0.05, 0.05)$ . Results in (d)–(i) show a plot matrix representation of the univariate marginals and bivariate marginals estimated using the final 50 000 iterations of the Markov chain in (a)–(c). For the univariate distribution, the posterior modes are  $\bar{D} = 707 \mu\text{m}^2/\text{h}$ ,  $\bar{k}_r = 0.0266/\text{h}$  and  $\bar{k}_g = 0.0727/\text{h}$ , and the 95% credible intervals are  $D \in [455, 1007]$ ,  $k_r \in [0.0187, 0.0362]$  and  $k_g \in [0.0628, 0.0727]$ . In the univariate marginals, the 95% credible intervals are shown in red dashed vertical lines; in the bivariate marginals the region of maximum density is shown in the darkest shade. Results in (j)–(m) superimpose 30 solutions of equations (3.1) and (3.2) where  $\theta$  is randomly sampled from the Markov chain. MCMC results use  $\sigma = 0.05$ .

marginal densities is given (figure 2d–i) [47], where the univariate marginals (figure 2d,g,i) are unimodal and approximately symmetric. The univariate posterior modes are  $\bar{\theta} = (705, 0.0266, 0.0727)$ , confirming that each component lies within the biologically feasible range, and the univariate 95% credible intervals are relatively narrow. A posterior predictive check [47] with 30 randomly sampled parameter choices from the converged region of the Markov chain provides visual confidence in the MCMC results and the appropriateness of the model (figure 2j–m) [47]. Additional comparison of the experimental data and solution of the model parametrized with  $\bar{\theta}$  also confirms this (electronic supplementary material). In summary, setting  $D = D_r = D_g$ , our model parameters appear to be identifiable according to the Bayesian approach.

In the second scenario, where  $D_r \neq D_g$ , the Markov chains for  $D_r$ ,  $D_g$ ,  $k_r$  and  $k_g$  (figure 3a–d) are very different from the first scenario. The Markov chains for  $k_r$  and  $k_g$  remain within biologically feasible bounds. However, the Markov chains for  $D_r$  and  $D_g$  often move beyond the previously stated

conservative bounds with no obvious convergence. Additional results for different choices of  $\theta_0$  (electronic supplementary material) indicate that these poor results for  $D_r$  and  $D_g$  are consistent across a number of choices of  $\theta_0$ . A plot matrix representation confirms that we obtain well-behaved posterior distributions for  $k_r$  and  $k_g$ , but that the results for  $D_r$  and  $D_g$  are problematic since the univariate posterior distributions are not unimodal and the 95% credible intervals are relatively wide (figure 3e–n). A posterior predictive check with 30 randomly sampled parameter choices from the latter part of the Markov chain are provided for completeness (figure 3o–r). The posterior predictive check compares model predictions with the experimental data, with additional comparisons using the posterior median (electronic supplementary material). However, these comparisons are less useful since they are based on parameters arising from a Markov chain that has not settled to a well-defined posterior. In summary, without assuming  $D_r = D_g$ , our model parameters are practically non-identifiable according to the Bayesian approach.



**Figure 3.** Typical Markov chain iterations, of length 51 000, for  $D_r$ ,  $D_g$ ,  $k_r$  and  $k_g$  in (a)–(d), respectively. In this case, the Markov chain is initiated with  $\theta_0 = (228, 803, 0.0334, 0.0790)$ . Results in (e)–(n) show a plot matrix representation of the univariate marginals and bivariate marginals estimated using the final 50 000 iterations of the Markov chain in (a)–(d). For the univariate distribution the posterior modes are  $\bar{D}_r = 758 \mu\text{m}^2/\text{h}$ ,  $\bar{D}_g = 1203 \mu\text{m}^2/\text{h}$ ,  $\bar{k}_r = 0.0259/\text{h}$  and  $\bar{k}_g = 0.0703/\text{h}$ . The 95% credible intervals are  $D_r \in [86, 1100]$ ,  $D_g \in [104, 1939]$ ,  $k_r \in [0.0188, 0.0371]$  and  $k_g \in [0.0617, 0.0863]$ . In the univariate marginals, the 95% credible intervals are shown in red vertical dashed lines; in the bivariate marginals the region of maximum density is shown in the darkest shade. Results in (o)–(r) superimpose 30 solutions of equations (3.1) and (3.2) where  $\theta$  is randomly sampled from the Markov chain. MCMC results use  $\sigma = 0.05$ .

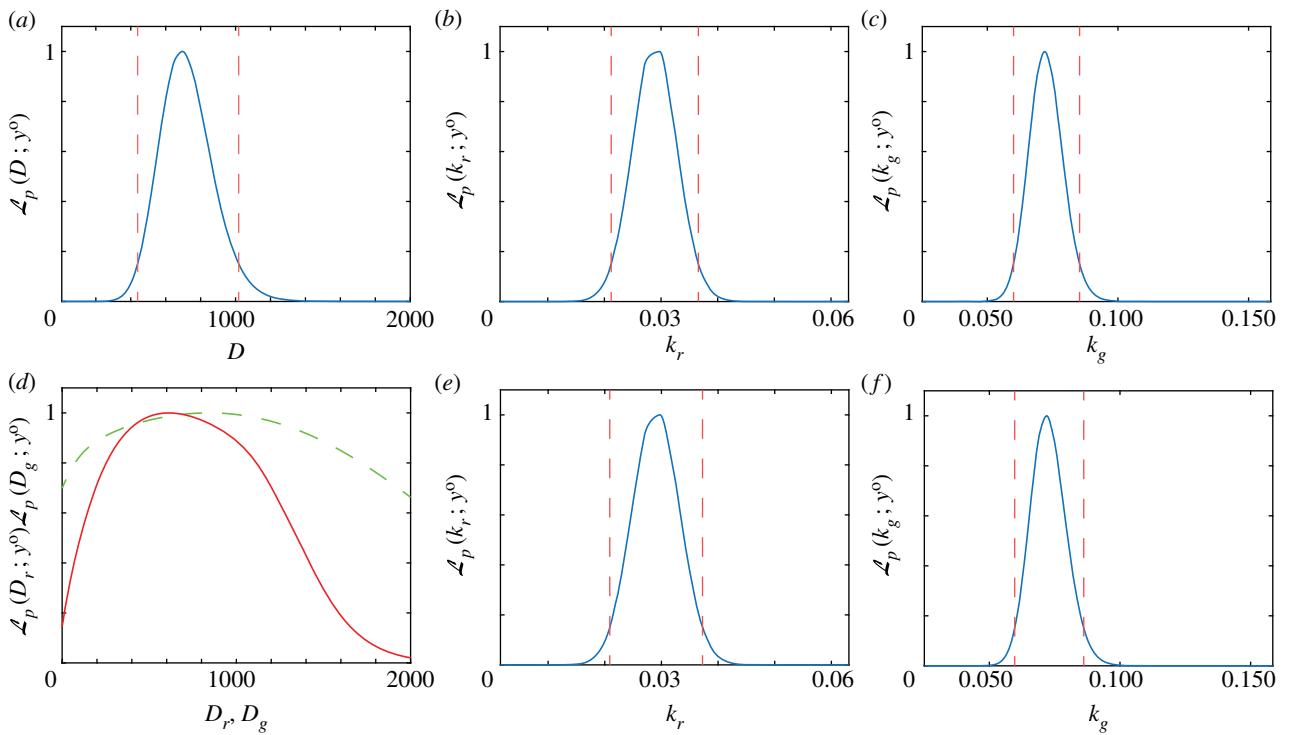
## 5.2. Profile likelihood analysis

Here we consider the results from our profile likelihood analysis. Each profile corresponds to taking a particular parameter as the interest parameter and the remaining parameters as nuisance parameters to be optimized out. More general partitions into an interest parameter and nuisance parameters are also considered in the electronic supplementary material. Here we present results for scenario 1 (figure 4a–c) and results for scenario 2 (figure 4d–f).

In the first scenario where  $D_r = D_g = D$ , we see regular-shaped profiles with clearly defined peaks for  $D$ ,  $k_r$  and  $k_g$  (figure 4a–c). These results are very similar to those obtained from the Bayesian analysis using MCMC, in terms of both interval estimates and maximum-likelihood estimates. The maximum-likelihood estimates are  $\theta^* = (D^*, k_r^*, k_g^*) = (691, 0.0289, 0.0719)$ .

In the second scenario, where  $D_r \neq D_g$ , we see very similar estimates for  $k_r$  and  $k_g$ . The diffusivities,  $D_r$  and  $D_g$ , are much less well determined, however. In particular, the profile-likelihood-based confidence interval for  $D_g$  covers the entire range

of the previously stated conservative bounds, while that for  $D_r$  only excludes a negligible part of this region. Thus, while  $k_r$  and  $k_g$  are well identified,  $D_r$  and  $D_g$  are not. These results are again very similar to those obtained from the Bayesian analysis using MCMC, though the diffusivity estimates are even more conservative here. The maximum-likelihood estimates are  $\theta^* = (D_r^*, D_g^*, k_r^*, k_g^*) = (612, 842, 0.0290, 0.0717)$ . We also generated profiles for the difference  $D_r - D_g$  (electronic supplementary material). The associated interval encompasses a broad range of values, including zero, but also includes very large positive and negative values. This is consistent with a lack of practical identifiability of  $D_r - D_g$ . Despite the fact that the diffusivities are practically non-identifiable with our relatively abundant experimental data, further experimentation with synthetic data indicates that all parameters are, in fact, strictly identifiable provided sufficient data (electronic supplementary material). These additional results (electronic supplementary material) allow us to explore how the choice of data resolution affects practical identifiability.



**Figure 4.** Profile likelihoods. Each profile likelihood is computed using a grid of 50 equally spaced points; cubic interpolation is used for display and to determine confidence intervals. Approximate 95% confidence intervals are indicated based on a relative likelihood threshold of 0.15 [40]. Results for the first scenario where  $D_r = D_g = D$  (*a–c*), and results for the second scenario where  $D_r \neq D_g$  (*d–f*). In (*d*), the profile likelihood for  $D_g$  and  $D_r$  are in red (solid) and green (dashed), respectively. Interval estimates are:  $D \in [439, 1018]$ ,  $k_r \in [0.0212, 0.0366]$ ,  $k_g \in [0.0601, 0.0853]$  for  $D = D_r = D_g$ . Maximum-likelihood estimates are:  $D^* = 691$ ,  $k_r^* = 0.0289$ ,  $k_g^* = 0.0719$  for  $D = D_r = D_g$ . Interval estimates are:  $D_r \in [1.4, 1647]$ ,  $D_g \in [0, 2000]$  (i.e. the whole interval),  $k_r \in [0.0209, 0.0373]$ ,  $k_g \in [0.0597, 0.0861]$  for  $D_r \neq D_g$ . Maximum-likelihood estimates are:  $D_r^* = 1.4$ ,  $D_g^* = 1647$ ,  $k_r^* = 0.0290$ ,  $k_g^* = 0.0717$  for  $D_r \neq D_g$ .

In both cases, the results of the profile likelihood analysis are consistent with the MCMC-based Bayesian analysis. However, the profiles are far cheaper to compute than full MCMC posterior distributions. For example, each profile here was computed in less than 1 min on a standard laptop, whereas the MCMC analysis took approximately 30 min of computation. Each optimization for a given value of the target parameter in the profile analysis took less than 1 s to solve, and profiles can be computed in parallel.

## 6. Conclusion and outlook

In this work, we assess the practical identifiability of a spatio-temporal reaction-diffusion model of a scratch assay where we have access to a relatively large amount of experimental data. In the literature, most scratch assays are reported by providing single images at the beginning and the conclusion of the experiment without making measurements of spatial distributions of cell density at multiple time points (e.g. [48,49]). By contrast, we consider detailed cell density measurements of two subpopulations at four time points with a relatively high spatial resolution of cell density at 24 spatial locations. This means that we work with  $24 \times 4 \times 2 = 192$  measurements of cell density. Given this dataset, a key decision when using a mathematical model is to determine how complicated the model ought to be. We first explored this question through a Bayesian, MCMC-based framework, and then with a profile likelihood-based approach.

MCMC-based Bayesian analysis is relatively common in the systems biology literature for temporal models [34]. Bayesian MCMC approaches for spatio-temporal models are less

common, but increasingly of interest. While it is established that practical parameter identifiability can often be diagnosed using an MCMC framework, many applications of Bayesian MCMC in the mathematical biology literature never explicitly consider the question of practical identifiability [35]. Furthermore, MCMC can be misleading in the presence of true non-identifiability [42], as can other methods such as the bootstrap or Fisher-information-based approaches [17]. Even assuming that MCMC can reliably determine non-identifiability, MCMC can be computationally expensive, especially for models with large  $|\theta|$ . Thus, we also assess identifiability using an optimization-based, profile likelihood approach. In contrast to previous work on temporal data and temporal processes, here we apply this approach to spatio-temporal reaction-diffusion models and spatio-temporal data. Algorithms are available on GitHub (<https://github.com/ProfMJSimpson/Identifiability>).

A key feature of our study is that we consider two slightly different modelling scenarios. In the first scenario, we assume  $D = D_r = D_g$  so that the diffusivity of cells in G1 phase is identical to the diffusivity of cells in S/G2/M phase. In this case, there are three unknown parameters,  $\theta = (D, k_r, k_g)$ . By contrast, for the second scenario we do not invoke this assumption, and there are four unknown parameters,  $\theta = (D_r, D_g, k_r, k_g)$ . Despite the relative abundance of experimental data and the apparent simplicity of the reaction-diffusion model, we find that the parameters are identifiable in the first scenario but are practically non-identifiable in the second.

Our results show that the profile likelihood provides similar results to the Bayesian MCMC approach, with the advantage of being an order of magnitude faster to implement.

Typical MCMC results in figures 2 and 3 require approximately 30 min of computation time, whereas the profile likelihood results in figure 4 require approximately 1 min to compute on a standard laptop (Dell, Intel Core i7 processor, 2.8 GHz, 16 GB RAM). Both methods indicate practical non-identifiability of the diffusivities in the second scenario. As mentioned, however, profile likelihood has previously been shown to be more reliable in the presence of true non-identifiability [17,42]; this is an important consideration as the reaction–diffusion models we used here are already relatively simple and neglect certain mechanisms. Potential mechanisms that might be incorporated in a further extended model include cell–cell adhesion [8,50] or directed motion such as chemotaxis [51,52].

Overall, we find that MCMC provides additional information about practical identifiability but it is more expensive to compute, whereas the profile likelihood analysis provides less information but is computationally cheaper. There are situations where this additional information, such as determining if parameter estimates are correlated, is useful and in this sense the MCMC is preferred. However, there are also other situations where MCMC is challenging or computationally impossible, such as if we are dealing with models that are very expensive to simulate, or models with a much larger  $|\theta|$ , but perhaps a smaller number of ‘interest’ parameters. In this case, MCMC could be computationally expensive, making the profile likelihood preferred. Our aim in this work is to present a balanced view, and not to promote one method over the other. Since the profile likelihood is often simpler and/or faster to compute, our general recommendation is that the profile likelihood makes an excellent screening tool that can provide rapid insight before more complicated MCMC results are generated, if necessary.

Visual inspection of the posterior predictive check in figure 2 shows that, while the overall comparison between the model and the experimental data is reasonable, there are some minor discrepancies, such as in the G1 (red) subpopulation near the centre of the scratch at  $t = 32$  h, suggesting some model inadequacy. This observation is consistent with the fact that we neglect some mechanisms that could potentially improve the quality of the match. Had we taken a more standard approach without considering an identifiability analysis, we might have been tempted to extend equations (3.1) and (3.2) to improve the quality of match between the data and the solution of the model. Such extensions (e.g. adhesion, chemotaxis), while conceptually straightforward to implement, involve increasing the dimension of the parameter space. We caution against naively implementing such extensions since our analysis shows that even the simpler models we consider are not always practically identifiable even with the extensive dataset that we work with here. A profile likelihood analysis of particular, lower dimensional *interest parameters* within a more complex model may still be feasible, however.

There are many ways our analysis could be extended, in terms of the process model, the observation model and the

experimental model. A key assumption here is that the observation model is Gaussian with a fixed variance. While this is a standard assumption [9,14], one possible extension would be to treat the observation model parameter  $\sigma$  as a variable to be estimated at the same time as the process model parameters  $\theta$ . Another extension would be to recognize that, since we are dealing with count data, a multinomial observation model could be more appropriate. Both of these extensions could be dealt with directly in the same MCMC framework but the profile likelihood optimization problems would become slightly more complicated than simple least squares. In terms of the process model, a further extension would be to apply the same reaction–diffusion models in a different geometry. For example, data reported by Jin *et al.* [53] use a different experimental approach to produce square- and triangular-shaped scratches. While the same reaction–diffusion models as those used here could be used to model Jin *et al.*’s [53] experiments, such models would need to be solved in two spatial dimensions because of the difference in initial conditions [54]. This extension could be handled by the same MCMC and the same profile likelihood approaches without increasing the dimension of the parameter space. However, this extension would increase the computational effort required to solve the reaction–diffusion model and this would amplify the computational advantage of the profile likelihood method over the Bayesian MCMC approach. In terms of the experimental model, in this work we use cell cycle labelling with two colours, red and green. More advanced cell cycle labelling technologies, such as FUCCI4 [55,56], are now available, and these technologies label the cell cycle with four fluorescent colours [55,56]. Data from a scratch assay where cells are labelled with FUCCI4 could be interpreted with an extended system of reaction–diffusion equations, similar to equations (3.1) and (3.2). Such an extended system of equations would be similar to equations (3.1) and (3.2), but the dimensionality of the parameter space would increase. With four colours we would have the possibility of working with four distinct cell diffusivities and four distinct rate parameters, thereby doubling the size of the parameter space, hence we leave this for future consideration.

**Data accessibility.** Experimental data are available in the electronic supplementary material.

**Authors’ contributions.** All authors designed the investigation. M.J.S. and O.J.M. conducted the analysis, S.T.V. performed the experiments, M.J.S. and O.J.M. wrote the manuscript and all authors approved the final version of the manuscript.

**Competing interests.** We declare we have no competing interest.

**Funding.** M.J.S. is supported by the Australian Research Council (DP170100474) and the University of Canterbury Erskine Fellowship. O.J.M. is supported by the University of Auckland, Faculty of Engineering James and Hazel D. Lord Emerging Faculty Fellowship. R.E.B. is supported by a Royal Society Wolfson Research Merit Award, a Leverhulme Trust Research Fellowship and BBSRC(BB/R000816/1). We thank three referees for helpful feedback.

## References

1. Simpson MJ, Treloar KK, Binder BJ, Haridas P, Manton KJ, Leavesley DI, McElwain DLS, Baker RE. 2013 Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *J. R. Soc. Interface* **10**, 20130007. (doi:10.1098/rsif.2013.0007)
2. Maini PK, McElwain DLS, Leavesley DI. 2004 Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.* **10**, 475–482. (doi:10.1089/107632704323061834)
3. Grada A, Otero-Vinas M, Prieto-Castrillo F, Obagi Z, Falanga V. 2016 Research techniques made simple: analysis of collective cell migration using the wound healing assay. *J. Invest. Dermatol.* **137**, e11–e16. (doi:10.1016/j.jid.2016.11.020)

4. Haass NK, Beaumont KA, Hill DS, Anfosso A, Mras P, Munoz MA, Kinjyo I, Weninger W. 2014 Realtime cell cycle imaging during melanoma growth, invasion, and drug response. *Pigment Cell Melanoma Invasion* **27**, 764–776. (doi:10.1111/pcmr.12274)
5. Murray JD. 2002 *Mathematical biology I: an introduction*. Heidelberg, Germany: Springer.
6. Sherratt JA, Murray JD. 1990 Models of epidermal wound healing. *Proc. R. Soc. Lond. B* **241**, 29–36. (doi:10.1098/rspb.1990.0061)
7. Sengers BG, Please CP, Oreffo ROC. 2007 Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration. *J. R. Soc. Interface* **4**, 1107–1117. (doi:10.1098/rsif.2007.0233)
8. Nardini JT, Chapnick DA, Liu X, Bortz DM. 2016 Modeling keratinocyte wound healing dynamics: cell–cell adhesion promotes sustained collective migration. *J. Theor. Biol.* **400**, 103–117. (doi:10.1016/j.jtbi.2016.04.015)
9. Warne DJ, Baker RE, Simpson MJ. 2019 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)
10. Johnston ST, Ross RV, Binder BJ, McElwain DLS, Haridas P, Simpson MJ. 2016 Quantifying the effect of experimental design choices for in vitro scratch assays. *J. Theor. Biol.* **400**, 19–31. (doi:10.1016/j.jtbi.2016.04.012)
11. Audoly S, Bellu G, D’Angiò L, Saccomani MP, Cobelli C. 2001 Global identifiability of nonlinear models of biological systems. *IEEE Trans. Biomed. Eng.* **48**, 55–65. (doi:10.1109/10.900248)
12. 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)
13. Cobelli C, Distefano JJ. 1980 Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **8**, R7–R24. (doi:10.1152/ajpregu.1980.239.1.R7)
14. Hines KE, Middendorf TR, Aldrich RW. 2014 Determination of parameter identifiability in nonlinear biophysical models: a Bayesian approach. *J. Gen. Physiol.* **143**, 401. (doi:10.1085/jgp.201311116)
15. 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)
16. 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)
17. Fröhlich F, Theis FJ, Hasenauer J. 2014 Uncertainty analysis for non-identifiable dynamical systems: profile likelihoods, bootstrapping and more. In *Computational Methods in Systems Biology: Proc. 12th Int. Conf., CMSB 2014, Manchester, UK, 17–19 November 2014*, pp. 61–72. Berlin, Germany: Springer.
18. Villaverde AF, Tsiantis N, Banga JR. 2019 Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models. *J. R. Soc. Interface* **16**, 20190043. (doi:10.1098/rsif.2019.0043)
19. Janzen DL, Bergenholm L, Jirstrand M, Parkinson J, Yates J, Evans ND, Chappell MJ. 2016 Parameter identifiability of fundamental pharmacodynamic models. *Front. Physiol.* **7**, 590. (doi:10.3389/fphys.2016.00590)
20. Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. 2014 Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* **v30**, 1440–1448. (doi:10.1093/bioinformatics/btu006)
21. 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)
22. Greenland S. 2011 Identifiability. In *International encyclopedia of statistical science* (ed. M Lovric), p. 645. Berlin, Germany: Springer-Verlag.
23. Bickel PJ, Doksum KA. 2015 *Mathematical statistics: basic ideas and selected topics*, vol. 1. Boca Raton, FL: CRC Press.
24. Evans SN, Stark PB. 2002 Inverse problems as statistics. *Inverse Prob.* **18**, R55. (doi:10.1088/0266-5611/18/4/201)
25. Maclare OJ, Nicholson R. 2019 What can be estimated? Identifiability, estimability, causal inference and ill-posed inverse problems. (<http://arxiv.org/abs/1904.02826>).
26. Bellu G, Saccomani MP, Audoly S, D’Angio L. 2007 DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* **88**, 52–61. (doi:10.1016/j.cmpb.2007.07.002)
27. Eisenberg MC, Robertson SL, Tien JH. 2013 Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. *J. Theor. Biol.* **234**, 84–102. (doi:10.1016/j.jtbi.2012.12.021)
28. Akaike H. 1974 A new look at the statistical model identification. *IEEE Trans. Automat. Control* **19**, 716–723. (doi:10.1109/TAC.1974.1100705)
29. Shmueli G. 2010 To explain or to predict? *Stat. Sci.* **25**, 289–310. (doi:10.1214/10-STS330)
30. Cox DR. 2006 *Principles of statistical inference*. Cambridge, UK: Cambridge University Press.
31. Bland JM. 2009 The tyranny of power: is there a better way to calculate sample size?. *Br. Med. J.* **339**, b3985. (doi:10.1136/bmj.b3985)
32. Rothman KJ, Greenland S. 2018 Planning study size based on precision rather than power. *Epidemiology* **29**, 599–603. (doi:10.1097/EDE.0000000000000876)
33. Warne DJ, Baker RE, Simpson MJ. 2019 Simulation and inference algorithms for biochemical stochastic reaction networks: from basic concepts to state-of-the-art. *J. R. Soc. Interface* **16**, 20180943. (doi:10.1098/rsif.2018.0943)
34. 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**, 187–202. (doi:10.1098/rsif.2008.0172)
35. Vo BN, Drovandi CC, Pettitt AN, Simpson MJ. 2015 Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. *Math. Biosci.* **263**, 133–142. (doi:10.1016/j.mbs.2015.02.010)
36. Browning AP, Haridas P, Simpson MJ. 2019 A Bayesian sequential learning framework to parameterise continuum models of melanoma invasion into human skin. *Bull. Math. Biol.* **81**, 676–698. (doi:10.1007/s11538-018-0532-1)
37. Maclare OJ, Parker A, Pin C, Carding SR, Watson A, Fletcher AG, Byrne HM, Maini PK. 2017 A hierarchical Bayesian model for understanding the spatiotemporal dynamics of the intestinal epithelium. *PLoS Comput. Biol.* **13**, e1005688. (doi:10.1371/journal.pcbi.1005688)
38. Lambert B, MacLean AL, Fletcher AG, Combes AN, Little MH, Byrne HM. 2018 Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. *J. Math. Biol.* **76**, 1673–1697. (doi:10.1007/s00285-018-1208-z)
39. Wilkinson RD. 2013 Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Stat. Appl. Genet. Mol. Biol.* **12**, 129–141. (doi:10.1515/sagmb-2013-0010)
40. Pawitan Y. 2001 *In all likelihood: statistical modelling and inference using likelihood*. Oxford, UK: Oxford University Press.
41. Vittadello ST, McCue SW, Gunasingh G, Haass NK, Simpson MJ. 2018 Mathematical models for cell migration with real-time cell cycle dynamics. *Biophys. J.* **114**, 1241–1253. (doi:10.1016/j.bpj.2017.12.041)
42. 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. *Phil. Trans. R. Soc. A* **371**, 20110544. (doi:10.1098/rsta.2011.0544)
43. Treloar KK, Simpson MJ, Haridas P, Manton KJ, Leavesley DL, McElwain DLS, Baker RE. 2013 Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies. *BMC Syst. Biol.* **7**, 137. (doi:10.1186/1752-0509-7-137)
44. Mathworks. 2020 lsqnonlin. Solve nonlinear least-squares. <https://au.mathworks.com/help/optim/ug/lsqnonlin.html> (accessed February 2020).
45. Edwards AW. 1992 *Likelihood*, expanded edn. Baltimore, MD: Johns Hopkins University Press.
46. Royall R. 2017 *Statistical evidence: a likelihood paradigm*. London, UK: Routledge.
47. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2014 *Bayesian data analysis*, 3rd edn. London, UK: Chapman & Hall.
48. Fattah F et al. 2016 Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. *Nature* **531**, 105–109. (doi:10.1038/nature16951)
49. Wang X et al. 2019 UDP-glucose accelerates *SNAI1* mRNA decay and impairs lung cancer metastasis. *Nature* **571**, 127–131. (doi:10.1038/s41586-019-1340-y)

50. Armstrong NJ, Painter KJ, Sherratt JA. 2006 A continuum approach to modelling cell-cell adhesion. *J. Theor. Biol.* **243**, 98–113. (doi:10.1016/j.jtbi.2006.05.030)
51. Keller EF, Segel LA. 1971 Model for chemotaxis. *J. Theor. Biol.* **30**, 225–234. (doi:10.1016/0022-5193(71)90050-6)
52. Simpson MJ, Landman KA, Newgreen DF. 2006 Chemotactic and diffusive migration on a nonuniformly growing domain: numerical algorithm development and applications. *J. Comput. Appl. Math.* **192**, 282–300. (doi:10.1016/j.cam.2005.05.003)
53. Jin W, Lo K-Y, Chou S-E, McCue SW, Simpson MJ. 2018 The role of initial geometry in experimental models of wound closing. *Chem. Eng. Sci.* **179**, 221–226. (doi:10.1016/j.ces.2018.01.004)
54. Simpson MJ, Landman KA, Hughes BD. 2009 Multi-species simple exclusion processes. *Physica A* **388**, 399–406. (doi:10.1016/j.physa.2008.10.038)
55. Bajar BT *et al.* 2016 Fluorescent indicators for simultaneous reporting of all four cell cycle phases. *Nat. Methods* **13**, 993–996. (doi:10.1038/nmeth.4045)
56. Liccardi G, Fava L. 2017 Fiat lux: illuminating the cell cycle. *Cell Death Discovery* **3**, 17042. (doi:10.1038/cddiscovery.2017.42)

## Supplementary Material

Matthew J Simpson<sup>1</sup>, Ruth E Baker<sup>1</sup>, Sean T Vittadello<sup>1</sup>, Oliver J Maclaren<sup>1</sup>

<sup>a</sup>*School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia.*

<sup>b</sup>*Mathematical Institute, University of Oxford, Oxford, UK.*

<sup>c</sup>*Department of Engineering Science, University of Auckland, Auckland 1142, New Zealand.*

---

---

---

\*Corresponding author: matthew.simpson@qut.edu.au

## 1. Numerical methods

Solutions of Equations (3)–(4) are obtained numerically. To obtain these solutions we uniformly discretise the spatial domain,  $0 \leq x \leq L$ , using a mesh of constant spacing,  $\delta x$ . Central differences are used to approximate the spatial derivatives in Equation (3)–(4), and no flux conditions at  $x = 0$  and  $x = L$  are approximated with a standard forward and backward difference approximation, respectively. This spatial discretisation leads to a system of  $M = L/\delta x + 1$  coupled nonlinear ordinary differential equations for  $r(x_i, t)$ , for  $i = 1, 2, \dots, M$  and a system of  $M$  coupled nonlinear ordinary differential equations for  $g(x_i, t)$ , for  $i = 1, 2, \dots, M$ . Temporal derivatives are approximated using a backward Euler approximation with constant time step of duration  $\delta t$ . The resulting systems of nonlinear algebraic equations are linearised using Picard iteration and are solved sequentially, first for  $r(x_i, t_j)$  and then for  $g(x_i, t_j)$ , using the Thomas algorithm for tridiagonal linear systems. Picard iteration is continued until the updated values for  $r(x_i, t_j)$  and then  $g(x_i, t_j)$  converge to within an absolute tolerance  $\varepsilon$ . All results correspond to  $L = 1242 \mu\text{m}$ ,  $\delta x = 1.0 \mu\text{m}$ ,  $\delta t = 1.0 \text{ h}$  and  $\varepsilon = 1 \times 10^{-10}$ .

All numerical solutions require the specification of an initial condition,  $r(x, 0)$  and  $g(x, 0)$ . For the results in the main document we specify these initial conditions by taking  $y_i^0$  at  $t = 0$  and we use linear interpolation to determine  $r(x_i, 0)$  and  $g(x_i, 0)$  at each point on the finite difference mesh,  $i = 1, 2, \dots, M$ .

## 2. Likelihood profiling

Here we give more details on how likelihood profiling is computationally implemented. We first give an overview of the basic concepts and show how, in our case, this methodology gives rise to a series of nonlinear least squares problems that can be solved using standard software. We then present the details of an example profiling calculation in the context of a linear regression model.

### 2.1. Reduction to nonlinear least squares for normally-distributed observation errors

Here we show how the profiling problem given in the main text can be reduced to solving a series of nonlinear least squares problems. We denote by  $y(\psi, \lambda)$  the vector containing all model solutions  $y(s_i, \psi, \lambda)$  at each observation point  $s_i, i = 1, \dots, n$ , for the parameter partition  $\theta = (\psi, \lambda)$ , i.e.

$$y(\psi, \lambda) = (r(s_1, \psi, \lambda), g(s_1, \psi, \lambda), \dots, r(s_n, \psi, \lambda), g(s_n, \psi, \lambda)). \quad (\text{S1})$$

We assume  $y^o$  is arranged similarly. Given that the noise is iid with a normal distribution and standard deviation  $\sigma$  (and is the same for observations of both red and green cells) we obtain the probability density

$$p(y^o; \psi, \lambda) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|y^o - y(\psi, \lambda)\|^2\right). \quad (\text{S2})$$

Hence, for fixed standard deviation  $\sigma$ , the profile (normalised) likelihood for  $\psi$  is

$$\mathcal{L}_p(\psi; y^o) = \frac{\sup_{\lambda} \left( \exp\left(-\frac{1}{2\sigma^2} \|y^o - y(\psi, \lambda)\|^2\right) \right)}{\sup_{(\psi, \lambda)} \left( \exp\left(-\frac{1}{2\sigma^2} \|y^o - y(\psi, \lambda)\|^2\right) \right)}, \quad (\text{S3})$$

where the common factor  $(2\pi\sigma^2)^{-\frac{n}{2}}$  cancels. As maximisation and taking the exponential commute, we have

$$\mathcal{L}_p(\psi; y^o) = \frac{\exp\left(\sup_{\lambda} \left( -\frac{1}{2\sigma^2} \|y^o - y(\psi, \lambda)\|^2 \right) \right)}{\exp\left(\sup_{(\psi, \lambda)} \left( -\frac{1}{2\sigma^2} \|y^o - y(\psi, \lambda)\|^2 \right) \right)}, \quad (\text{S4})$$

and hence the key computations involve solving

$$\sup_{\lambda} -\frac{1}{2\sigma^2} \|y^o - y(\psi, \lambda)\|^2 = \inf_{\lambda} \frac{1}{2\sigma^2} \|y^o - y(\psi, \lambda)\|^2 \quad (\text{S5})$$

for each value of  $\psi$ . Equation (??) is a nonlinear least squares problem that can be solved using standard numerical software. In this work we choose to solve such problems using the lsqnonlin routine in MATLAB [? ]. After solving

each least squares problem we take exponentials to regain the appropriate likelihoods. The overall optimisation for the denominator can be computed by a subsequent optimisation over the fixed  $\psi$  sub-problems, as  $\sup_{(\psi,\lambda)} f(\psi,\lambda) = \sup_\psi (\sup_\lambda (f(\psi,\lambda)))$  for any function  $f$ .

## 2.2. Profiling example: linear regression

Here we consider profiling for the simple linear (in parameters) model

$$y_i = f(x_i, a, b) + \epsilon_i = ax_i + bx_i^2 + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1), \quad (\text{S6})$$

i.e.  $y_i \sim \mathcal{N}(f(x_i, a, b), 1)$ , for  $i = 1, \dots, N$ , meaning that we have Gaussian observation model with  $\sigma = 1$ . We wish to compute the profiles  $\mathcal{L}_p(a; y^o)$  and  $\mathcal{L}_p(b; y^o)$ , where  $y^o$  denotes the combined vector of observations.

Following the arguments outlined above, the key computations for  $\mathcal{L}_p(a; y^o)$ , for example, involve solving least squares problems of the form

$$\inf_b \frac{1}{2\sigma^2} \| (y^o - f(a, b)) \|^2 \quad (\text{S7})$$

for a grid of  $a$  values, and where model outputs at all  $x$  observation locations have been collected (as before) into a vector, here denoted by  $f(a, b)$ . The relative likelihood function is then computed by taking the maximum across all  $a$  values to obtain the overall minimiser for the denominator, and then taking exponentials to regain the likelihoods. We assume that the true likelihood is sufficiently smooth so that we can extend the likelihood as calculated at gridded values to a likelihood defined at all values via interpolation between grid points. Similarly, we compute  $\mathcal{L}_p(b; y^o)$  by switching the roles of  $a$  and  $b$  as interest and nuisance parameters.

A simple example of profiling the two-dimensional likelihood function corresponding to the model in Equation (??) is shown Figure ???. Here we generated 40 samples from the model over a regularly spaced grid for  $x \in [0, 1]$ . The true parameters used were  $a = 1, b = 3$ . The two-dimensional likelihood was generated by direct calculation over a  $200 \times 200$  grid.

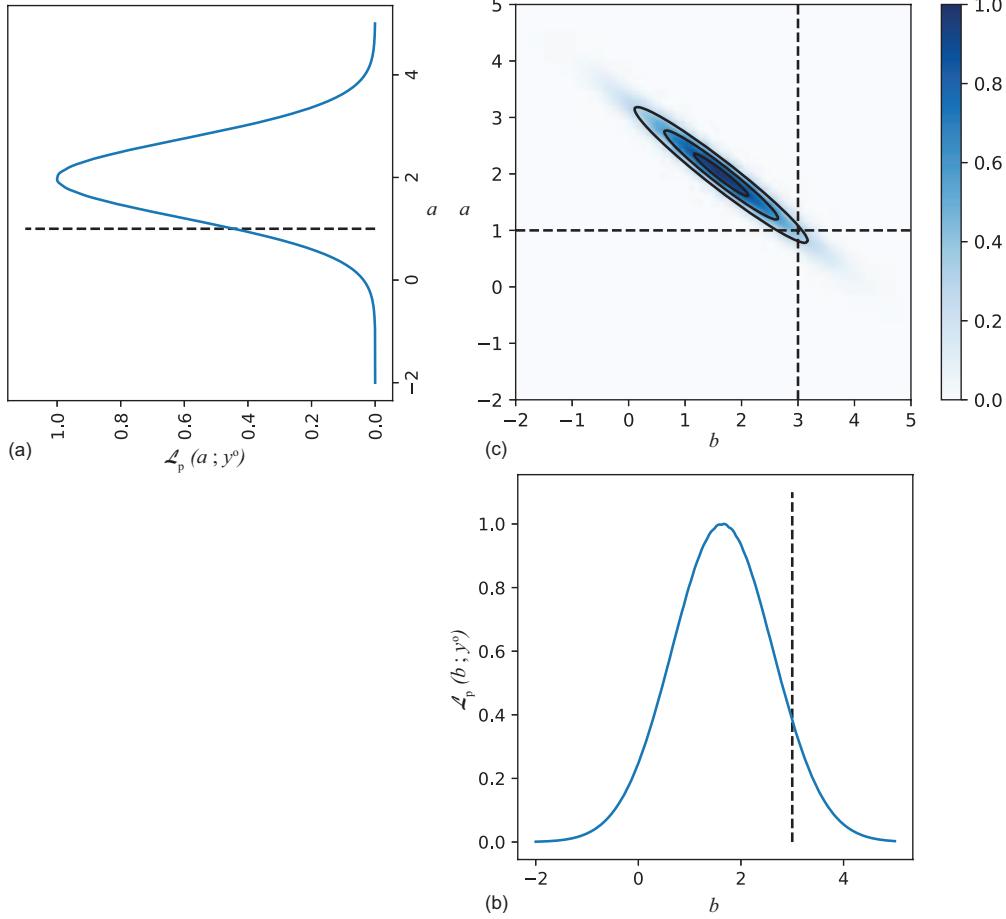


Figure S1: Example of profiling the two-dimensional likelihood function corresponding to the model ?? to obtain profile likelihoods for each parameter  $a$  (a) and  $b$  (b). Here we generated 40 samples from the model over a regularly spaced grid for  $x \in [0, 1]$ . The true parameters used were  $a = 1, b = 3$  and these are indicated by dashed lines. The observation model is Gaussian with  $\sigma = 1$ . Each profile corresponds to the maximum taken over the other dimension and hence to a ‘profile’ view of the two-dimensional likelihood (c). The two-dimensional likelihood is evaluated over a  $200 \times 200$  grid.

### **3. Additional MCMC results for the experimental data**

Results in Figures 2-3 report the Bayesian MCMC results for the two scenarios where  $D = D_r = D_g$  and  $D_r \neq D_g$ , respectively. In those results we present a visual posterior predictive check. Additional results in Figure ??–?? show identical figures except that instead of the posterior predictive check we compare the experimental data with the solution of the model using the univariate mode for the case where  $D = D_r = D_g$  (Figure ??) and the median for the case where  $D_r \neq D_g$  (Figure ??)

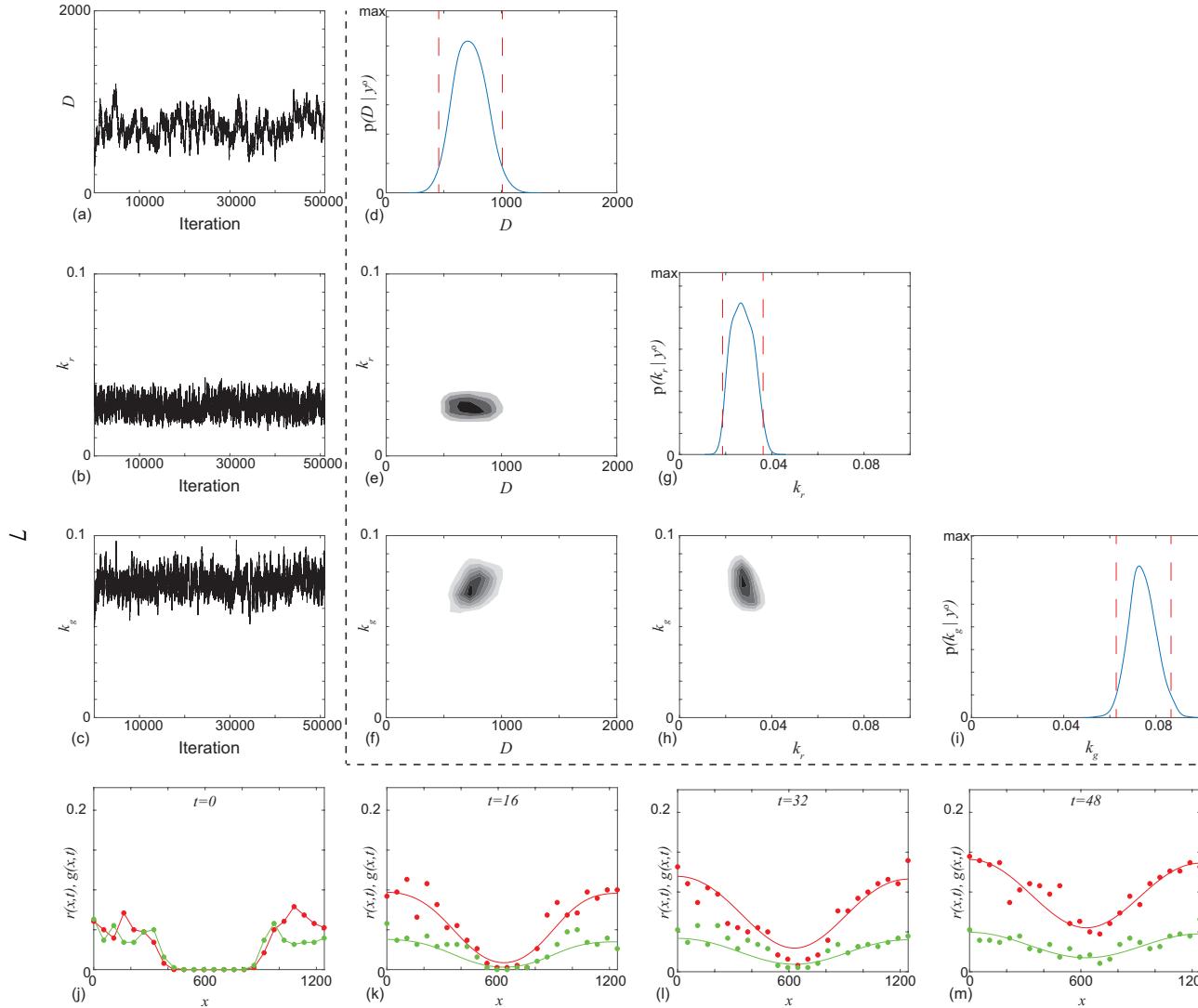


Figure S2: Typical Markov chain iterations, of length 51,000, for  $D$ ,  $k_r$  and  $k_g$  in (a)–(c). In this case the Markov chain is initiated with  $\theta_0 = (500, 0.05, 0.05)$ . Results in (d)–(i) show a plot matrix representation of the univariate marginals and bivariate marginals estimated using the final 50,000 iterations of the Markov chain in (a)–(c). For the univariate distribution the posterior modes are  $\bar{D} = 707 \mu\text{m}^2/\text{h}$ ,  $\bar{k}_r = 0.0266/\text{h}$ , and  $\bar{k}_g = 0.0727/\text{h}$ , and the 95% credible intervals are  $D \in [455, 1007]$ ,  $k_r \in [0.0187, 0.0362]$  and  $k_g \in [0.0628, 0.0727]$ . In the univariate marginals the 95% credible intervals are shown in red dashed vertical lines, in the bivariate marginals the region of maximum density is shown in the darkest shade. Results in (j)–(m) superimpose the solution of Equations (1)–(2) with the median  $\theta = (707, 0.0266, 0.0727)$  superimposed on the experimental data. MCMC results use  $\sigma = 0.05$ .

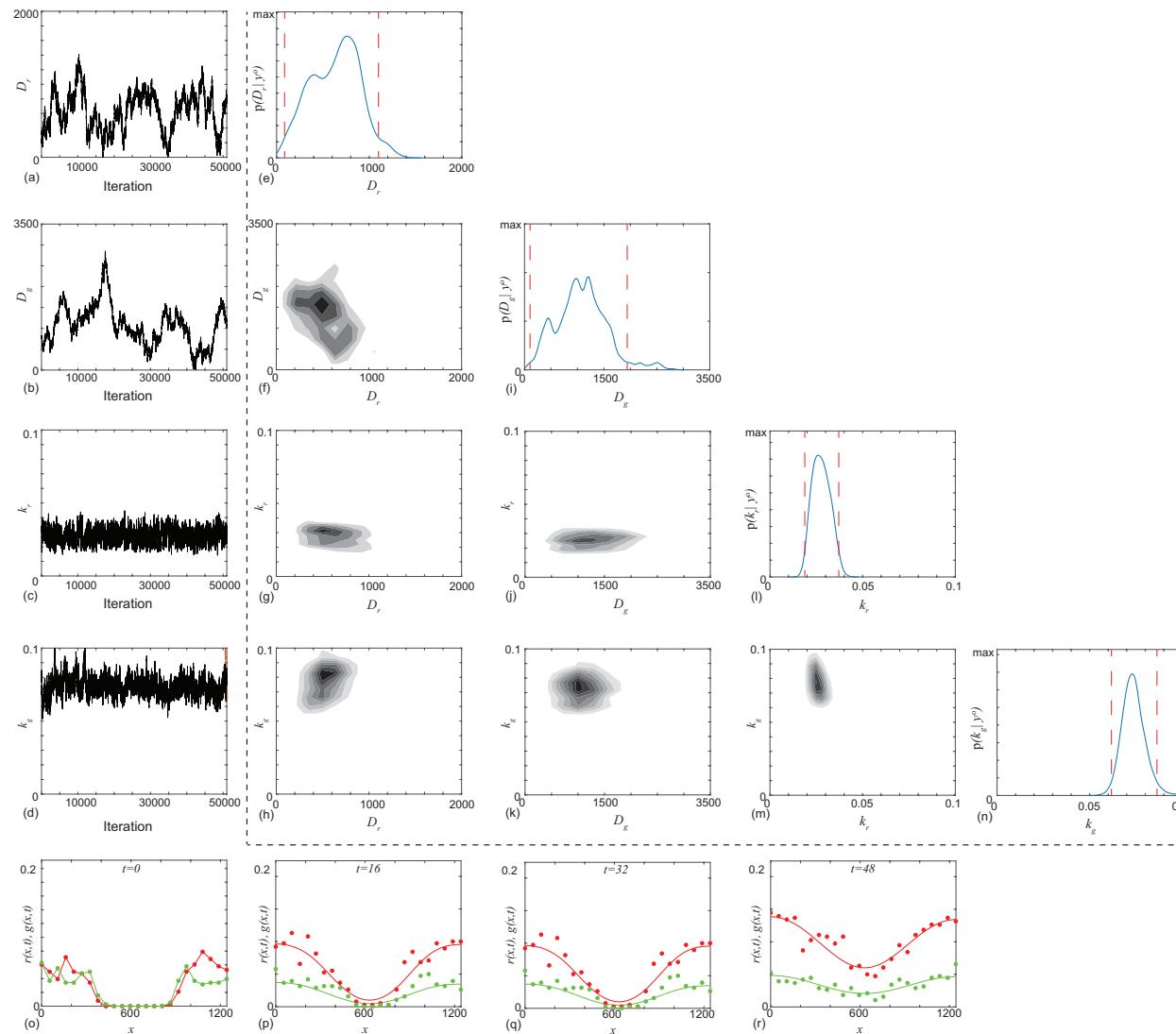


Figure S3: Typical Markov chain iterations, of length 51,000, for  $D_r$ ,  $D_g$ ,  $k_r$  and  $k_g$  in (a)–(d), respectively. In this case the Markov chain is initiated with  $\theta_0 = (228, 803, 0.0334, 0.0790)$ . Results in (e)–(n) show a plot matrix representation of the univariate marginals and bivariate marginals estimated using the final 50,000 iterations of the Markov chain in (a)–(d). For the univariate distribution the posterior modes are  $\bar{D}_r = 758 \mu\text{m}^2/\text{h}$ ,  $\bar{D}_g = 1203 \mu\text{m}^2/\text{h}$ ,  $\bar{k}_r = 0.0259/\text{h}$ , and  $\bar{k}_g = 0.0703/\text{h}$ . The 95% credible intervals are  $D_r \in [86, 1100]$ ,  $D_g \in [104, 1939]$ ,  $k_r \in [0.0188, 0.0371]$  and  $k_g \in [0.0617, 0.0863]$ . In the univariate marginals the 95% credible intervals are shown in red dashed vertical lines, in the bivariate marginals the region of maximum density is shown in the darkest shade. Results in (o)–(r) superimpose the solution of Equations (1)–(2) with the median  $\theta = (660, 1063, 0.0274, 0.0733)$  superimposed on the experimental data. MCMC results use  $\sigma = 0.05$ .

As noted in the main document, the Markov chain in Figure 2a indicates that the first scenario with  $D = D_r = D_g$  is identifiable because the trace plot appears to converge whereas results in Figure 3a-b indicate practical non-identifiability in the second scenario because the trace plots do not converge for  $D_r$  and  $D_g$ . Since the MCMC results in Figure 2 and Figure 3 are for one particular choice of  $\theta_0$ , we provide additional results here to explore the impact of varying  $\theta_0$ . Additional results in Figures ??–?? provides further evidence to support these conclusions by showing additional MCMC results for different choices of  $\theta_0$  for the first case where  $D = D_r = D_g$  (Figure ??) and in the second case where  $D_r \neq D_g$  (Figure ??). Results in Figure ?? indicate rapid convergence for  $D$ ,  $k_r$  and  $k_g$  regardless of  $\theta_0$ . In contrast, results in Figure ?? indicate rapid convergence for  $k_r$  and  $k_g$  regardless of  $\theta_0$ , but we see that the Markov chains for  $D_r$  and  $D_g$  depend upon  $\theta_0$  and do not converge.

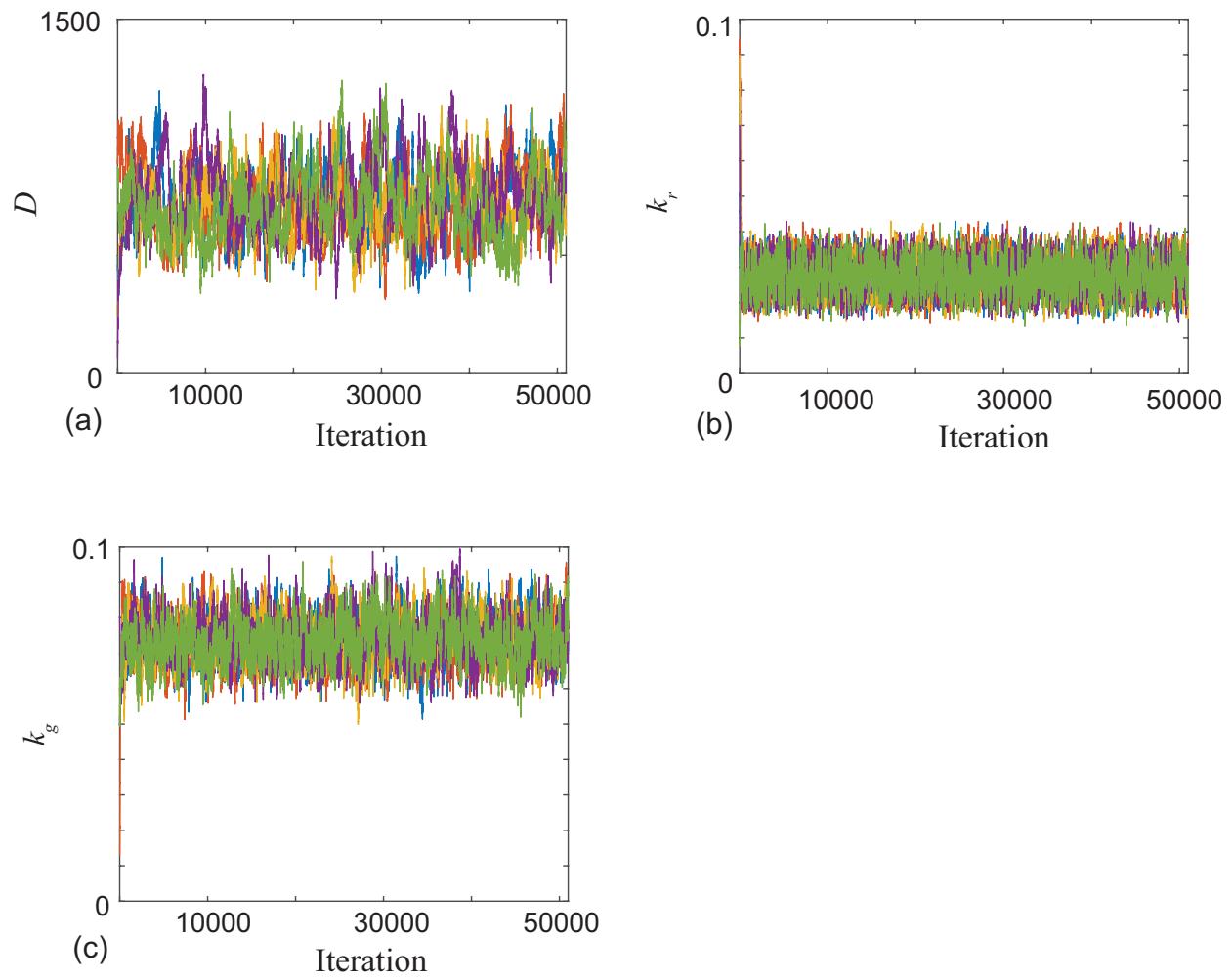


Figure S4: Five realizations of the Markov Chain from Figure 2. The initial choice of  $\theta_0$  are (500, 500, 0.0500, 0.0500) (blue), (906, 0.0900, 0.0130) (orange), (259, 0.0900, 0.0600) (yellow), (67, 0.0700, 0.0800) (purple) and (866, 0.0096, 0.0550) (green). Results in Figure 2 correspond to the blue chain.

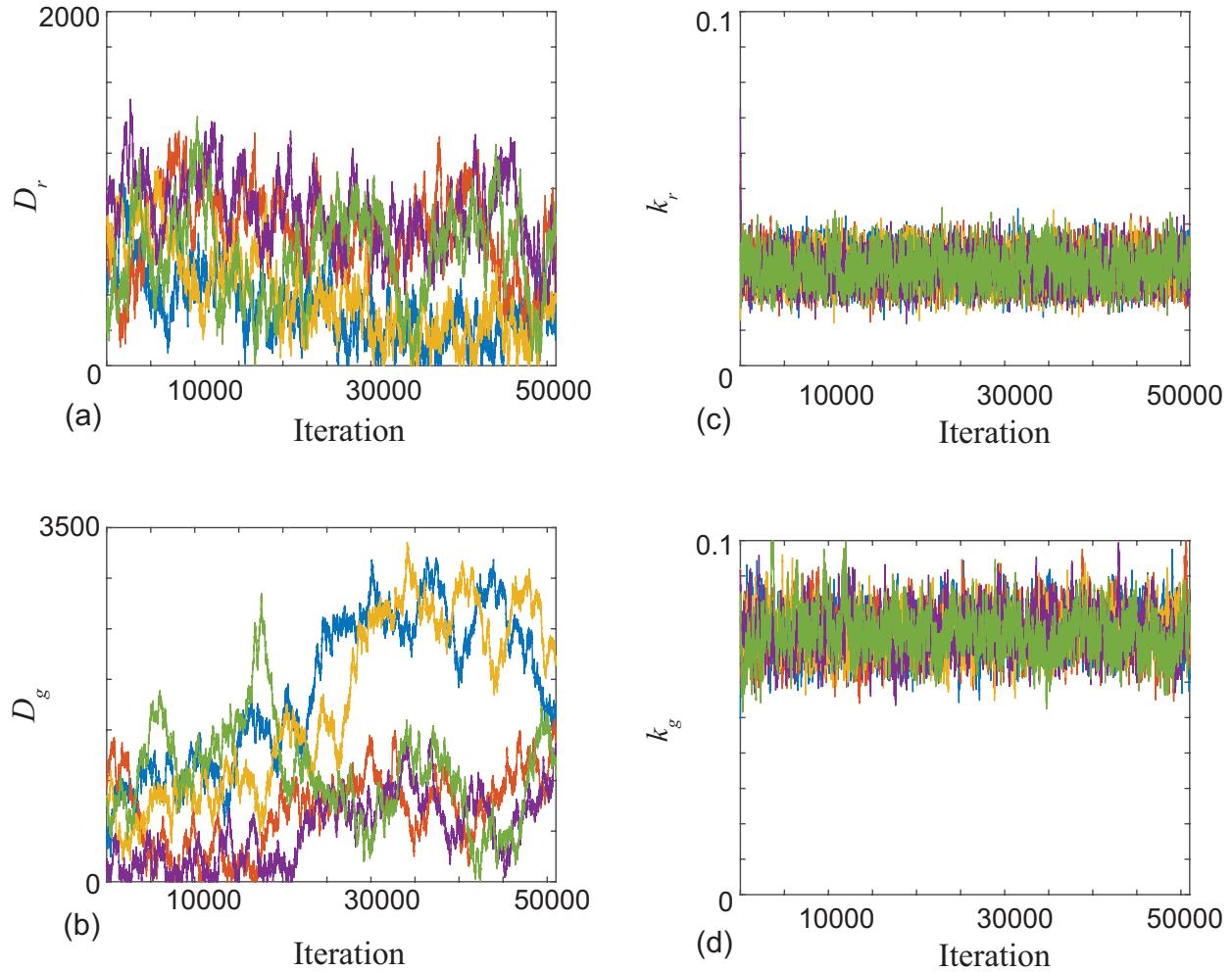


Figure S5: Five realizations of the Markov Chain from Figure 3. The initial choice of  $\theta_0$  are (500, 500, 0.0500, 0.0500) (blue), (705, 979, 0.0573, 0.0875) (orange), (905, 814, 0.0127, 0.0913) (yellow), (905, 215, 0.0727, 0.0914) (purple) and (228, 803, 0.0334, 0.0790) (green). Results in Figure 3 correspond to the green chain.

#### 4. Additional profile likelihood results for the experimental data

Results in Figure ?? show the profile for the experimental data under the second scenario where  $D_r \neq D_g$ . Here the interest parameter is the difference in diffusivities, i.e.  $\psi = D_r - D_g$ . Results indicate that compatible results include zero, but also differences at least as large as  $+/- 1500$ , and the cut-off 0.15 lies outside of this interval.

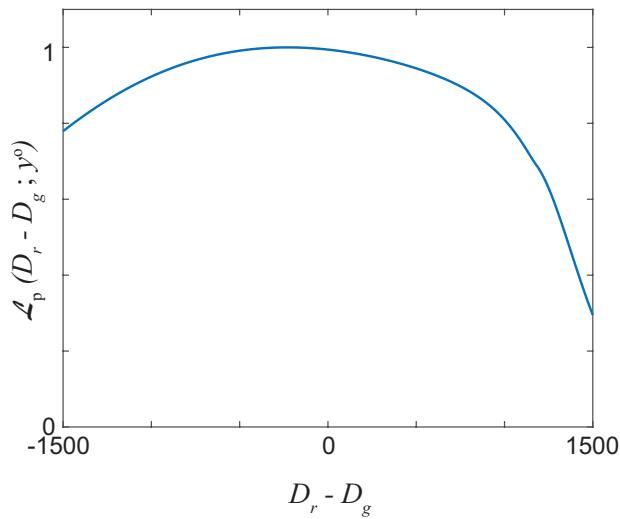


Figure S6: Profile likelihood for  $D_r - D_g$  for the second scenario where  $D_r \neq D_g$ . The profile likelihood is computed using a grid of 50 equally-spaced points; cubic interpolation is used for display and to determine confidence intervals.

## 5. Additional MCMC results for finely resolved synthetic data

To aid in demonstrating the key difference between practical and structural identifiability we now repeat all results from the main document using synthetic data for which we specify known parameter values. This approach allows us to collect as much data as we prefer since we are not limited by experimental constraints. Since we can collect as much data as we like we will refer to these results as *finely resolved synthetic data*. This synthetic data corresponds to solutions of Equations (1)–(2) on the same domain. The initial condition is  $r(x, 0) = g(x, 0) = 0.2$  for  $0 < x < 400$ ,  $r(x, 0) = g(x, 0) = 0.2$  for  $842 < x < 1242$ , and  $r(x, 0) = g(x, 0) = 0$  for  $400 < x < 842$ . We consider two scenarios, the first has  $D_r = D_g$  with  $\theta = (500, 500, 0.05, 0.10)$ , and the second has  $D_r \neq D_g$  with  $\theta = (700, 300, 0.05, 0.10)$ . In both scenarios we collect synthetic data at each point on the finite difference mesh, and at  $t = 0, 16, 32$  and  $48$ . Thus  $y^0$  consists of  $N = 1243 \times 4 = 4972$  observations of both of  $r$  and  $g$ , that is 9944 measurements of density in total. Therefore the synthetic data cases involves dealing with almost two orders of magnitude more observations than was possible with the experimental data. In the next section we also consider using a finer temporal resolution, along with the finer spatial resolution, using profile likelihood.

Results in Figures ??-?? show the Bayesian MCMC results for the two scenarios respectively. In both cases we see that the Markov chains appear to converge and the univariate posterior distributions are regular shaped with clearly defined maxima that are close to the expected values. These results indicate that the parameters are identifiable in both synthetic scenarios provided we have access to sufficient observations, i.e. we expect that with even more data our estimates will converge to the true values. However, even in this dense data limit, when we compare results in Figure ?? for  $D_r = D_g$  with results in Figure ?? for  $D_r \neq D_g$ , we see that the credible intervals in second scenario are relatively wider than in the first scenario. This indicates that identifiability in the second scenario is more challenging than in the first, and this is consistent with the results in Figure 2-3 for the experimental data. Nonetheless, with this finely resolved synthetic data we see that the MCMC results leads to well-formed posteriors and each marginal posterior is approximately symmetric and unimodal with the point estimates close to the known values in each case.

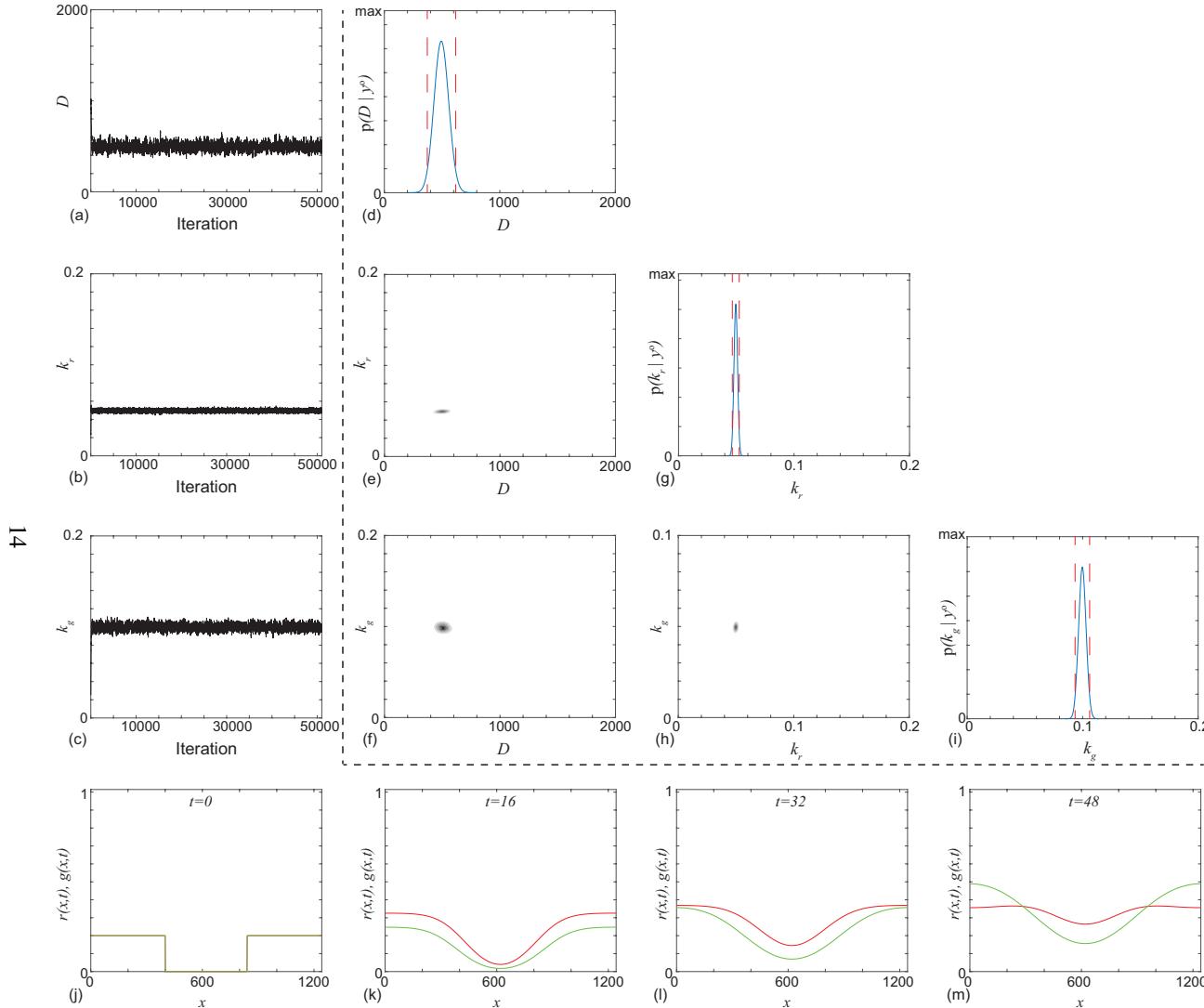


Figure S7: Typical Markov chain iterations, of length 51,000, for  $D$ ,  $k_r$  and  $k_g$  in (a)–(c), respectively. In this case the Markov chain is initiated with  $\theta = (1000, 0.025, 0.025)$ . In this case the true value is  $\theta = (500, 0.050, 0.10)$ . Results in (d)–(i) show a plot matrix representation of the univariate marginals and bivariate marginals estimated using the final 50,000 iterations of the Markov chain in (a)–(c). For the univariate distribution the posterior modes are  $\bar{D} = 493 \mu\text{m}^2/\text{h}$ ,  $\bar{k}_r = 0.0496/\text{h}$ , and  $\bar{k}_g = 0.0997/\text{h}$ , and the 95% credible intervals are  $D \in [372, 517]$ ,  $k_r \in [0.0466, 0.0526]$  and  $k_g \in [0.0935, 0.1061]$ . In the univariate marginals the 95% credible intervals are shown in red vertical dashed lines, in the bivariate marginals the region of maximum density is shown in the darkest shade. Results in (j)–(m) show the solution of Equations (1)–(2) with  $\theta = (1000, 0.025, 0.025)$ . MCMC results use  $\sigma = 0.5$ .

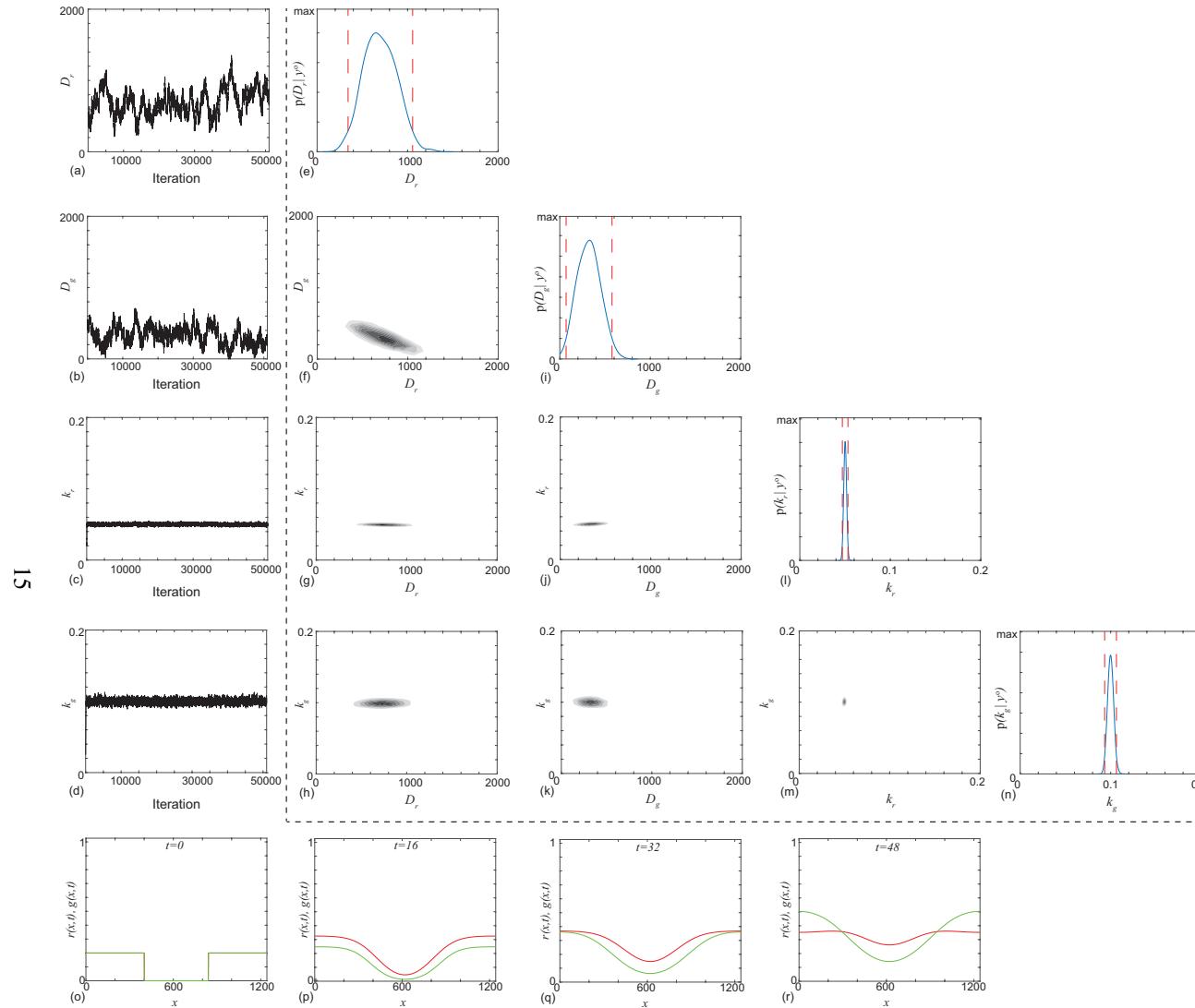


Figure S8: Typical Markov chain iterations, of length 51,000, for  $D_r$ ,  $D_g$ ,  $k_r$  and  $k_g$  in (a)–(d), respectively. In this case the Markov chain is initiated with  $\theta = (500, 500, 0.0250, 0.0250)$ . In this case the true value is  $\theta = (700, 300, 0.050, 0.10)$ . Results in (e)–(n) show a plot matrix representation of the univariate marginals and bivariate marginals estimated using the final 50,000 iterations of the Markov chain in (a)–(d). For the univariate distribution the posterior modes are  $\bar{D}_r = 658 \mu\text{m}^2/\text{h}$ ,  $\bar{D}_g = 320 \mu\text{m}^2/\text{h}$ ,  $\bar{k}_r = 0.0501/\text{h}$ , and  $\bar{k}_g = 0.0999/\text{h}$ . The 95% credible intervals are  $D_r \in [342, 1056]$ ,  $D_g \in [69, 577]$ ,  $k_r \in [0.0470, 0.03532]$  and  $k_g \in [0.0933, 0.1064]$ . In the univariate marginals the 95% credible intervals are shown in red vertical dashed lines, in the bivariate marginals the region of maximum density is shown in the darkest shade. Results in (o)–(r) show the solution of Equations (1)–(2) with  $\theta = (700, 300, 0.050, 0.100)$ . MCMC results use  $\sigma = 0.5$ .

## 6. Additional profile results for synthetic data

Results in Figures ??-?? show additional profiling results for the first and second scenarios discussed above, i.e. for  $D_r = D_g$ , with  $\theta = (500, 500, 0.05, 0.10)$ , and for  $D_r \neq D_g$ , with  $\theta = (700, 300, 0.05, 0.10)$ . As above, here we use finely resolved synthetic data, though now we also take finer measurements in time: we take observations at each spatial location of the solution grid, as well as observations at each hour between  $t = 0$  and  $t = 48$  h. This gives  $N = 1243 \times 48 = 59664$  observations (excluding the initial condition) of both of  $r$  and  $g$ , that is 119328 measurements of density in total. In Figure ?? we also consider data collected every hour but using the coarse experimental grid, given parameters set according to the second scenario. For all results we used the same measurement precision,  $\sigma = 0.05$ , as used for the experimental data.

In the first scenario where  $D_r = D_g$  we see sharply peaked profiles for  $D_r$ ,  $D_g$ ,  $k_r$  and  $k_g$  centred at the known values (Figure ??a-b, d-e). We also profile for  $D_r - D_g$  and obtain sharp profiles with a clear defined peak near zero, as expected (Figure ??c). In the second scenario where  $D_r \neq D_g$  and  $D_r - D_g = 400$ , and for high resolution spatiotemporal data (Figure ??), we again see sharply peaked profiles centred at the known values. The maximum likelihood estimates correspond to the true values in each scenario, i.e.  $\theta^* = (500, 0.050, 0.100)$  and  $\theta^* = (700, 300, 0.050, 0.100)$ , respectively.

In the case of high temporal resolution but coarse spatial resolution, we see better results than those from the experimental data, but that it is still difficult to distinguish  $D_r$  and  $D_g$  (Figure ??). For example, the approximate 95% confidence interval for the difference  $D_r - D_g$  is [-436, 902], whereas the real difference is 400.

These numerical results indicate that the model parameters are likely structurally identifiable, i.e. that the estimates can be expected to converge to the true values in the limit of increasing amounts of spatiotemporal data. The convergence rates of the diffusion coefficients, and the convergence rates of the difference in diffusion coefficients, appear to be slower than those for the rate parameters, however, thus indicating that they are more difficult to estimate in general. This is consistent with the practical identifiability results with real data in the main text. We also see that, while the finest resolution results indicate that the model is structurally identifiable, high temporal resolution by itself does not appear to be sufficient for good practical identifiability.

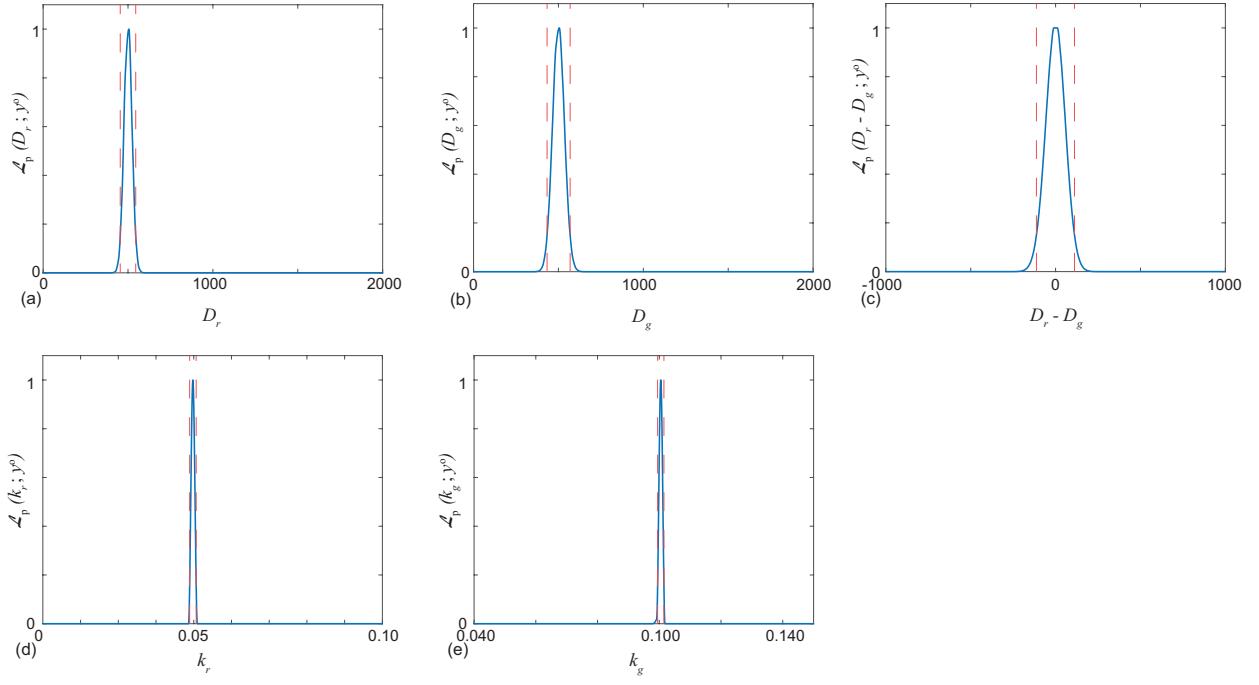


Figure S9: Profile likelihoods for synthetic data with  $\theta = (500, 500, 0.05, 0.10)$ , representing increasing amounts of spatiotemporal data. Measurements are collected at each spatial and temporal point of the numerical solution, between  $t = 0$  to  $t = 48$  h. Each profile likelihood is computed using a grid of 100 equally-spaced points for the target parameter; cubic interpolation is used for display and to determine confidence intervals. Approximate 95% confidence intervals are indicated based on a relative likelihood cutoff of 0.15 [? ]. Interval estimates are:  $D_r \in [454, 545]$ ,  $D_g \in [433, 569]$ ,  $D_r - D_g \in [-113, 111]$ ,  $k_r \in [0.0489, 0.0506]$  and  $k_g \in [0.0995, 0.1015]$ . Maximum likelihood estimates are:  $D_r^*, D_g^*, (D_r - D_g)^*, k_r^*, k_g^* = 500, 500, 0.0, 0.05, 0.10$ . Our estimates can thus be seen to be converging to the true values in the limit of increasing amounts of spatiotemporal data. Results used  $\sigma = 0.05$ .

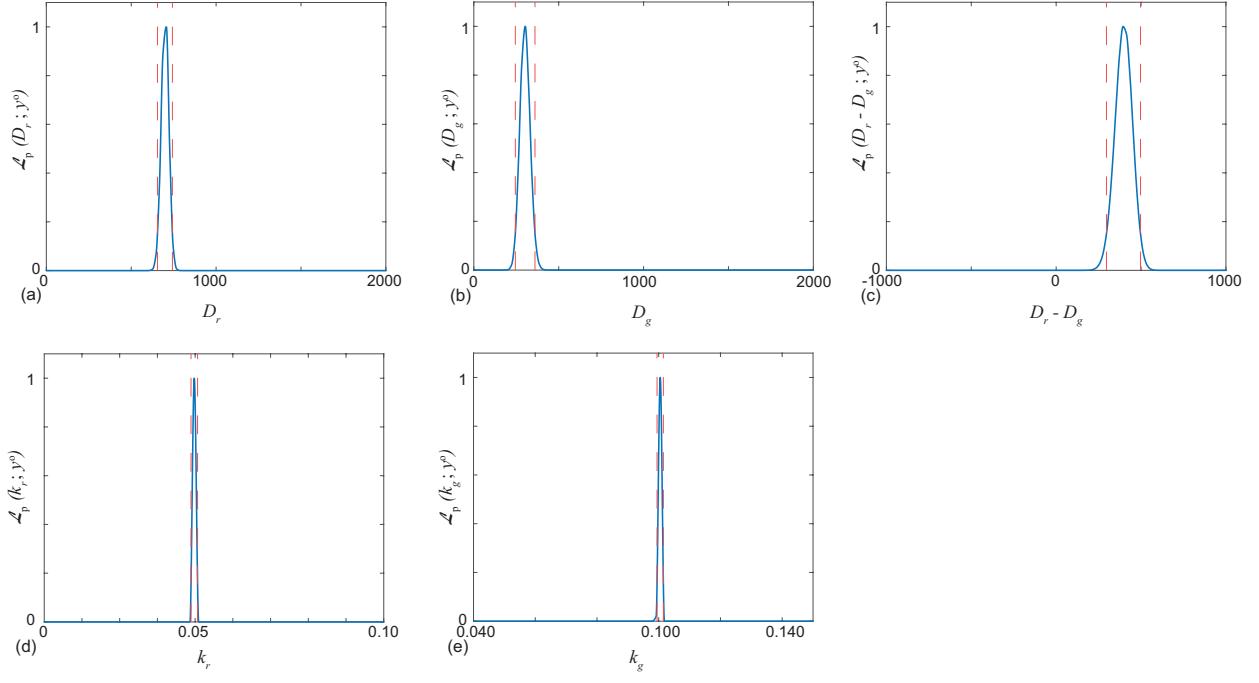


Figure S10: Profile likelihoods for synthetic data with  $\theta = (700, 300, 0.05, 0.10)$ , representing increasing amounts of spatiotemporal data. Measurements are collected at each spatial and temporal point of the numerical solution, between  $t = 0$  to  $t = 48$  h. Each profile likelihood is computed using a grid of 100 equally-spaced points; cubic interpolation is used for display and to determine confidence intervals. Approximate 95% confidence intervals are indicated based on a relative likelihood cutoff of 0.15 [?]. Interval estimates are:  $D_r \in [655, 743]$ ,  $D_g \in [245, 360]$ ,  $D_r - D_g \in [296, 497]$ ,  $k_r \in [0.0489, 0.0506]$  and  $k_g \in [0.0995, 0.1015]$ . Maximum likelihood estimates are:  $D_r^*, D_g^*, (D_r - D_g)^*, k_r^*, k_g^* = 700, 300, 400, 0.05, 0.10$ . Our estimates can thus be seen to be converging to the true values in the limit of increasing amounts of spatiotemporal data. Results used  $\sigma = 0.05$ .

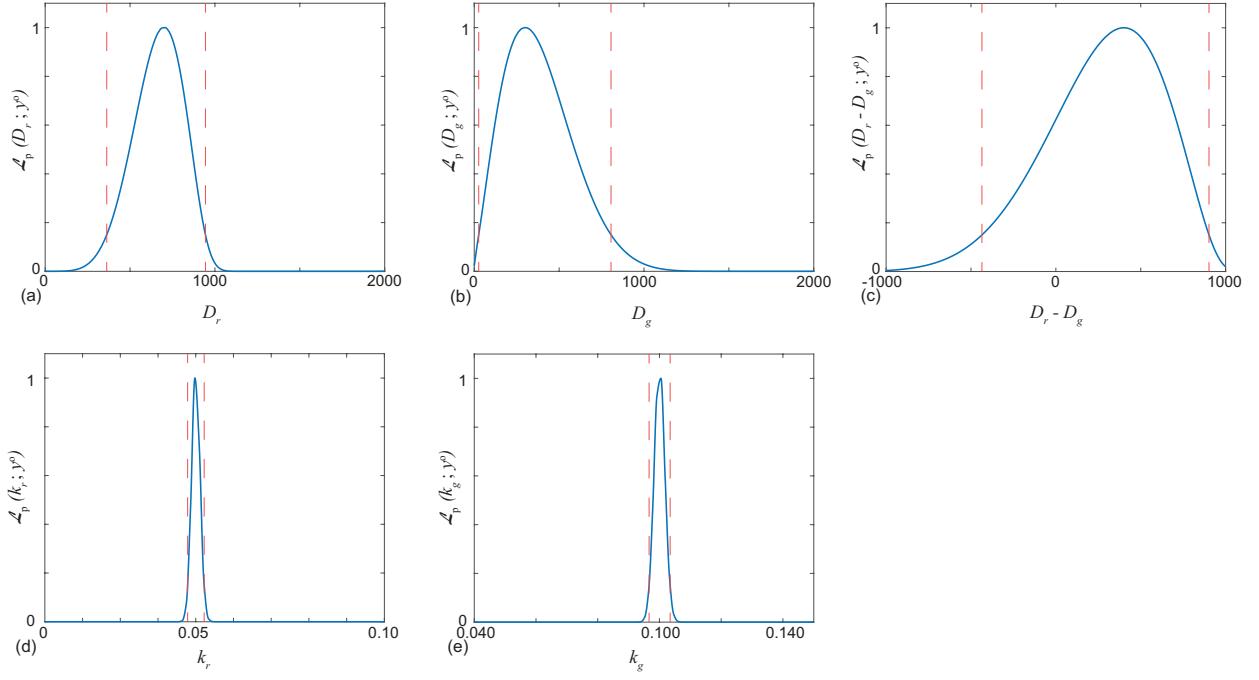


Figure S11: Profile likelihoods for synthetic data with  $\theta = (700, 300, 0.05, 0.10)$ , representing the case of increasing temporal resolution and fixed, experimental spatial resolution. Measurements are collected every hour between  $t = 0$  to  $t = 48$  at the same spatial grid points as the experimental results in the main text. Each profile likelihood is computed using a grid of 100 equally-spaced points; cubic interpolation is used for display and to determine confidence intervals. Approximate 95% confidence intervals are indicated based on a relative likelihood cutoff of 0.15 [? ]. Interval estimates are:  $D_r \in [363, 944]$ ,  $D_g \in [26.4, 805]$ ,  $D_r - D_g \in [-436, 902]$ ,  $k_r \in [0.0478, 0.0522]$  and  $k_g \in [0.0966, 0.1035]$ . Maximum likelihood estimates are:  $D_r^*, D_g^*, (D_r - D_g)^*, k_r^*, k_g^* = 700, 300, 400, 0.05, 0.10$ . Our estimates can thus be seen to be converging to the true values in the limit of increasing amounts of temporal data, but at a slower rate than when spatial resolution is also increased. Results used  $\sigma = 0.05$ .

- [1] Mathworks. 2020. lsqnonlin. Solve nonlinear least-squares. <https://au.mathworks.com/help/optim/ug/lsqnonlin.html>. (Accessed January 2020).
- [2] Pawitan Y. 2001. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press.