Supplement to

Analysis of CFSE proliferation assay using division-, age- and label-structured population models

Sabrina Hross[1,2] and Jan Hasenauer[1,2]

[1]Helmholtz Zentrum München - German Research Center for Environmental Health, Institute of Computational Biology, 85764 Neuherberg, Germany

[2]Technische Universität München, Center for Mathematics, Chair of Mathematical Modeling of Biological Systems, 85748 Garching, Germany

Contents

1 Modeling of CFSE proliferation assay

In this section, we summarize the division-, age- and label-structured population (DALSP) model, a generalization of existing modeling approaches for CFSE time-series data (Figure 1). We provide the model and the sensitivity equations along with decompositions which simplify the numerical evaluation.

1.1 Model

We start by introducing the DALSP model for the proliferation dynamics of cell populations (Metzger *et al.*, 2012) as well as a model for the measurement processes (Hasenauer, 2013).

Notation: We denote the set of nonnegative real numbers by $\mathbb{R}_+ := [0, \infty)$ and the set of natural numbers with zero by $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. The units used in the following equations are number of cells (cells), unit of concentration (UC), unit of fluorescence intensity (UFI) and unit of time (UT). For simplicity, we assume that age and time are measured in the same units, generalizations are however straight forward.

1.1.1 Proliferation dynamics

We use the number density $n(a, x, i|t)$ (in cells/UC/UT) to describe the state of a proliferating cell population. The different variables are the cell age a (in UT), the label concentration x (in UC), the number of cell divisions i (without unit) and the time t (in UT). The number density $n(a, x, i|t)$ evolves according to a systems of partial differential equations (PDEs),

$$\forall i \geq 0: \quad \frac{\partial n(a,x,i|t)}{\partial t} + \frac{\partial n(a,x,i|t)}{\partial a} + \frac{\partial (\nu(t,x) n(a,x,i|t))}{\partial x} = -\left(\alpha_i(t,a) + \beta_i(t,a)\right) n(a,x,i|t) \tag{1}$$

with initial conditions (ICs)

$$\begin{aligned} i = 0: &\quad n(a,x,i|0) \equiv n_0(a) p_0(x), \\ \forall i \geq 1: &\quad n(a,x,i|0) \equiv 0, \end{aligned} \tag{2}$$

and boundary conditions (BCs)

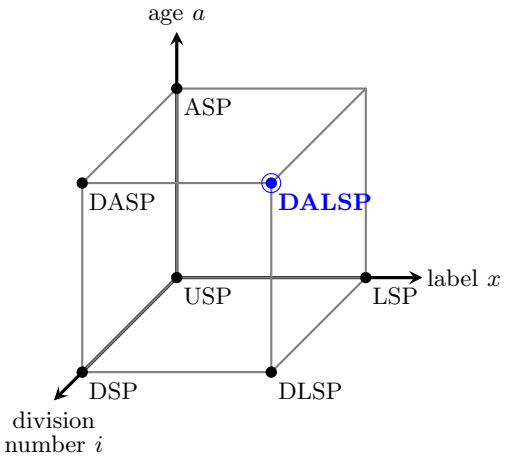
$$\begin{aligned} &i = 0: n(0,x,0|t) \equiv 0, \\ &i \geq 1: n(0,x,i|t) \equiv 4 \int_{\mathbb{R}_+} \alpha_{i-1}(t,a) n(a,2x,i-1|t) da. \end{aligned} \tag{3}$$

The i-th PDE describes the dynamics of the subpopulation of cells with i cell divisions. The product structure of the initial condition, $n(a, x, i|0) \equiv n_0(a)p_0(x)$, is assumed to allow for a simple solution algorithm (see Section 1.5). Furthermore, it is biologically plausible as there are no indications that the labelling efficiency depends on the cell age.

1.1.2 Measurement process

To assess the proliferation dynamics, cell populations are analysed at different time points using cell counting and flow cytometry. Via cell counting the overall size of the cell population, $N(t)$ (in cells),

$$N(t) = \sum_{i=1}^{\infty} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} n(a,x,i|t) dx da, \tag{4}$$



Supplement Figure 1: Illustration of the common models (USP, DSP, LSP, DLSP, ASP and DASP model) and the DALSP model considered in this work. The axis represent properties of individual cell for which some of the model account.

is determined. Flow cytometry measures in addition the fluorescence y_m (in UFI) of individual cells. The measured fluorescence y_m is the sum of the CFSE induced fluorescence, $y = cx$ with proportionality constant $c > 0$ (UFI/UC) and background fluorescence y_b, hence $y_m = y + y_b$ (Hasenauer, 2013). Information about cell age a and division number i is generally not assessable. Accordingly, the distribution of y_m is given by the convolution integral

$$p_m(y_m|t) = \int_0^{y_m} p(y|t)p_b(y_m - y)dy, \tag{5}$$

in which $p_b(y_b)$ denotes the distribution of background fluorescence and $p(y|t)$ denotes the distribution of the label induced fluorescence. The distribution of the label induced fluorescence is

$$p(y|t) = \frac{1}{c} p\left(x = \frac{y}{c} \middle| t\right), \tag{6}$$

with label concentration distribution

$$p(x|t) = \frac{1}{N(t)} \sum_{i \in \mathbb{N}_0} \int_{\mathbb{R}_+} n(a, x, i|t)da. \tag{7}$$

In most studies, neither the label concentration, x, nor the proportionality constant, c, are of interest. Merely, the information about the label induced fluorescence, and its distribution within the cell population shall be used to infer the proliferation properties. This can be employed to avoid the estimation of c. It can be shown that the PDE model (1)-(3) also governs the dynamics of the distribution of the label induced fluorescence, when substituting x (in UC) by y (in UFI) and rescaling the initial distribution. This proves that the estimation of c can be avoided by restating the model and implies that any choice of c is valid, e.g., $c = 1$ (in UFI/UC).

1.2 Sensitivity equation

For parameter estimation the sensitivities of the measured quantities, $N(t)$ and $p_m(y_m|t)$, with respect to the model parameters $\theta \in \mathbb{R}^{n_\theta}$ are required. We parametrise the potentially unknown rates of cell division and cell death, the rate of label degradation as well as the initial age distribution, the initial fluorescence distribution and the autofluorescence distribution. In this section we derive the sensitivity equation of the state variables for the DALSP model, $n(a, x, i|t)$, and subsequently the sensitivity equations for the measured quantities, $N(t)$ and $p_m(y_m|t)$.

1.2.1 Proliferation dynamics

The sensitivity of the number density $n(a, x, i|t)$ with respect to parameter θ_j is

$$s_j^n(a, x, i|t) := \frac{\partial n(a, x, i|t)}{\partial \theta_j}. \tag{8}$$

This sensitivity evolves according to the forward sensitivity equation

$$\begin{aligned}\forall i \geq 0 : &\frac{\partial s_j^n(a,x,i|t)}{\partial t} + \frac{\partial s_j^n(a,x,i|t)}{\partial a} + \frac{\partial}{\partial x}\left(\frac{\partial \nu(t,x)}{\partial \theta_j} n(a,x,i|t) + \nu(t,x) s_j^n(a,x,i|t)\right) \\ &= -\left(\frac{\partial \alpha_i(t,a)}{\partial \theta_j} + \frac{\partial \beta_i(t,a)}{\partial \theta_j}\right) n(a,x,i|t) - (\alpha_i(t,a) + \beta_i(t,a))\, s_j^n(a,x,i|t)\end{aligned} \tag{9}$$

with initial conditions (ICs)

$$\begin{aligned} i = 0 : &\quad s_j^n(a,x,i|0) \equiv \frac{\partial n_0(a)}{\partial \theta_j} p_0(x) + n_0(a) \frac{\partial p_0(x)}{\partial \theta_j}, \\ \forall i \geq 1 : &\quad s_j^n(a,x,i|0) \equiv 0, \end{aligned} \tag{10}$$

and boundary conditions (BCs)

$$\begin{aligned} &i = 0 : s_j^n(0,x,0|t) \equiv 0, \\ &i \geq 1 : s_j^n(0,x,i|t) \equiv 4 \int_{\mathbb{R}_+} \left(\frac{\partial \alpha_{i-1}(t,a)}{\partial \theta_j} n(a,2x,i-1|t) + \alpha_{i-1}(t,a) s_j^n(a,2x,i-1|t\right) da, \end{aligned} \tag{11}$$

This forward sensitivity equation is obtained by differentiating (1)-(3) with respect to the parameter θ_j and subsequent reordering. By simultaneous solution of the systems (1)-(3) and the forward sensitivity equation (9)-(11) we obtain a more robust estimate compared to finite differences and the computation is often more efficient.

1.2.2 Measurement process

For a parameter θ_j the sensitivities of the population size $N(t)$ and the label distribution $p(x|t)$ are obtained by differentiating (4) and (5) with respect to θ_j. We obtain

$$\frac{\partial N(t)}{\partial \theta_j} = \sum_{i=1}^{\infty} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \frac{\partial n(a,x,i|t)}{\partial \theta_j} dx da \tag{12}$$

and

$$\frac{\partial p_m(y_m|t)}{\partial \theta_j} = \int_0^{y_m} \left(\frac{\partial p(y|t)}{\partial \theta_j} p_b(y_m - y) + p(y|t) \frac{\partial p_b(y_m - x)}{\partial \theta_j}\right) dy, \tag{13}$$

$$\text{with} \qquad \frac{\partial p(y|t)}{\partial \theta_j} = -\frac{1}{c^2} \frac{\partial c}{\partial \theta_j} p\left(x = \frac{y}{c} \middle| t\right) + \frac{1}{c} \frac{\partial p\left(x = \frac{y}{c} \middle| t\right)}{\partial \theta_j}, \tag{14}$$

$$\text{and} \qquad \frac{\partial p(x|t)}{\partial \theta_j} = \frac{1}{N(t)} \left(\sum_{i=1}^{\infty} \int_{\mathbb{R}_+} \frac{\partial n(a,x,i|t)}{\partial \theta_j} da - p(x|t) \frac{\partial N(t)}{\partial \theta_j}\right). \tag{15}$$

If the proportionality constant is known or not fitted, (14) simplifies to

$$\frac{\partial p(y|t)}{\partial \theta_j} = \frac{1}{c} \left.\frac{\partial p(x|t)}{\partial \theta_j}\right|_{x=\frac{y}{c}} \tag{16}$$

The sensitivities can be evaluated given the state sensitivities $s_j^n(a,x,i|t) = \partial n(a,x,i|t)/\partial \theta_j$ and a parametrised distribution for $p_b(y_b)$. If the background fluorescence distribution is not parametrised but modelled based on control experiments, the derivative $\partial p_b(y_m - x)/\partial \theta_j$ is zero.

1.3 Decomposition of solution

The DALSP model is a systems of coupled 2-dimensional PDEs. The numerical simulation of its state variables and sensitivities can be time-consuming. To reduce the computational complexity, we reuse a dimensionally reducing decomposition method we introduced for similar models (Hasenauer *et al.*, 2012).

1.3.1 Model

We proved previously that for degradation rates $\nu(t,x) = -k(t)x$ the solution of (1)-(3) can be written as a product,

$$n(a,x,i|t) = n(a,i|t)p(x|i,t), \tag{17}$$

of the solution of a division- and age-structured population (DASP) model, $n(a,i|t)$, and the solution to a simple label-structured population (LSP) model, $p(x|i,t)$ (Metzger *et al.*, 2012). The DASP model governing the evolution of $n(a,i|t)$ is

$$\forall i \geq 0 : \frac{\partial n(a,i|t)}{\partial t} + \frac{\partial n(a,i|t)}{\partial a} = -(\alpha_i(t,a) + \beta_i(t,a))n(a,i|t), \tag{18}$$

with ICs

$$\begin{aligned} i = 0 &: n(a,i|0) \equiv n_0(a), \\ i \geq 1 &: n(a,i|0) \equiv 0, \end{aligned} \tag{19}$$

and BCs

$$\begin{aligned} i = 0 &: n(0,i|t) \equiv 0, \\ i \geq 1 &: n(0,i|t) \equiv 2\int_{\mathrm{R}_+} \alpha_{i-1}(t,a)n(a,i-1|t)da. \end{aligned} \tag{20}$$

The simple LSP models governing the evolution of $p(x|i,t)$ are

$$\forall i \geq 0 : \frac{\partial p(x|i,t)}{\partial t} + \frac{\partial(\nu(t,x)p(x|i,t))}{\partial x} = 0, \qquad p(x,i|0) \equiv 2^i p_0(2^i x). \tag{21}$$

1.3.2 Sensitivity equation

A similar decomposition can also be applied to the solution of the sensitivity equation (9)-(11) by exploiting that

$$s_j^n(a,x,i|t) = \frac{\partial n(a,i|t)p(x|t)}{\partial \theta_j} = \frac{\partial n(a,i|t)}{\partial \theta_j}p(x|t) + n(a,i|t)\frac{\partial p(x|t)}{\partial \theta_j}, \tag{22}$$

with

$$s_j^n(a,i|t) \coloneqq \frac{\partial n(a,i|t)}{\partial \theta_j} \qquad \text{and} \qquad s_j^p(x|t) \coloneqq \frac{\partial p(x|t)}{\partial \theta_j}. \tag{23}$$

The sensitivities $s_j^n(a,i|t)$ and $s_j^p(x|t)$ denote the sensitivities of the solutions of the aforementioned DASP (18)-(20) and LSP (21) models. The sensitivity equations are obtained by differentiating (18)-(20) and (21), respectively. The resulting evolution equation for the sensitivity of the solution of the DASP model is

$$\forall i \geq 0 : \frac{\partial s_j^n(a,i|t)}{\partial t} + \frac{\partial s_j^n(a,i|t)}{\partial a} = -\left(\frac{\partial \alpha_i(t,a)}{\partial \theta_j} + \frac{\partial \beta_i(t,a)}{\partial \theta_j}\right) n(a,i|t) - (\alpha_i(t,a) + \beta_i(t,a))s_j^n(a,i|t), \tag{24}$$

with ICs

$$\begin{aligned} i = 0 &: s_n(a,i|0) \equiv \frac{\partial n_0(a)}{\partial \theta_j}, \\ i \geq 1 &: s_n(a,i|0) \equiv 0, \end{aligned} \tag{25}$$

and BCs

$$\begin{aligned} i = 0 &: n(0,i|t) \equiv 0, \\ i \geq 1 &: n(0,i|t) \equiv 2\int_{\mathrm{R}_+} \frac{\partial \alpha_{i-1}(t,a)}{\partial \theta_j} n(a,i-1|t) + \alpha_{i-1}(t,a)s_j^n(a,i-1|t)da. \end{aligned} \tag{26}$$

The evolution of the sensitivity of the solution of the simple LSP models is governed by

$$\forall i \geq 0 : \frac{\partial s_j^p(x|i,t)}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\partial \nu(t,x)}{\partial \theta_j}p(x|i,t) + \nu(t,x)s_j^p(x|i,t)\right) = 0, \qquad s_j^p(x,i|0) \equiv 2^i \frac{\partial p_0(2^i x)}{\partial \theta_j}. \tag{27}$$

The decomposition of solution and sensitivity enables the calculation of the solution by solving a systems of coupled 1-dimensional PDEs for the division and age structure and a set of decoupled PDEs for the label structure instead of a system of coupled 2-dimensional PDEs. This does not introduce an approximation error and due to the reduced dimensionality it can be simulated more efficiently.

1.4 Analytical results

The decompositions yields a DASP model and set of simple LSP models. These models allow for an analytical analysis. In this section we present analytical solutions of the models and sensitivity equations.

1.4.1 Division- and age-structured population model

The dynamics of the individual subpopulations in the DASP model (18)-(20) are governed by a *Von Foerster*-like equation (von Foerster, 1959). By using the solution of Trucco (Trucco, 1965) for *Von Foerster*-like equations repeatedly for the individual subpopulations, we could prove that

$$\begin{aligned} i=0: n(a,0|t) &= \begin{cases} n_0(a-t)\exp\left(-\int_0^t \chi_0(\tilde{t},\tilde{t}+a-t)d\tilde{t}\right) & ,t\leq a \\ 0 & ,t>a \end{cases} \\ i\geq 1: n(a,i|t) &= \begin{cases} 0 & ,t\leq a \\ 2\int_{\mathrm{R}_+}\alpha_{i-1}(t-a,\tilde{a})n(\tilde{a},i-1|t-a)d\tilde{a}\exp\left(-\int_0^a \chi_i(\tilde{a}+t-a,\tilde{a})d\tilde{a}\right) & ,t>a, \end{cases} \end{aligned} \tag{28}$$

with $\chi_i(t,a)=\alpha_i(t,a)+\beta_i(t,a)$ (Metzger *et al.*, 2012). By differentiating expression (28) with respect to θ_j, we obtain an expression for the solution of the sensitivity equation (24)-(26),

$$i=0: s_j^n(a,0|t) = \begin{cases} \exp\left(-\int_0^t \chi_0(\tilde{t},\tilde{t}+a-t)d\tilde{t}\right)\left(\frac{\partial n_0(a-t)}{\partial\theta_j} - n_0(a-t)\int_0^t \frac{\partial\chi_0(\tilde{t},\tilde{t}+a-t)}{\partial\theta_j}d\tilde{t}\right) & ,t\leq a \\ 0 & ,t>a \end{cases}$$

$$i\geq 1: s_j^n(a,i|t) = \begin{cases} 0 & ,t\leq a \\ 2\exp\left(-\int_0^a \chi_i(\tilde{a}+t-a,\tilde{a})d\tilde{a}\right) & ,t>a, \\ \quad\cdot\left(\int_{\mathrm{R}_+}\frac{\partial\alpha_{i-1}(t-a,\tilde{a})}{\partial\theta_j}n(\tilde{a},i-1|t-a)+\alpha_{i-1}(t-a,\tilde{a})s_j^n(\tilde{a},i-1|t-a)d\tilde{a}\right. & \\ \qquad\left.-\int_{\mathrm{R}_+}\alpha_{i-1}(t-a,\tilde{a})n(\tilde{a},i-1|t-a)d\tilde{a}\int_0^a\frac{\partial\chi_i(\tilde{a}+t-a,\tilde{a})}{\partial\theta_j}d\tilde{a}\right) & \end{cases} \tag{29}$$

with $\dfrac{\partial\chi_i(t,a)}{\partial\theta_j}=\dfrac{\partial\alpha_i(t,a)}{\partial\theta_j}+\dfrac{\partial\beta_i(t,a)}{\partial\theta_j}$.

The expressions (28) and (29) will be important for the evaluation of states and the sensitivities of the DALSP model. These expressions provide closed-form solutions of $n(a,i|t)$ and $s_j^n(a,i|t)$ for special cases of $\alpha(t,a)$ and $\beta(t,a)$, e.g., if both are constant. For general $\alpha(t,a)$ and $\beta(t,a)$, solutions of the integrals contained in (28) and (29) cannot be expressed analytically and, accordingly, we do not get closed-form solutions for $n(a,i|t)$ and $s_j^n(a,i|t)$.

1.4.2 Simple label-structured population model

For the solution of the simple LSP models (21) we derived previously the expression

$$\forall i\geq 0:\ p(x|i,t)=2^i\exp\left(\int_0^t k(\tilde{t})d\tilde{t}\right)p_0\left(2^i\exp\left(\int_0^t k(\tilde{t})d\tilde{t}\right)x\right) \tag{30}$$

(Metzger *et al.*, 2012). Differentiation of this expression with respect to the parameter θ_j yields an expression for the solution of the sensitivity equation (27), namely

$$\forall i\geq 0:\ s_j^p(x|i,t)=2^i\exp\left(\int_0^t k(\tilde{t})d\tilde{t}\right)\left(\left(\int_0^t\frac{\partial k(\tilde{t})}{\partial\theta_j}d\tilde{t}\right)p_0\left(2^i\exp\left(\int_0^t k(\tilde{t})d\tilde{t}\right)x\right)+\frac{dp_0\left(2^i\exp\left(\int_0^t k(\tilde{t})d\tilde{t}\right)x\right)}{d\theta_j}\right). \tag{31}$$

The expressions (30) and (31) provide closed-form solution for some choices of $k(t)$. Constant degradation $k(t)=k$ and Gompertz decay process $k(t)=k_{\max}\exp(-k_T t)$ yield

$$\forall i\geq 0:\ p(x|i,t)=2^i\exp(kt)p_0\left(2^i\exp(kt)x\right) \tag{32}$$

and

$$\forall i \geq 0 : \; p(x|i,t) = 2^i \exp\left(\frac{k_{\max}}{k_T}\left(1-\exp(-k_T t)\right)\right) p_0\left(2^i \exp\left(\frac{k_{\max}}{k_T}\left(1-\exp(-k_T t)\right)\right) x\right), \tag{33}$$

respectively. For general $k(t)$, numerical methods have to be used to compute the integral $\int_0^t k(\tilde{t})d\tilde{t}$.

1.5 Numerical evaluation

As closed-form solutions are generally not available, this section addresses the numerical calculation of state and outputs of the models.

1.5.1 Proliferation dynamics

Division-, age- and label-structured population model: To compute the solution of the DALSP model and its sensitivities, we exploit the decompositions (17) and (22). Accordingly, states and sensitivities can be computed by multiplication/addition of the solutions of the DASP model and the simple LSP models. If for the given rates $\alpha(t,a)$, $\beta(t,a)$ and $k(t)$ closed-form solutions for the DASP model and the simple LSP models are available, these solutions are used. Otherwise, numerical methods are employed.

Division- and age-structured population model: The numerical solution of the DASP model is obtained by iterative evaluation of the expression (28). Starting with $i = 0$ the solutions $n(a,i|t)$ for the individual subpopulations are computed. We use a constant discretization of time, space and age. For numerical integration the trapezoidal rule is used, ensuring positivity of the resulting solutions. The integration error can be evaluated be refining the discretization.

We exploit a similar iterative procedure to evaluate the sensitivity of the DASP model. As the calculation of the sensitivities requires the solution for the state, states and sensitivities are evaluated simultaneously. This bears the advantages that we can reuse several parts of the expressions (28) and (29), e.g. the integrals over χ. The derivatives of $\alpha_i(t,a)$ and $\beta_i(t,a)$ are computed using symbolic differentiation.

Simple label-structured population model: The expressions for solution of the simple LSP models (30) and the corresponding sensitivity equations (31) is rather simple. Using symbolic expressions for the initial label distribution $p_0(x)$ and the degradation rate $k(t)$, merely integrals of $k(t)$ and its derivative have to be computed numerically to evaluate the expressions (30) and (31). The remaining parts can be obtained by symbolic calculations. As in this manuscript only linear and Gompertz decay is considered, even closed-form solutions for the integrals over $k(t)$ and $\partial k(t)/\partial\theta_j$ are available.

1.5.2 Measurement process

Population size: The population size $N(t)$ and its derivative $\partial N/\partial\theta_j$ is computed from the solution of the DASP model and its sensitivity by integration over the age a and subsequent summation over the division number i. For numerical integration the trapezoidal rule is employed.

Fluorescence distribution: The measured fluorescence distribution $p_m(y_m|t)$ is given by the convolution integral (5). This convolution integral can be computed using standard numerical integration methods.

In this study we have chosen a different approach. We assume that the initial distribution of CFSE is a weighted sum of log-normal distributions,

$$p_0(x) = \sum_{l=1}^{L} w_l \log\mathcal{N}(x|\mu_{l,0},\sigma_{l,0}^2), \tag{34}$$

with weights $w_l \in [0,1]$, with $\sum_{l=1}^{L} w_l = 1$, location parameter $\mu_{l,0}$ and scale parameter $\sigma_{l,0}^2$. Additionally, we assume that the autofluorescence is log-normally distributed $p_b(y_b) = \log\mathcal{N}(y_b|\mu_b,\sigma_b^2)$. These assumptions are in general not limiting as smooth distributions with positive support can be approximated by a sum of log-normal distributions and autofluorescence levels are known to be often roughly log-normally distributed (see, e.g., (Hawkins *et al.*, 2007)).

We proved previously that given (34), the time-dependent distribution of CFSE in the subpopulations i is

$$p(x|i,t) = \sum_{l=1}^{L} w_l \log \mathcal{N}(x|\mu_{i,l}^x(t), \sigma_{l,0}^2), \tag{35}$$

with

$$\mu_{i,l}^x(t) = -i\log(\gamma) - \int_0^t k(\tilde{t})d\tilde{t} + \mu_{l,0}. \tag{36}$$

This follows from the expression (30) for $p(x|i,t)$ and for details we refer to (Hasenauer, 2013). This property has been denoted as log-normal invariance. Due to the fact that log-normal distributions are conserved under the flow of the simple LSP model.

From (35) and (36), we obtain the distribution of CFSE induced fluorescence in the subpopulation for $y = cx$,

$$p(y|i,t) = \sum_{l=1}^{L} w_l \log \mathcal{N}(y|\mu_{i,l}(t), \sigma_{l,0}^2), \tag{37}$$

with

$$\mu_{i,l}(t) = -i\log(\gamma) - \int_0^t k(\tilde{t})d\tilde{t} + \mu_{l,0} + \log(c). \tag{38}$$

Using this formulation, the convolution integral (5) can be restated as

$$p(y_m|t) = \sum_{i\in\mathbb{N}_0} \frac{N(i|t)}{N(t)} \sum_{l=1}^{L} w_l \int_{\mathbb{R}_+} \log \mathcal{N}(y|\mu_{i,l}(t), \sigma_{l,0}^2) \log \mathcal{N}(y - y_m|\mu_b, \sigma_b^2)dy, \tag{39}$$

with $N(i|t) = \int_{\mathbb{R}_+} n(a,i|t)da$ denoting the number of cells in subpopulation i at time point t. The cell number $N(i|t) = \int_{\mathbb{R}_+} n(a,i|t)da$ is computed using numerical integration with the trapezoidal rule. The convolution integral of the product of the log-normal distributions is approximated using the method proposed by Fenton (1960). Fenton's approximation employs that the sum of log-normally distributed random variables is approximately log-normally distributed with expectation and variance of the actual distribution. Expectation and variance of $p(y|i,t)$ and $p_b(y_b)$ are

$$\mathrm{E}_y^{i,l}(t,\theta) = \exp\left(\mu_{i,l}(t) + \frac{\sigma_{l,0}^2}{2}\right), \quad \mathrm{Var}_y^{i,l}(t,\theta) = \exp\left(2\mu_{i,l}(t) + \sigma_{l,0}^2\right)\left(\exp\left(\sigma_{l,0}^2\right) - 1\right), \tag{40}$$

and

$$\mathrm{E}_{y_b} = \exp\left(\mu_b + \frac{\sigma_b^2}{2}\right), \quad \mathrm{Var}_{y_b} = \exp\left(2\mu_b + \sigma_b^2\right)\left(\exp\left(\sigma_b^2\right) - 1\right), \tag{41}$$

respectively. Accordingly, expectation and variance of the actual distribution of the sum are

$$\mathrm{E}_{y_m}^{i,l}(t) = \mathrm{E}_y^{i,l}(t) + \mathrm{E}_{y_b}, \quad \mathrm{Var}_{y_m}^{i,l}(t) = \mathrm{Var}_y^{i,l}(t) + \mathrm{Var}_{y_b}. \tag{42}$$

The log-normal distribution with this expectation and variance possesses the location parameter

$$\hat{\mu}_{y_m,i,l}(t) = \log(\mathrm{E}_{y_m}^{i,l}(t)) - \frac{1}{2}\log\left(\frac{\mathrm{Var}_{y_m}^{i,l}(t)}{\left(\mathrm{E}_{y_m}^{i,l}(t)\right)^2} + 1\right) \tag{43}$$

and the scale parameter

$$\hat{\sigma}_{y_m,i,l}^2(t) = \log\left(\frac{\mathrm{Var}_{y_m}^{i,l}(t)}{\left(\mathrm{E}_{y_m}^{i,l}(t)\right)^2} + 1\right), \tag{44}$$

yielding the approximation

$$\int_{\mathbb{R}_+} \log \mathcal{N}(y|\mu_{i,l}(t), \sigma_{l,0}^2) \log \mathcal{N}(y - y_m|\mu_b, \sigma_b^2)dy \approx \log \mathcal{N}(y_m|\hat{\mu}_{y_m,i,l}(t), \hat{\sigma}_{y_m,i,l}^2(t)). \tag{45}$$

This approximation of the individual convolution integrals in (39) yields the overall fluorescence distribution

$$p(y_m|t) \approx \sum_{i\in\mathbb{N}_0} \frac{N(i|t)}{N(t)} \sum_{l=1}^{L} w_l \log \mathcal{N}(y_m|\hat{\mu}_{y_m,i,l}(t), \hat{\sigma}^2_{y_m,i,l}(t)), \tag{46}$$

While this is an approximation, our studies revealed that it is in general very accurate. Indeed, it is often more than a rough numerical integration. For an evaluation of Fenton's approximation and a comparison to alternative approximation we refer to Kapraun (2014).

The sensitivity of the approximate fluorescence distribution (47) is obtained by differentiation,

$$\begin{aligned}\frac{\partial p(y_m|t)}{\partial \theta_j} \approx \sum_{i\in\mathbb{N}_0} \frac{1}{N(t)^2} \left(\frac{\partial N(i|t)}{\partial \theta_j} N(t) - N(i|t) \frac{\partial N(t)}{\partial \theta_j} \right) \sum_{l=1}^{L} w_l \log \mathcal{N}(y_m|\hat{\mu}_{y_m,i,l}(t), \hat{\sigma}^2_{y_m,i,l}(t)) \\ + \sum_{i\in\mathbb{N}_0} \frac{N(i|t)}{N(t)} \sum_{l=1}^{L} \log \mathcal{N}(y_m|\hat{\mu}_{y_m,i,l}(t), \hat{\sigma}^2_{y_m,i,l}(t)) \cdot \left(\frac{\partial w_l}{\partial \theta_j} + w_l \frac{\log y_m - \hat{\mu}_{y_m,i,l}(t)}{\hat{\sigma}^2_{y_m,i,l}(t)} \frac{\partial \hat{\mu}_{y_m,i,l}(t)}{\partial \theta_j} \right. \\ \left. + w_l \left(\frac{(\log y_m - \hat{\mu}_{y_m,i,l}(t))^2}{\hat{\sigma}^3_{y_m,i,l}(t)} - \frac{1}{\hat{\sigma}_{y_m,i,l}(t)} \right) \frac{\partial \hat{\sigma}_{y_m,i,l}(t)}{\partial \theta_j} \right)\end{aligned} \tag{47}$$

The derivative $\partial N(i|t)/\partial\theta_j$ is computed by integration $\partial n(a,i|t)/\partial\theta_j$ with respect to age a and the derivative $\partial N(t)/\partial\theta_j$ is discussed above. The weights w_l are parameters and their derivatives $\partial w_l/\partial\theta_j$ can be computed directly. To compute the derivatives of $\hat{\mu}_{y_m,i,l}(t)$ and $\hat{\sigma}_{y_m,i,l}(t)$, we differentiate (43) and (44). The resulting expression can be evaluated analytically in the case of linear and Gompertz decay.

1.5.3 Implementation

All steps of the numerical evaluation are implemented in MATLAB (version R2013b).

2 Parameter Estimation and Model Selection

In this section, we outline the parameter estimation and model selection methods used in this study. Among others, we introduce the likelihood function and its gradient, optimisation using multi-start and particle-swarm methods and uncertainty analysis using profile likelihoods and Markov chain Monte Carlo sampling.

2.1 Likelihood function

2.1.1 Noise models

Data for CFSE proliferation assays are collected using cell counting and flow cytometry. Cell counting provides the overall population size. In this study we assume that the counting data are corrupted by multiplicative, log-normally distributed measurement noise. After log-transformation we obtain

$$\log \bar{N}_k = \log N(t_k) + \epsilon, \qquad \text{with} \qquad \epsilon \sim \mathcal{N}(0, \sigma^2_{\text{noise},N}). \tag{48}$$

Assuming independence of time points, we obtain the likelihood function

$$\mathbb{P}\left(\{\bar{N}_k\}_{k=1}^K|\theta\right) = \prod_{k=1}^{K} \mathbb{P}\left(\bar{N}_k|\theta\right) = \prod_{k=1}^{K} \mathcal{N}(\log \bar{N}_k|\log N(t_k), \sigma^2_{\text{noise},N}), \tag{49}$$

in which the predicted cell number $N(t_k)$ depends implicitly on the parameters θ.

Flow cytometry measurements provide information about the fluorescence of individual cells. As cytometry data are mostly binned into finite many intervals the data possess a limited resolution and are usually presented as histograms. For each bin, $j = 1,\ldots,J$, and time point t_k, $k = 1,\ldots,K$, the number of cells with measured fluorescence intensities between the lower bound $\xi_{j,\text{lb}}$ and upper bound $\xi_{j,\text{ub}}$ of the j-th histogram bin is denoted by $\bar{H}^j_k$. The likelihood of observing a particular histogram $\{\bar{H}^j_k\}_{j=1}^J$ follows a multinomial distribution (Merran *et al.*, 2000),

$$\mathbb{P}\left(\{\bar{H}^j_k\}_{j=1}^J|\theta\right) = \frac{\left(\sum_{j=1}^J \bar{H}^j_k\right)!}{\prod_{j=1}^J \bar{H}^j_k!} \prod_{j=1}^{J} \left(p(x_{m,k} \in I_j|\theta)\right)^{\bar{H}^j_k}, \tag{50}$$

with $p(x_{m,k} \in I_j|\theta)$ denoting the probability of observing a cell at time point t_k a cell with fluorescence $x_{m,k} \in I_j = (\xi_{j,\mathrm{lb}}, \xi_{j,\mathrm{ub}}]$. This likelihood function for cytometry data has been introduced simultaneously by Thompson (2012) and ourselves (Hasenauer *et al.*, 2011).

In the absence of outliers, the probability $p(y_m \in I_j|t_k)$ is the integral of $p(y_m|t_k)$ over I_j. In the presence of outliers, we obtain

$$p(y_m \in I_j|t_k) = \int_{\xi_{j,\mathrm{lb}}}^{\xi_{j,\mathrm{ub}}} ((1 - w_{\mathrm{outliers}})p(y_m|t_k) + w_{\mathrm{outliers}}\, p_{\mathrm{outliers}}(y_m))\, dy_m, \tag{51}$$

$$= (1 - w_{\mathrm{outliers}}) \int_{\xi_{j,\mathrm{lb}}}^{\xi_{j,\mathrm{ub}}} p(y_m|t_k) dy_m + w_{\mathrm{outliers}} \int_{\xi_{j,\mathrm{lb}}}^{\xi_{j,\mathrm{ub}}} p_{\mathrm{outliers}}(y_m) dy_m, \tag{52}$$

in which w_{outliers} denotes the probability that a measurement is an outlier and $p_{\mathrm{outliers}}(y_m)$ denote the distribution of outliers. As the outlier distribution is in general unknown, we assume it to be a uniform distribution between lower and upper bound, $\xi_{1,\mathrm{lb}}$ and $\xi_{J,\mathrm{ub}}$. The outlier probability w_{outliers} can either be estimated from the data or chosen beforehand. Note that according to this outlier model, with a probability of w_{outliers} any value can be observed.

Given the likelihood function $\mathbb{P}\left(\{\bar{H}_k^j\}_{j=1}^J|\theta\right)$, the distribution of bin counts can be assessed as each individual bin $\bar{H}_k^j$ is binomially distributed,

$$\bar{H}_k^j \sim \mathrm{bino}\left(\sum_{j=1}^J \bar{H}_k^j, p(y_m \in I_j|t_k)\right). \tag{53}$$

This can be used to determine confidence intervals for the respective measurements. The $100(1-\alpha)\%$ confidence interval for the measured histogram counts is

$$\mathrm{CI}_\alpha(\bar{H}_k^j) = \left[\mathrm{binoinv}\left(\frac{\alpha}{2}\middle|\sum_{j=1}^J \bar{H}_k^j, p(y_m \in I_j|t_k)\right), \mathrm{binoinv}\left(1-\frac{\alpha}{2}\middle|\sum_{j=1}^J \bar{H}_k^j, p(y_m \in I_j|t_k)\right)\right], \tag{54}$$

in which 'binoinv' denotes the inverse of the binomial cumulative distribution function.

2.1.2 Likelihood function

To estimate the model parameters from measurements of the overall population size and binned snapshot data, $\mathcal{D} = \{\{\bar{H}_k^j\}_{j=1}^J, \bar{N}_k\}_{k=1}^K$, we use the likelihood function

$$\mathbb{P}(\mathcal{D}|\theta) = \prod_{k=1}^K \mathbb{P}(\{\bar{H}_k^j\}_{j=1}^J|\theta)\mathbb{P}(\bar{N}_k|\theta). \tag{55}$$

This likelihood function assumes independence of the measurement noise, which should in general be fulfilled. For optimisation and sampling we will in the following use the log-likelihood functions,

$$\log \mathbb{P}(\mathcal{D}|\theta) = \sum_{k=1}^K \log \mathbb{P}(\{\bar{H}_k^j\}_{j=1}^J|\theta) + \sum_{k=1}^K \log \mathbb{P}(\bar{N}_k|\theta), \tag{56}$$

with

$$\log \mathbb{P}\left(\bar{N}_k|\theta\right) = -\frac{1}{2}\left(\log(2\pi\sigma^2_{\mathrm{noise},N}) + \frac{(\log \bar{N}_k - \log N(t_k))^2}{\sigma^2_{\mathrm{noise},N}}\right), \tag{57}$$

$$\log \mathbb{P}\left(\{\bar{H}_k^j\}_{j=1}^J|\theta\right) = \sum_{n=1}^{\sum_{j=1}^J \bar{H}_k^j} \log k - \sum_{j=1}^J \sum_{k=1}^{\bar{H}_k^j} \log k + \sum_{j=1}^J \bar{H}_k^j \log p(y_m \in I_j|t_k). \tag{58}$$

The log-likelihood facilitates a more robust numerical evaluation. Furthermore, it is more suited for optimisations due to improved curvature characteristics.

The model properties on which the likelihood depends are the parameter-dependent population size $N(t)$ and the parameter-dependent bin probabilities $p(y_m \in I_j|t_k)$. The calculation of the former has been addressed in Section 1. The probability $p(y_m \in I_j|t_k)$ is computed using expression (47). Numerical integration yields directly the first term in (52). The second term of (52) is evaluated analytically using the analytical formulae of the outlier distribution.

For a detailed discussion of similar likelihood functions we refer to Hasenauer (2013).

2.1.3 Gradient of likelihood function

For efficient local optimisation, the gradient of the log-likelihood function is required. This gradient is

$$\frac{\partial \log \mathbb{P}(\mathcal{D}|\theta)}{\partial \theta_j} = \sum_{k=1}^{K} \frac{\partial \log \mathbb{P}(\{\bar{H}_{t_k}^j\}_{j=1}^J|\theta)}{\partial \theta_j} + \sum_{k=1}^{K} \frac{\partial \log \mathbb{P}(\bar{N}_k|\theta)}{\partial \theta_j}, \tag{59}$$

with

$$\frac{\partial \log \mathbb{P}\left(\bar{N}_k|\theta\right)}{\partial \theta_j} = \frac{1}{2}\frac{1}{\sigma^2_{\text{noise},N}}\left(\frac{(\log \bar{N}_k - \log N(t_k))^2}{\sigma^2_{\text{noise},N}} - 1\right)\frac{\partial \sigma^2_{\text{noise},N}}{\partial \theta_j} + \frac{\log \bar{N}_k - \log N(t_k)}{N(t_k)\sigma^2_{\text{noise},N}}\frac{\partial N(t_k)}{\partial \theta_j}, \tag{60}$$

$$\frac{\partial \log \mathbb{P}\left(\{\bar{H}_k^j\}_{j=1}^J|\theta\right)}{\partial \theta_j} = \sum_{j=1}^{J} \frac{\bar{H}_k^j}{p(y_m \in I_j|t_k)}\frac{\partial p(y_m \in I_j|t_k)}{\partial \theta_j}. \tag{61}$$

The calculation of the derivative $\partial N(t_k)/\partial \theta_j$ is discussed in Section 1.2.2. To calculate $\partial p(y_m \in I_j|t_k)/\partial \theta_j$, we differentiate (52) with respect to θ_j. This yields

$$\begin{aligned}\frac{\partial p(y_m \in I_j|t_k)}{\partial \theta_j} = &+ (1 - w_{\text{outliers}})\int_{\xi_{j,\text{lb}}}^{\xi_{j,\text{ub}}} \frac{\partial p(y_m|t_k)}{\partial \theta_j} dy_m + w_{\text{outliers}}\int_{\xi_{j,\text{lb}}}^{\xi_{j,\text{ub}}} \frac{\partial p_{\text{outliers}}(y_m)}{\partial \theta_j} dy_m \\ &- \frac{\partial w_{\text{outliers}}}{\partial \theta_j}\left(\int_{\xi_{b,\text{lb}}}^{\xi_{j,\text{ub}}} p(y_m|t_k)dy_m - \int_{\xi_{j,\text{lb}}}^{\xi_{j,\text{ub}}} p_{\text{outliers}}(y_m)dy_m\right).\end{aligned} \tag{62}$$

We computed the different integrals in this work using the trapezoidal rule, thereby ensuring positivity of the numerically calculated probabilities. Alternatively, the integrals over $p(y_m|t_k)$ and its derivative might be computed by exploiting the fact that it is a log-normal distribution and the cumulative distribution function is known. Similarly, for $p_{\text{outliers}}(y_m)$ and its derivative also analytical formulas might be available.

2.2 Optimisation

To estimate the parameters of the DALSP model from experimental data, we solve the maximum likelihood estimation problem

$$\theta = \arg\max_{\theta \in \Omega} \mathbb{P}(\mathcal{D}|\theta)$$

by minimizing the negative log-likelihood function (55),

$$\theta = \arg\min_{\theta \in \Omega} \{\mathcal{J}(\theta) \coloneqq -\log \mathbb{P}(\mathcal{D}|\theta)\}.$$

The optima of both optimisation problems coincide, the use of the negative log-likelihood function is numerically however often advantageous. The search domain is the hypercube $\Omega = \{\theta \in \mathbb{R}^{n_\theta} | \theta_{\min} \leq \theta \leq \theta_{\max}\}$.

The negative log-likelihood function $\mathcal{J}$ can possess local minima, therefore, we exploit global optimisation methods.

2.2.1 Multi-start local optimisation

The most simplistic global optimisation method is a multi-start local optimisation approach. This approach achieves global exploration by initializing (deterministic) local optimisers at many different starting points.

In this study, the starting points for the local optimisations were generated using latin hypercube sampling. Local optimisation was performed using the interior point algorithm implemented in the MATLAB function `fmincon.m`. To facilitate good convergence we set function tolerances (`TolFun`) and parameter tolerances (`TolX`) to 10^{-9}. Furthermore, we increased the maximum number of iterations (`MaxIter`) to 10^3 and the maximum number of function evaluations (`MaxFunEvals`) to $10^3 \times n_\theta$, in which n_θ denotes the number of parameters of the considered model. If the gradient of the negative log-likelihood function J was computed, we set `GradObj` to `on`. Otherwise, `fmincon.m` internally computes a finite difference approximations of the gradient.

The multi-start implementation is implemented in our in-house Parameter EStimation TOolbx (PESTO).

2.3 Particle swarm optimisation

In addition to multi-start local optimisation methods we considered the particle swarm optimisation method implemented in the MATLB toolbox PSwarm (Vaz & Vicente, 2007). We have chosen PSwarm as the implemented global optimisation method proved effective for a variety of optimisation problems. It outperformed other global optimisers in a series of tests (Vaz & Vicente, 2007).

For our evaluation we used a population size (`options.Size`) of $10 \times n_\theta$ and a maximum number of function evaluations (`options.MaxObj`) of either $10^3 \times n_\theta$ or $10^4 \times n_\theta$. To ensure that the number of iterations was not limiting, we set the maximum number of function evaluations (`options.MaxIter`) to the same values as the maximum number of function evaluations (`options.MaxObj`). All value were higher than the default setting to improve the convergence of the optimiser.

2.4 Comparison of optimisers

We compared the performance of optimisers in terms of the computation time per 'converged start'. Unlike many other publications, we exploit a statistical definition for 'converged start'. We call a start converged if it cannot be rejected compared to the best result. For this hypothesis testing problem we used the likelihood ratio test. Accordingly, a start was converged to significance level α if the inequality

$$\log \mathbb{P}(\mathcal{D}|\theta) > \log \mathbb{P}(\mathcal{D}|\theta_{\text{global}}) - \frac{\Delta_\alpha}{2}, \tag{63}$$

holds. Here θ was the result of the current optimisation run, θ_{global} was the best available estimate and Δ_α is the $100(1-\alpha)$th percentile of the χ^2-distribution with one degree of freedom. The average computation time per converged start $t_{\text{effective}}$ is the overall computation time for all starts divided by the number of converged starts,

$$t_{\text{effective}} = \frac{\text{overall computation time for all starts}}{\#\text{ number converged starts}}. \tag{64}$$

2.5 Uncertainty analysis

Parameter uncertainties can be assessed using Bayesian and frequentist methodologies. In this study, we combined both approaches to ensure reliability of our results (Hug *et al.*, 2013; Raue *et al.*, 2013).

2.5.1 Profile likelihood calculation

To compute parameter confidence intervals profile likelihoods, a frequentist methodology, can be used. The profile likelihood for parameter θ_i is the maximal feasible value of the likelihood function for a given value of θ_i (Murphy & van der Vaart, 2000). Mathematically, the profile likelihood is defined by a constrained optimisation problem for each value of θ_i,

$$\text{PL}_i(\theta_i) = \arg\max_{\substack{\bar{\theta}\in\Omega \\ \bar{\theta}_i=\theta_i}} \mathbb{P}(\mathcal{D}|\bar{\theta}). \tag{65}$$

For the profile likelihood calculation we employed sequential local optimisation. Starting at the optimum, the parameter θ_i was slightly increased or decreased and the remaining parameters were optimised using the MATLAB function `fmincon.m` with the aforementioned settings. This procedure was repeated until the full profile likelihood was calculated. Parameter confidence intervals were then derived from the profile likelihood $\text{PL}_i(\theta_i)$ using significance cut-offs determined using the likelihood ratio test. For further details we refer to (Raue *et al.*, 2009).

The profile likelihood calculation method we used is implemented in PESTO and similar to other available codes (Hug *et al.*, 2013; Raue *et al.*, 2009, 2015).

2.5.2 Markov chain Monte-Carlo sampling

Bayesian uncertainty analysis relies on Bayes' theorem,

$$\pi(\theta|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|\theta)\pi(\theta)}{\mathbb{P}(\mathcal{D})}. \tag{66}$$

Bayes' theorem states that the posterior probability density $\pi(\theta|\mathcal{D})$ is the product of the likelihood $\mathbb{P}(\mathcal{D}|\theta)$ and the prior probability density $\pi(\theta)$ divided by the marginal probability $\mathbb{P}(\mathcal{D}) = \int \mathbb{P}(\mathcal{D}|\theta)\pi(\theta)d\theta$. Accordingly, besides the data it can account for prior information

In this study, we use a simple uniform prior in Ω. The resulting posterior distribution was sampled using the delayed rejection adaptive Metropolis Hastings algorithm implemented in the MATLAB toolbox DRAM (Haario *et al.*, 2006) which is called using PESTO. DRAM is a self-tuning Markov chain Monte-Carlo (MCMC) sampler which provided good convergence properties. Convergence after burn-in was assessed using the Geweke test. The parameter chain $\{\theta^{(l)}\}_l$ obtained using DRAM can be used to analyse parameter uncertainties as well as uncertainties of functions $g(\theta)$ of the parameters. To assess the latter, samples from the distribution $p(g|\mathcal{D}) = \int g(\theta)p(\theta|\mathcal{D})d\theta$ are generated by evaluating g for a representative parameter sample $\{\theta^{(l)}\}_l$. This is possible for any function g including, e.g., functions which require the simulation of the DALSP model.

3 Application

We exploit the DALSP model to quantify CFSE proliferation data presented by Luzyanina *et al.* (2007a). We used Data Set 2, which has also been analysed in other publications (see, e.g., Banks *et al.* (2011); Luzyanina *et al.* (2007b)) and is therefore well suited for the evaluation and comparison of our approach.

In this study, an ensemble of DASLP models is used to study the CFSE data collected on days 1-5. As a cell loss might have occurred between day 0 and day 1 (Luzyanina *et al.*, 2007b), experimental data for day 0 are disregarded and day 1 is considered as starting point for the analysis.

In the following, we shortly describe the setup and report the findings.

3.1 Lower and upper bounds of model parameters

For this study, we used the parameter bounds which are reported in Table 1. These lower and upper bounds have been chosen based on the following considerations:

Division rates: From day 1 to day 5 the number of cells increases roughly by a factor of 6. Assuming exponential growth, $\alpha_i(a) = k_\alpha$, without cell loss, $\beta_i(a) = 0$, this implies $k_\alpha = \log(6)/4 \approx 0.45\,\mathrm{d}^{-1}$. This provides a rough scale for k_α and $k_{\alpha,i}$. As this value might however be far off, lower bounds of $10^{-6}\,\mathrm{d}^{-1}$ and upper bounds of $10^{3}\,\mathrm{d}^{-1}$ are used for k_α and $k_{\alpha,i}$. These bounds are more than two orders of magnitude away from the simple estimate.

For the Hill exponent, n_α, and the Hill coefficient, K_α, lower and upper bounds of $10^{-6}\,\mathrm{d}$ and $10^{2}\,\mathrm{d}$ are used. This range of eight orders of magnitude allows for division rates which are almost constant ($K_\alpha \ll 1\,\mathrm{d}$ or $n_\alpha \ll 1$), far from saturation ($K_\alpha \gg 1\,\mathrm{d}$) and step-like ($n_\alpha \gg 1$).

Death rates: Bounds for the death rates are difficult to assess from the data. Therefore, we used the same bounds for k_β, $k_{\beta,i}$, K_β and n_β as for the corresponding parameters of the proliferation rates. These bounds allow for scenarios without cell death ($k_\alpha, k_{\alpha,i} \ll 1$), constant rates of cell death ($K_\beta \ll 1\,\mathrm{d}$ or $n_\beta \ll 1$) and a well defined survival time ($n_\alpha \gg 1$).

Label degradation: For the parameters k_{deg} and c_{deg}, lower bounds of $10^{-4}\,\mathrm{d}^{-1}$ and upper bounds of $10^{0}\,\mathrm{d}^{-1}$ are used. For $k_{\mathrm{deg}} = 1\,\mathrm{d}^{-1}$ and $c_{\mathrm{deg}} = 10^{-4}\,\mathrm{d}^{-1}$, more than 98% of the label are degraded from day 1 to day 5. For $k_{\mathrm{deg}} = 10^{-4}\,\mathrm{d}^{-1}$ and $c_{\mathrm{deg}} = 1\,\mathrm{d}^{-1}$, less than 0.1% of the label is degraded from day 1 to day 5. Accordingly, these lower and upper bounds allow for a spectrum of label properties, ranging from fast label degradation to high label stability.

Background fluorescence: The background fluorescence is assumed to be log-normally distributed with location parameter, μ_{noise}, and scale parameter, σ_{noise},

$$p_b(y_b) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{noise}} y_b} \exp\left\{-\frac{1}{2}\left(\frac{\log(y_b) - \mu_{\mathrm{noise}}}{\sigma_{\mathrm{noise}}}\right)^2\right\}. \tag{67}$$

Supplement Table 1: Nominal values, lower bounds, upper bounds and units for the model parameters. The nominal values are only used for model parameters which are not fitted.

	Parameter	Nominal value	Lower bound	Upper bound	Unit	Fitted
division rates	k_α	-	10^{-6}	10^3	d^{-1}	yes
	$k_{\alpha,i}$	-	10^{-6}	10^3	d^{-1}	yes
	K_α	-	10^{-6}	10^2	d	yes
	n_α	-	10^{-6}	10^2	-	yes
death rates	k_β	-	10^{-6}	10^3	d^{-1}	yes
	$k_{\beta,i}$	-	10^{-6}	10^3	d^{-1}	yes
	K_β	-	10^{-6}	10^2	d	yes
	n_β	-	10^{-6}	10^2	-	yes
label degradation	k_{deg}	-	10^{-4}	10^0	d^{-1}	yes
	c_{deg}	-	10^{-4}	10^0	d^{-1}	yes
background	μ_{noise}	-	0	4	-	yes
	σ_{noise}	-	10^{-1}	10^1	-	yes
initial cell population	N_0	-	10^0	10^4	cell	yes
	$r_{x,1}$	-	10^{-1}	10^1	-	yes
	$\mu_{x,1}$	-	6	8	-	yes
	$\mu_{x,2}$	-	6	8	-	yes
	$\sigma_{x,1}$	-	10^{-1}	10^0	-	yes
	$\sigma_{x,2}$	-	10^{-1}	10^0	-	yes
proportionality const.	c	1	-	-	UFI/UC	no

To obtain a plausible regime for these parameters, we fitted the fluorescence distribution on day 5 with a log-normal distribution, yielding $\mu_{\mathrm{noise}} \approx 3.5$ and $\sigma_{\mathrm{noise}} \approx 0.8$. As the fluorescence distribution observed on day 5 contains a significant portion of label induced fluorescence, wide lower and upper bounds are chosen which contain the estimates, namely, $\mu_{\mathrm{noise}} \in [0, 4]$ and $\sigma_{\mathrm{noise}} \in [10^{-1}, 10^1]$.

Initial conditions: The initial label distribution is modeled as a weighted sum of two log-normal distributions,

$$p_0(x) = \sum_{j=1}^{2} r_{x,j} \frac{1}{\sqrt{2\pi}\sigma_{x,j} x} \exp\left\{ -\frac{1}{2} \left(\frac{\log(x) - \mu_{x,j}}{\sigma_{x,j}} \right)^2 \right\}. \tag{68}$$

A plausible regime for the location parameters $\mu_{x,j}$ and scale parameter $\sigma_{x,j}$, with $j = 1, 2$, was obtained by fitting the fluorescence distribution of day 1 with a single log-normal distribution assuming the proportionality constant $c = 1$ UFI/UC. We obtained a location parameter of ≈ 6.6 and a scale parameter of ≈ 0.4. Based on this, we chose $\mu_{x,j} \in [6, 8]$ and $\sigma_{x,j} \in [10^{-1}, 10^0]$. The weighting of the log-normal distributions is $r_{x,1}$ and $r_{x,2} = (1 - r_{x,1})$ with $r_{x,1} \in [0, 1]$.

A plausible range for the size of the initial population, N_0, was derived from the measurement data. We used $N_0 \in [10^0, 10^4]$ cell.

3.2 Comparison of optimisation methods

We evaluated three different optimisation methods:

- multi-start local optimisation with sensitivity equations;
- multi-start local optimisation with finite differences; and
- particle swarm optimisation.

For model $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_{16}$ we report all results while for the remaining models merely the results for multi-start local optimisation with sensitivity equations are depicted.

For the model $\mathcal{M}_1$ which possessed merely 12 unknown parameters, all these optimisation method provided reasonable convergence result (Supplement Figure 2A). The highest percentage of converged starts was for

model $\mathcal{M}_1$ achieved using particle swarm optimisations. The highest computational efficiency, measured in run time per start (Supplement Figure 2B) as well as in average computation time per converged start (Supplement Figure 2C), is attained using multi-start local optimisation with sensitivity equations.

As the dimensionality of the parameter space increased, the convergence of all methods suffered. Particle swarm optimisation did not yield a single converged start for model $\mathcal{M}_{16}$ (Supplement Figure 2D). For multi-start local optimisation with sensitivity equations and finite differences we still observed convergence and merely the number of converged starts was reduced (Figure 3A and Supplement Figure 2D). This indicated that the particle swarm optimisation implemented in PSwarm scales significantly worse with the number of parameters than simple gradient-based optimisation methods.

For all models multi-start local optimisation with sensitivity equations achieved the best objective function value (see zooms in Figure 3A and Supplement Figure 2A,D). Furthermore, the average computation time per converged start, which is in our opinion the most important performance measured for a global optimiser, was at least 10-fold lower than for the other methods (Figure 3B and Supplement Figure 2C,F).

3.3 Multi-start local optimisation with forward sensitivity equation

As multi-start local optimisation with sensitivity equations yielded the best convergence and the highest computational efficiency, this methods has been used to optimise the remaining models, $\mathcal{M}_3$ to $\mathcal{M}_{15}$. The optimisation results for all models are depicted in Supplement Figure 3. As reference the relative difference to the likelihood of the most general model, $\mathcal{M}_{16}$, is shown. The y-axis in Supplement Figure 3 indicated $p(\mathcal{D}|\theta^{(16)}, \mathcal{M}_{16}) - p(\mathcal{D}|\theta^{(m)}, \mathcal{M}_m) + 1$. Note that we added the one toe allow for visualization in log-space.

Our analysis of the optimisation results indicated reproducibility of the results. For a few models, the global optimum was however found only a few times. The best fits of the individual models to the data are depicted in Supplement Figure 4 and 5.

3.4 Profile likelihood and Bayesian uncertainty analysis

For the models $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_{16}$ the parameter uncertainties were assessed. Therefore, we used profile likelihoods as well as samples from the posterior distribution. While the profile likelihoods were computed using PESTO, the MCMC samples were obtained using DRAM. As no convergence has been observed for random initial conditions, we started the sampling at the maximum likelihood estimate computed using multi-start local optimisation. In this case a burn-in of $2 \cdot 10^4$ samples and a sample size of 10^5 was sufficient for model $\mathcal{M}_1$ and $\mathcal{M}_2$ to achieve Geweke values > 0.9. For model $\mathcal{M}_{16}$ we used a burn-in of $1.5 \cdot 10^5$ samples and sampled $3.5 \cdot 10^5$ points, yielding Geweke values > 0.85. The slower convergence of DRAM for model $\mathcal{M}_{16}$ is probably related to the increased dimensionality of the parameter space and the parameter non-identifiabilities. A visualization of the chains is provided in Supplement Figure 6.

The profiles and sampling results for the models $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_{16}$ are depicted in Supplement Figure 7, 8 and 9. It is apparent that likelihood profiles and marginals of the MCMC samples agree well, indicating reliability of the results. The study of the scatter plot matrices depicted in Supplement Figure 10, 11 and 12 revealed parameter correlations. The path of the profile likelihood seemed to be roughly aligned with the main axis of the MCMC samples.

3.5 Division number-dependence of the rates of cell division

The model selection revealed that the age-dependence of the rates of cell division is most important for a good description of the experimental data. Second most important is the division number-dependence of the rate of cell division. As previous modelling (Banks *et al.*, 2014, 2015, 2013; Kapraun, 2014) and experimental work (Gett & Hodgkin, 2000) reported that the difference in the rates of cell division between the 0th-subpopulation (= no divisions) and the remaining subpopulations are most pronounced, we considered the model $\mathcal{M}_{2/4}$ with

$$\alpha_i(a) = \begin{cases} \frac{k_{\alpha,0} a^{n_\alpha}}{K_\alpha^{n_\alpha} + a^{n_\alpha}} & \text{for } i = 0 \\ \frac{k_{\alpha,1} a^{n_\alpha}}{K_\alpha^{n_\alpha} + a^{n_\alpha}} & \text{otherwise,} \end{cases} \tag{69}$$

and $\forall i \geq 0 : \beta_i(a) = k_\beta$. Model $\mathcal{M}_{2/4}$ is an intermediate between model $\mathcal{M}_2$ (age-dependent rate of cell division) and model $\mathcal{M}_4$ (division number- and age-dependent rate of cell division).

Model $\mathcal{M}_{2/4}$ was fitted to the experimental data using multi-start local optimisation with sensitivity equations, yielding an BIC value of 2.994×10^4. The comparison of the BIC values for models $\mathcal{M}_2$, $\mathcal{M}_{2/4}$ and

$\mathcal{M}_4$:

- BIC for $\mathcal{M}_2$: 3.058×10^4
- BIC for $\mathcal{M}_{2/4}$: 2.994×10^4
- BIC for $\mathcal{M}_4$: 2.974×10^4

reveals that

$$\frac{3.058 \times 10^4 - 2.994 \times 10^4}{3.058 \times 10^4 - 2.974 \times 10^4} = 76.2\% \tag{70}$$

of the improvement in the BIC value from $\mathcal{M}_2$ to $\mathcal{M}_4$ can be achieved by allowing different values of $k_{\alpha,i}$ for the 0th-subpopulation and the remaining subpopulations. The fit of the models $\mathcal{M}_2$, $\mathcal{M}_{2/4}$ and $\mathcal{M}_4$ is depicted in Supplement Figure 13.

References

Banks, H. T., Kapraun, D. F., Link, K. G., Thompson, W. C., Peligero, C., Argilaguet, J., & Meyerhans, A. (2014). Analysis of variability in estimates of cell proliferation parameters for cyton-based models using cfse-based flow cytometry data. *J. Inverse Ill-Posed Probl.*, *23*(2), 135–171.

Banks, H. T., Kapraun, D. F., Peligero, C., Argilaguet, J., & Meyerhans, A. (2015). Evaluating the importance of mitotic asymmetry in cyton-based models for cfse-based flow cytometry data. *Int. J. Pure Appl. Math.*, *100*(1), 131–156.

Banks, H. T., Kapraun, D. F., Thompson, W. C., Peligero, C., Argilaguet, J., & Meyerhans, A. (2013). A novel statistical analysis and interpretation of flow cytometry data. *J. Biol. Dyn.*, *7*(1), 96–132.

Banks, H. T., Sutton, K. L., Thompson, W. C., Bocharov, G., Doumic, M., Schenkel, T., Argilaguet, J., Giest, S., Peligero, C., & Meyerhans, A. (2011). A new model for the estimation of cell proliferation dynamics using CFSE data. *J. Immunological Methods*, *373*(1–2), 143–160.

Fenton, L. F. (1960). The sum of lognormal probability distributions in scatter transmission systems. *IRE Trans. Commun. Syst.*, *8*(1), 57—67.

Gett, A. V. & Hodgkin, P. D. (2000). A cellular calculus for signal integration by T cells. *Nat. Immunol.*, *1*(3), 239–244.

Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). DRAM: Efficient adaptive MCMC. *Stat. Comp.*, *16*(4), 339–354.

Hasenauer, J. (2013). *Modeling and parameter estimation for heterogeneous cell populations*. Ph.D. thesis, University of Stuttgart.

Hasenauer, J., Radde, N., Doszczak, M., Scheurich, P., & Allgöwer, F. (2011). Parameter estimation for the cme from noisy binned snapshot data: Formulation as maximum likelihood problem. Extended abstract at *Conf. of Stoch. Syst. Biol.*, Monte Verita, Switzerland.

Hasenauer, J., Schittler, D., & Allgöwer, F. (2012). Analysis and simulation of division- and label-structured population models: A new tool to analyze proliferation assays. *Bull. Math. Biol.*, *74*(11), 2692–2732.

Hawkins, E. D., Hommel, M., Turner, M. L., Battye, F. L., Markham, J. F., & Hodgkin, P. D. (2007). Measuring lymphocyte proliferation, survival and differentiation using CFSE time-series data. *Nat. Protoc.*, *2*(9), 2057–2067.

Hug, S., Raue, A., Hasenauer, J., Bachmann, J., Klingmüller, U., Timmer, J., & Theis, F. J. (2013). High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling. *Math. Biosci.*, *246*(2), 293–304.

Kapraun, D. F. (2014). *Cell proliferation models, CFSE-based flow cytometry data, and quantification of uncertainty*. Phd thesis, North Carolina State University, Raleigh, North Carolina, USA.

Luzyanina, T., Mrusek, S., Edwards, J., Roose, D., Ehl, S., & Bocharov, G. (2007a). Computational analysis of CFSE proliferation assay. *J. Math. Biol.*, *54*(1), 57–89.

Luzyanina, T., Roose, D., Schenkel, T., Sester, M., Ehl, S., Meyerhans, A., & Bocharov, G. (2007b). Numerical modelling of label-structured cell population growth using CFSE distribution data. *Theor. Biol. Med. Model.*, *4*, 26.

Merran, E., Hastings, N., & Peacock, B. (2000). *Statistical Distributions*. Wiley.

Metzger, P., Hasenauer, J., & Allgöwer, F. (2012). Modeling and analysis of division-, age-, and label-structured cell populations. In *Proceedings of 9th International Workshop on Computational Systems Biology*. Ulm, Germany: Tampere International Center for Signal Processing, 55–58.

Murphy, S. A. & van der Vaart, A. W. (2000). On profile likelihood. *J. Am. Stat. Assoc.*, *95*(450), 449–485.

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. *Bioinf.*, *25*(25), 1923–1929.

Raue, A., Kreutz, C., Theis, F. J., & Timmer, J. (2013). Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Phil. Trans. Royal Soc. A*, *371*(1984).

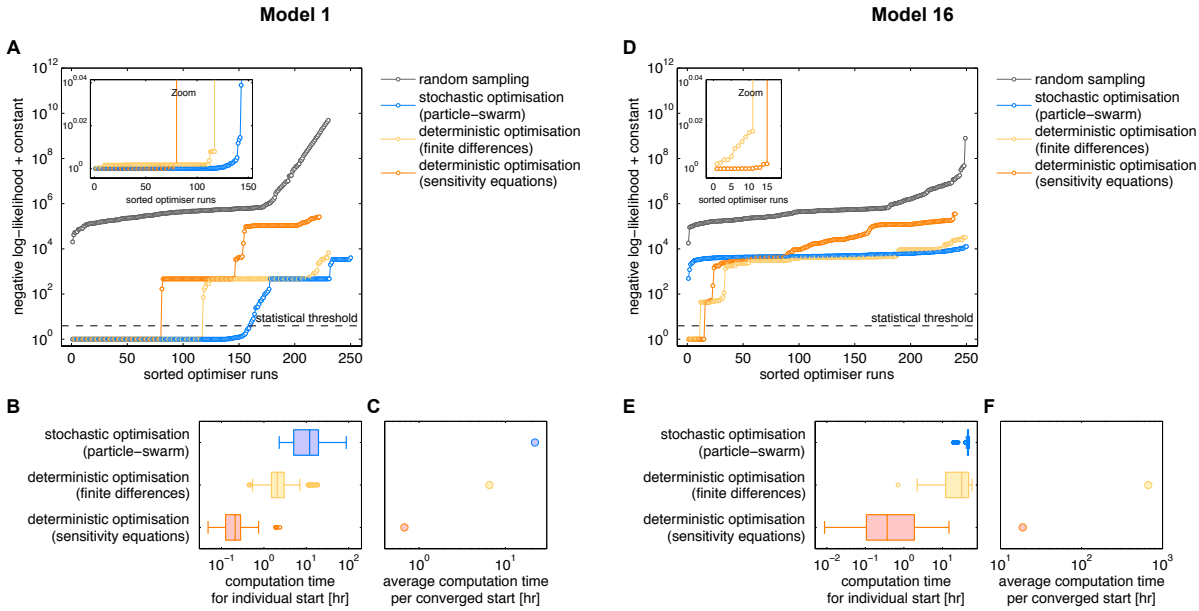
Raue, A., Steiert, B., Schelker, M., Kreutz, C., Maiwald, T., Hass, H., Vanlier, J., Tönsing, C., Adlung, L., Engesser, R., Mader, W., Heinemann, T., Hasenauer, J., Schilling, M., Höfer, T., Klipp, E., Theis, F. J., Klingmüller, U., Schöberl, B., & J.Timmer (2015). Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*.

Thompson, W. C. (2012). *Partial differential equation modeling of flow cytometry data from CFSE-based proliferation assays*. Ph.d. thesis, North Carolina State University.

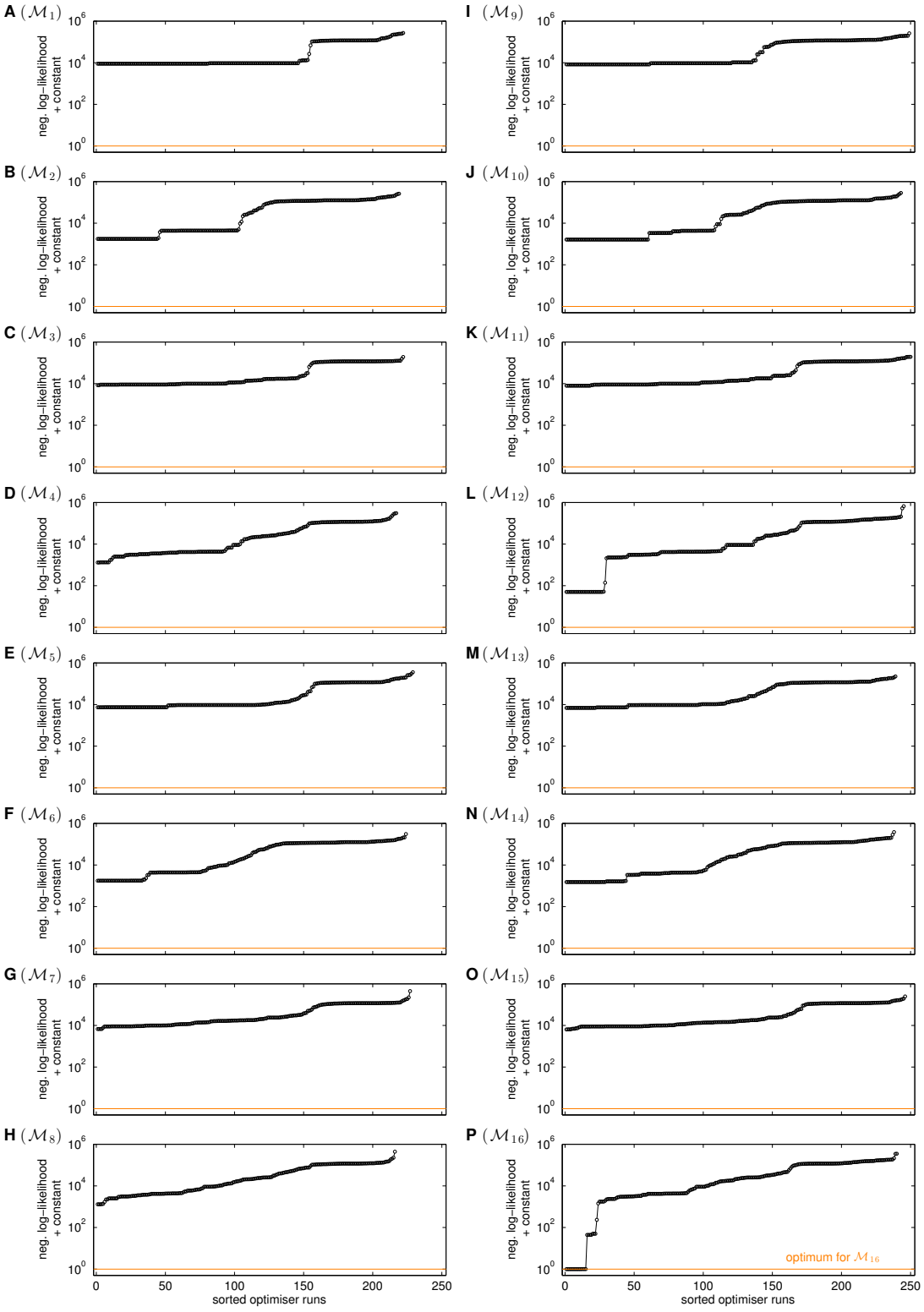
Trucco, E. (1965). Mathematical models for cellular systems the von foerster equation. Part i. *Bull. Math. Biol.*, *27*(3), 285–304.

Vaz, A. & Vicente, L. (2007). A particle swarm pattern search method for bound constrained global optimization. *J. Global Optim.*, *39*(2), 197–219.

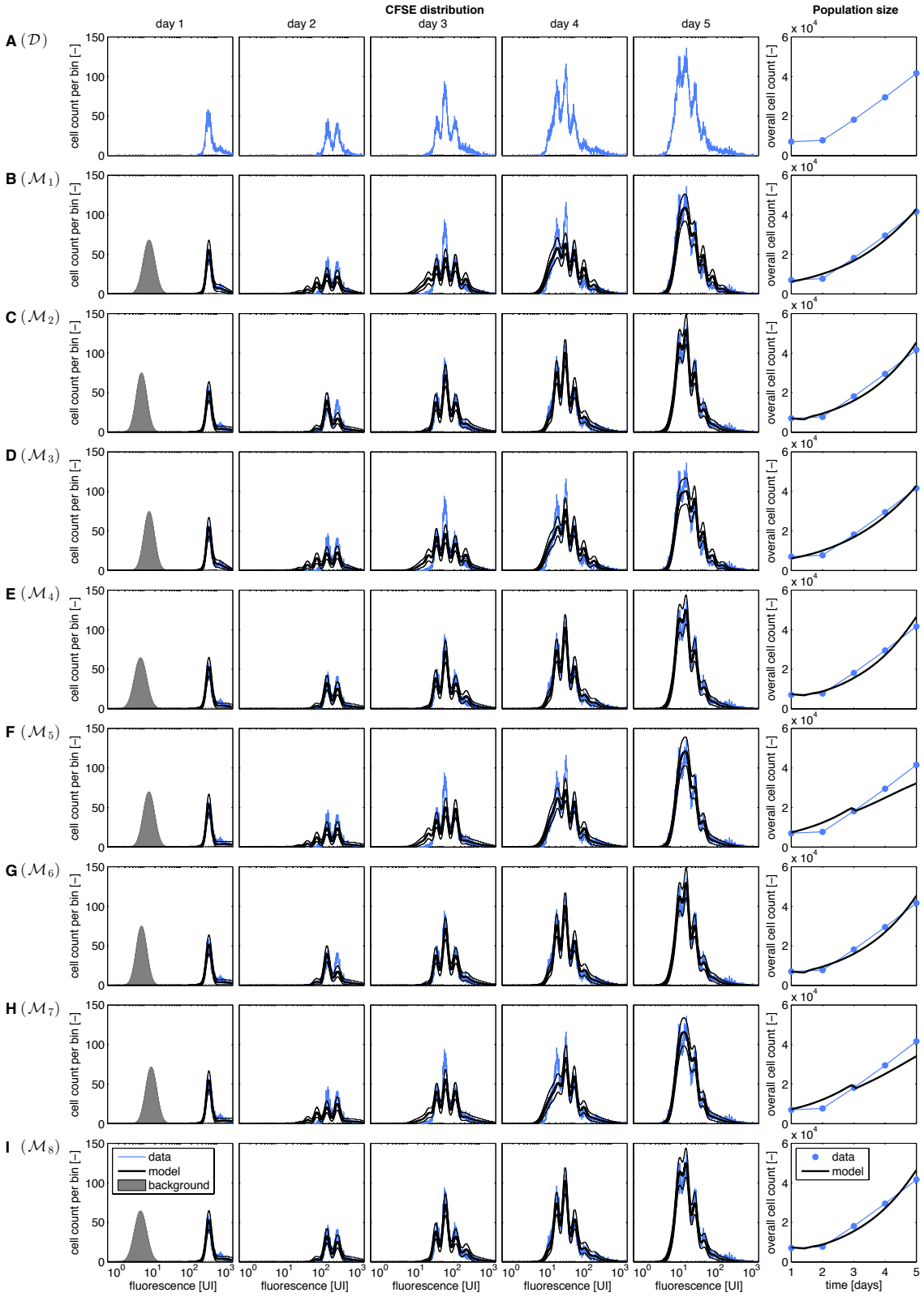
von Foerster, H. (1959). Some remarks on changing populations. *The kinetics of cellular proliferation*, New York: Grune and Stratton. 382–407.



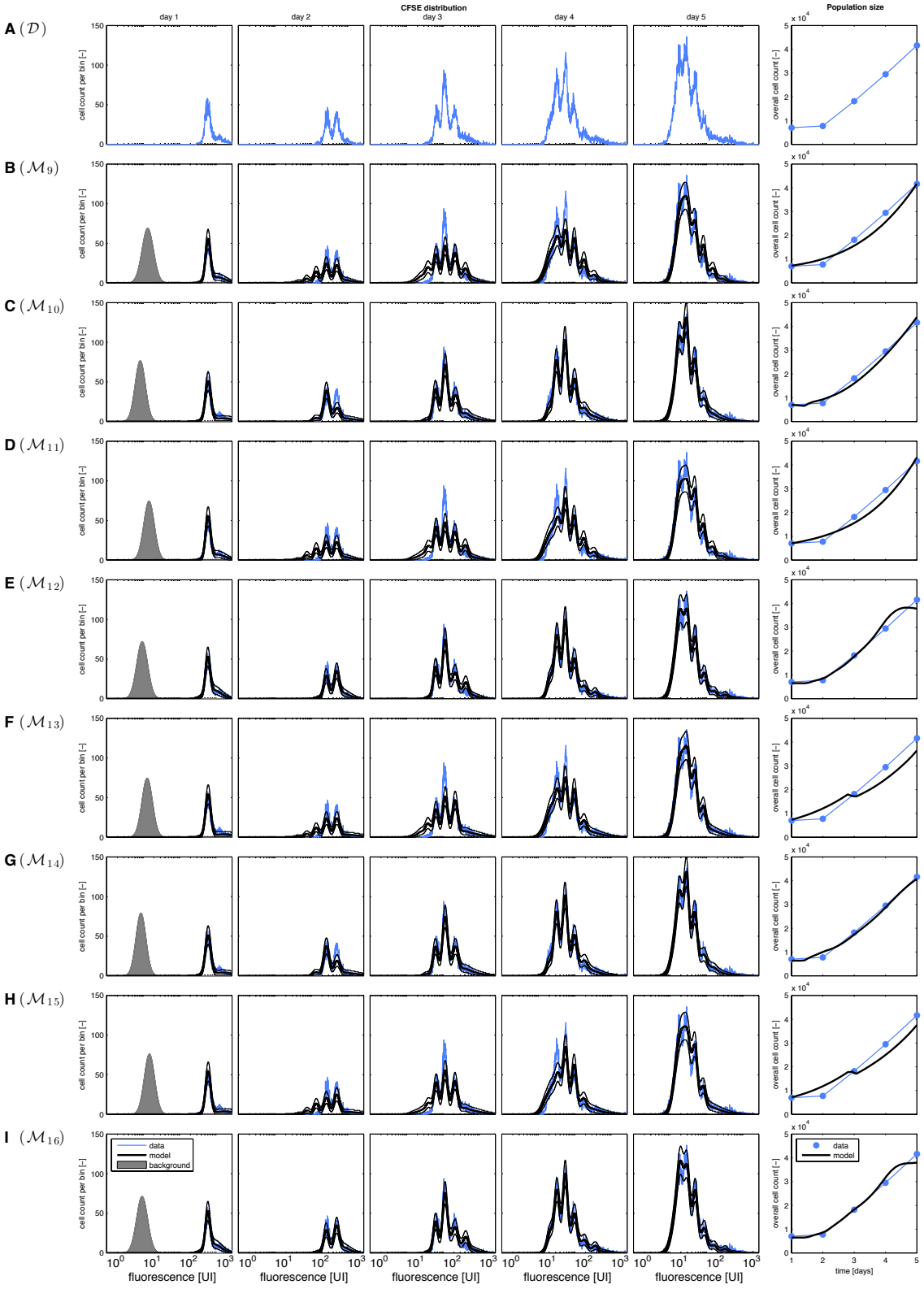
Supplement Figure 2: Performance of different optimisation methods for (**A**, **B**, **C**) model $\mathcal{M}_1$ and (**D**, **E**, **F**) model $\mathcal{M}_{16}$. (**A**, **D**) Negative log-likelihoods for 250 runs of deterministic local and global optimisers and 250 randomly samples parameter values. Missing points indicate failed objective function evaluations and optimisation runs. The dashed line indicates the significance threshold for converged starts. (**B**, **E**) Box plot of the computation time per optimisation run. (**C**, **F**) Average computation time per converged start. In (**F**) the average computation time per converged start of particle swarm optimisation is missing as not a single start converged.



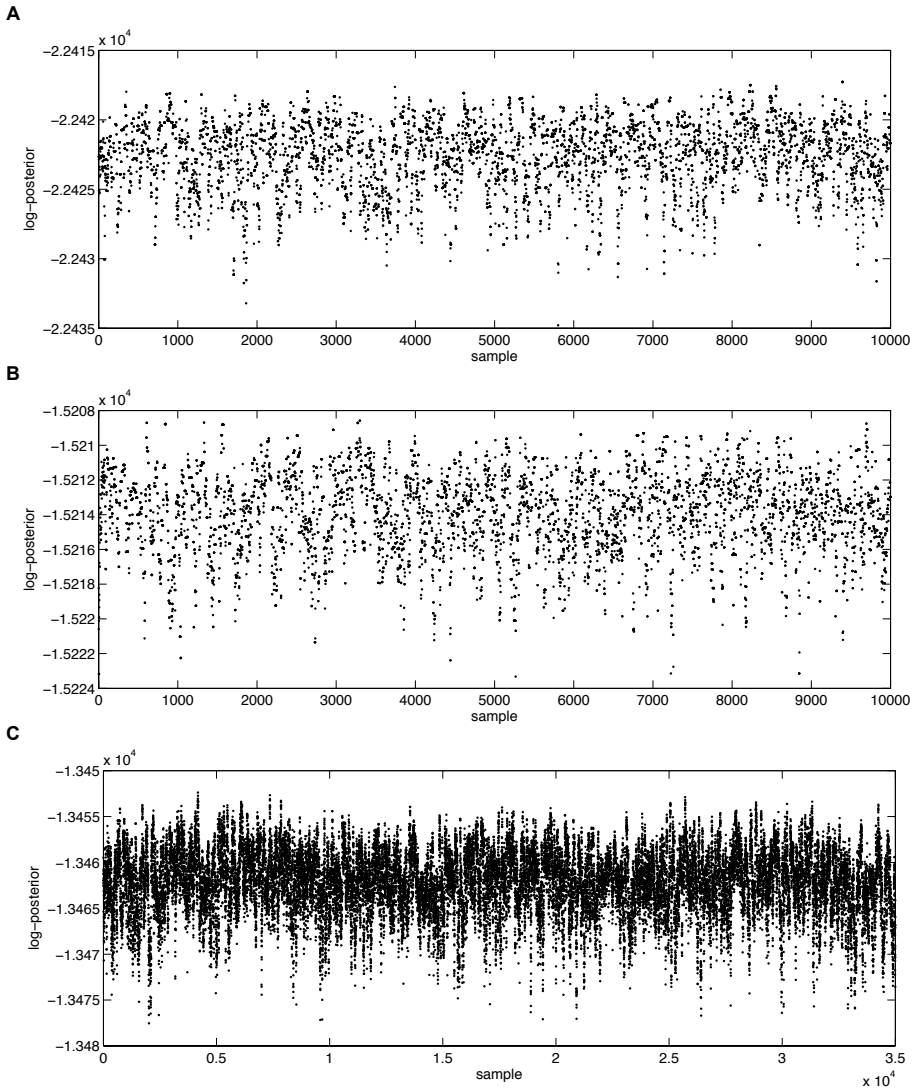
Supplement Figure 3: Result of multi-start local optimisation with forward sensitivity equation models $\mathcal{M}_1$ to $\mathcal{M}_{16}$. The individual subplots depict the negative log-likelihood values achieved after optimising from 250 different starting points, sorted in decreasing order, in comparison to the best result achieved using model $\mathcal{M}_{16}$. The local optimisation identified for most models suboptimal solutions (plateaus in the plots), as well as other stopping points. For all models the optimum seems to have been found several times.



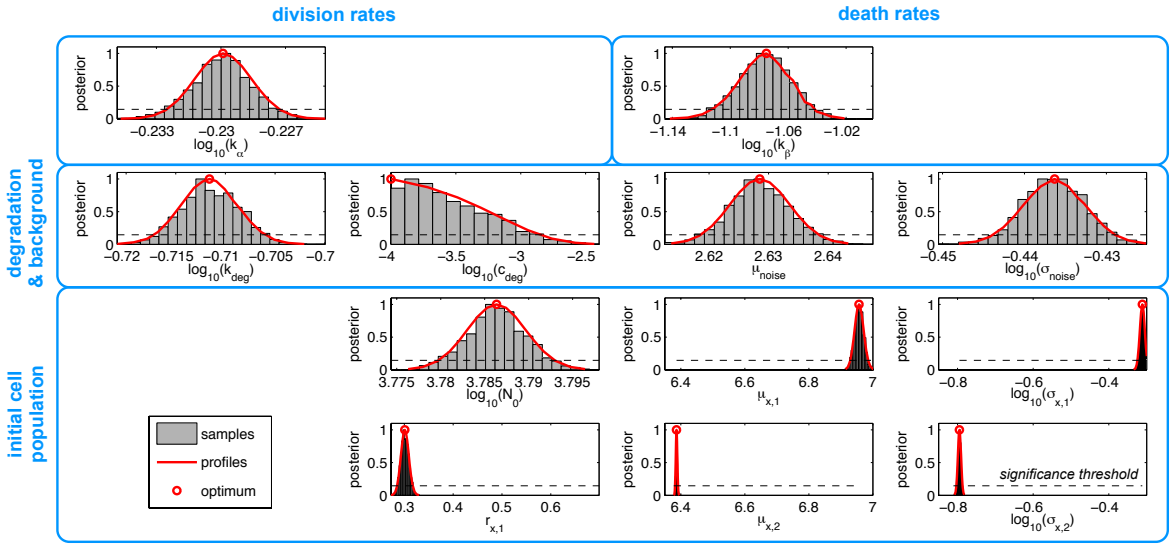
Supplement Figure 4: Comparison of (**A**, left) measured CFSE distributions and (**A**, right) measured overall cell counts with the best fits of model (**B**) $\mathcal{M}_1$, (**C**) $\mathcal{M}_2$, (**D**) $\mathcal{M}_3$, (**E**) $\mathcal{M}_4$, (**F**) $\mathcal{M}_5$, (**G**) $\mathcal{M}_6$, (**H**) $\mathcal{M}_7$ and (**I**) $\mathcal{M}_8$. The region in between the fine black lines (=) indicates the 90% confidence interval (5-th to 95-th percentile) of the bin counts for the particular number of measured cells for the optimal parameters. For a plausible model most data should be contained in the confidence intervals.



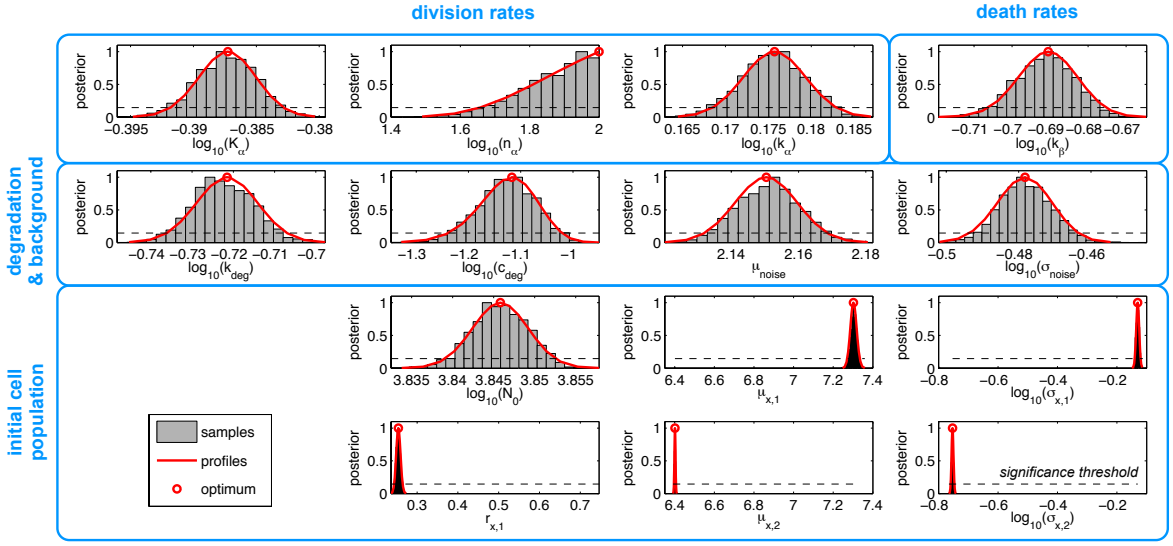
Supplement Figure 5: Comparison of (**A**, left) measured CFSE distributions and (**A**, right) measured overall cell counts with the best fits of model (**B**) $\mathcal{M}_9$, (**C**) $\mathcal{M}_{10}$, (**D**) $\mathcal{M}_{11}$, (**E**) $\mathcal{M}_{12}$, (**F**) $\mathcal{M}_{13}$, (**G**) $\mathcal{M}_{14}$, (**H**) $\mathcal{M}_{15}$ and (**I**) $\mathcal{M}_{16}$. The region in between the fine black lines (=) indicates the 90% confidence interval (5-th to 95-th percentile) of the bin counts for the particular number of measured cells for the optimal parameters. For a plausible model most data should be contained in the confidence intervals.



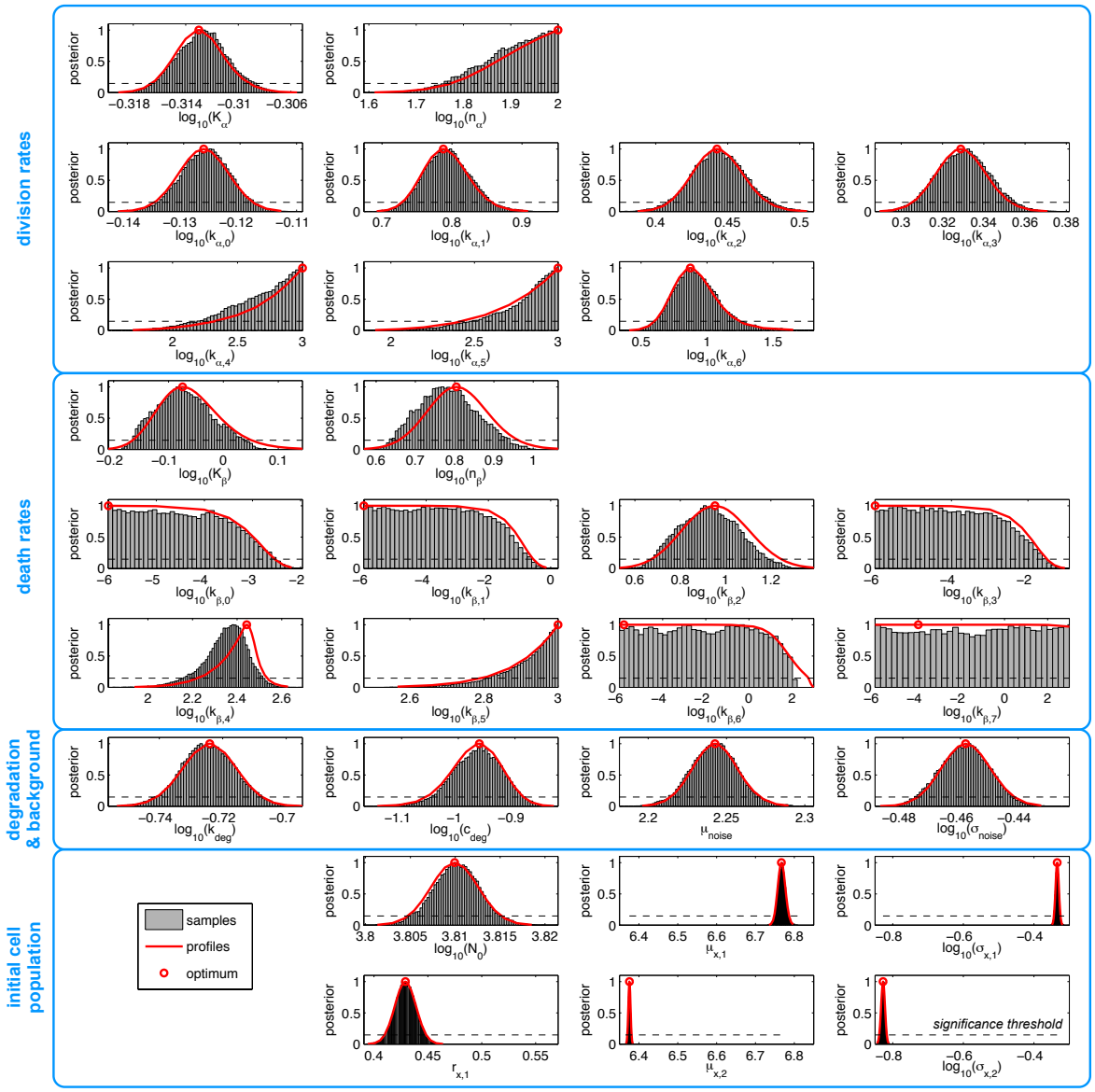
Supplement Figure 6: Diagnostic plot for convergence of MCMC sampling for (**A**) $\mathcal{M}_1$, (**B**) $\mathcal{M}_2$ and (**C**) $\mathcal{M}_{16}$. A thin sample containing every tenth member of the original sample is depicted.



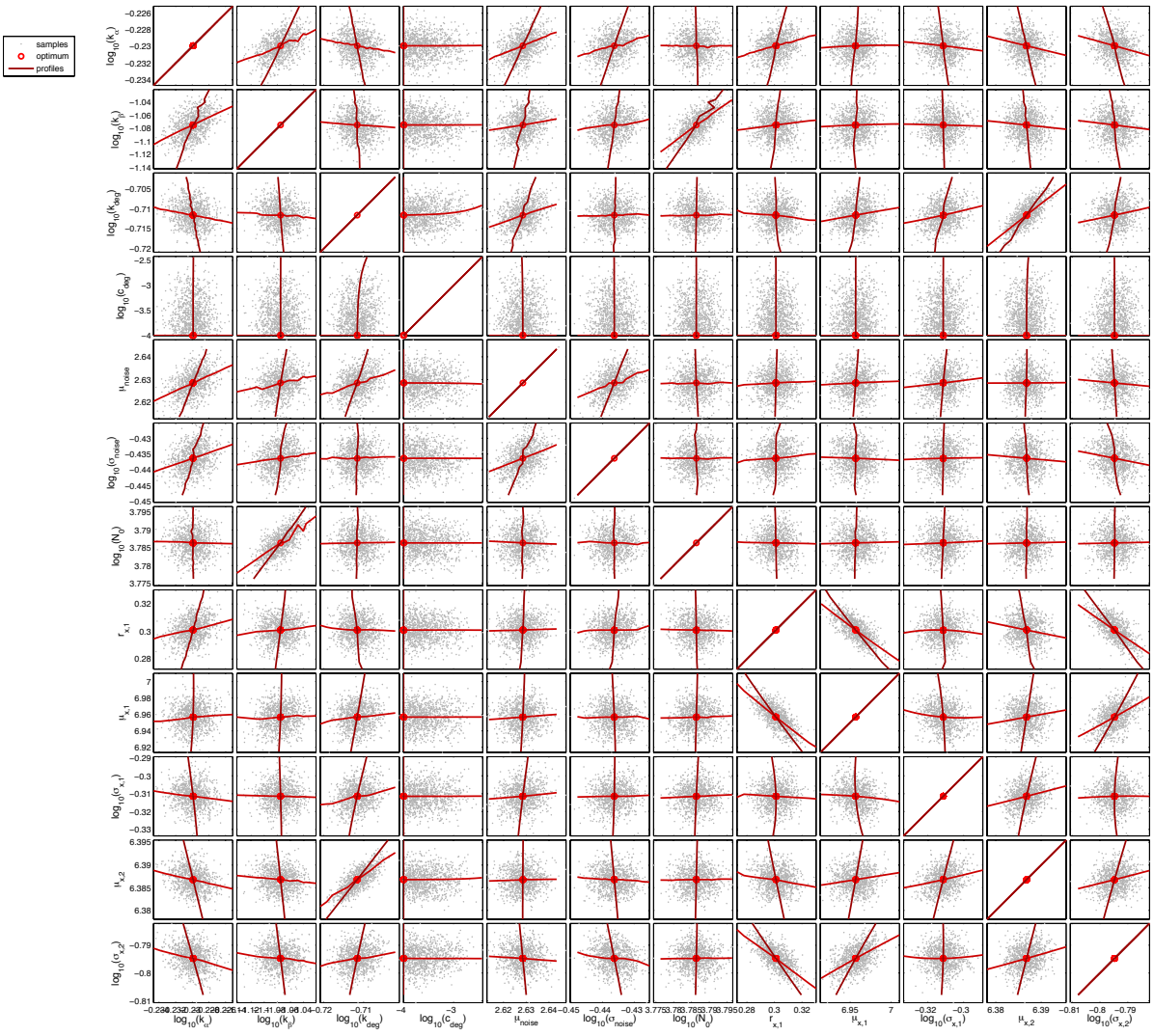
Supplement Figure 7: Parameter uncertainties for model $\mathcal{M}_1$. The maximum likelihood estimate (○), the profile likelihoods (—) and the histograms of the MCMC samples (■) are indicated. To handle the symmetry of $(r_{x,1}, \mu_{x,1}, \sigma_{x,1})$ and $(r_{x,2} = 1 - r_{x,1}, \mu_{x,2}, \sigma_{x,2})$, the estimation results are transformed such that $r_{x,1} < 0.5$ for all parameter vectors.



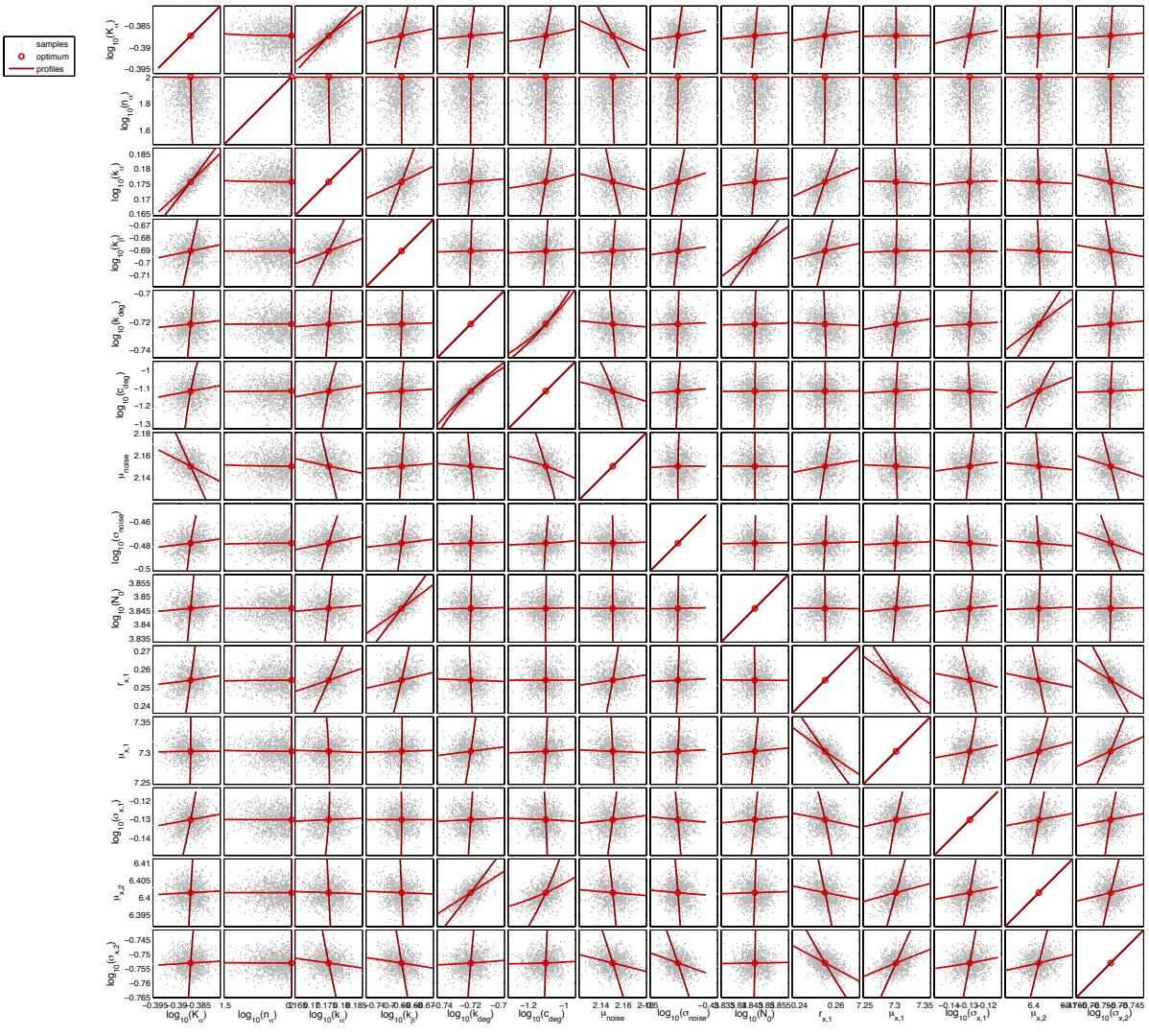
Supplement Figure 8: Parameter uncertainties for model $\mathcal{M}_2$. The maximum likelihood estimate (○), the profile likelihoods (—) and the histograms of the MCMC samples (■) are indicated. To handle the symmetry of $(r_{x,1}, \mu_{x,1}, \sigma_{x,1})$ and $(r_{x,2} = 1 - r_{x,1}, \mu_{x,2}, \sigma_{x,2})$, the estimation results are transformed such that $r_{x,1} < 0.5$ for all parameter vectors.



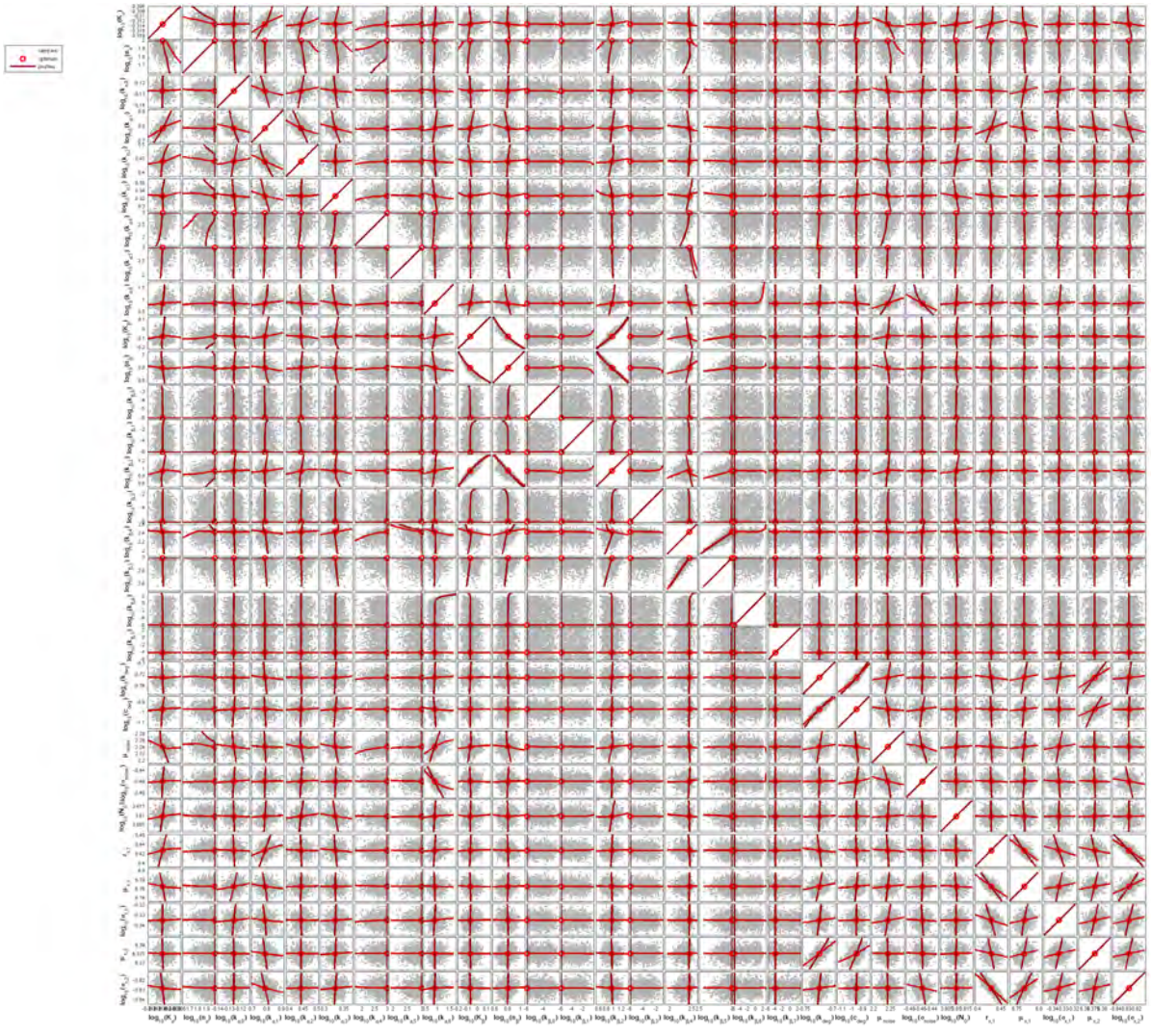
Supplement Figure 9: Parameter uncertainties for model $\mathcal{M}_{16}$. The maximum likelihood estimate (○), the profile likelihoods (—) and the histograms of the MCMC samples (■) are indicated. To handle the symmetry of $(r_{x,1}, \mu_{x,1}, \sigma_{x,1})$ and $(r_{x,2} = 1 - r_{x,1}, \mu_{x,2}, \sigma_{x,2})$, the estimation results are transformed such that $r_{x,1} < 0.5$ for all parameter vectors.



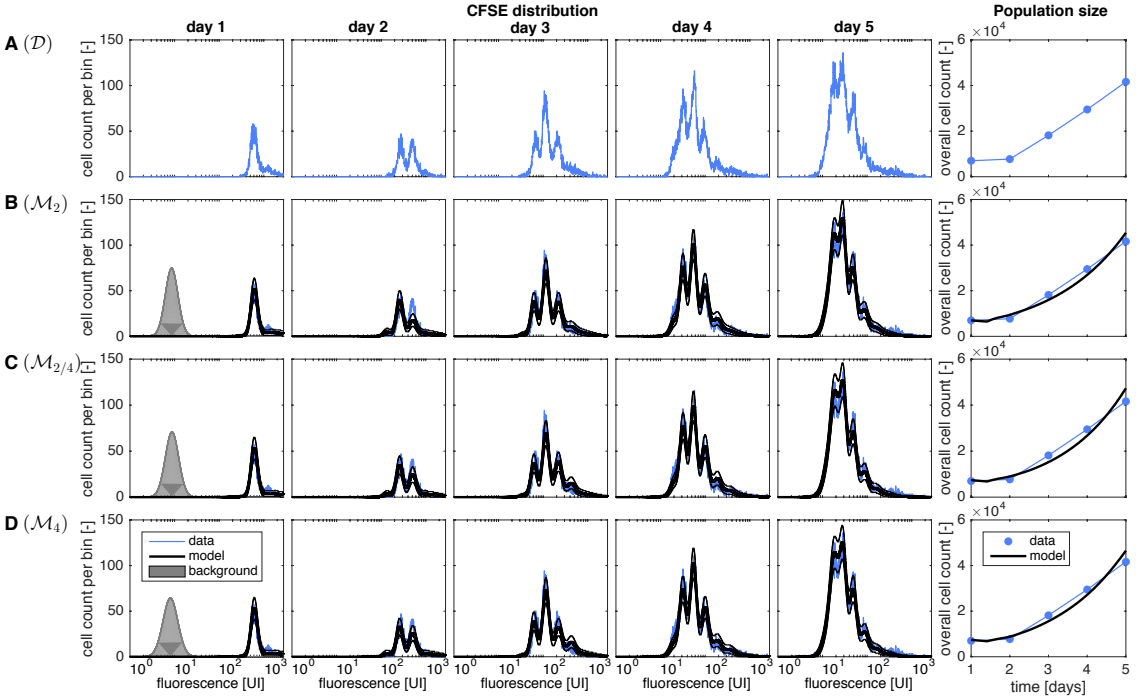
Supplement Figure 10: Correlation of parameter estimates for model $\mathcal{M}_1$. The scatterplot matrix for the MCMC samples ($\cdot$) are depicted. The maximum likelihood estimate ($\circ$) and the profile likelihoods with respect to the parameter in the x-axis (light red line, —) and y-axis (dark red line, —) are indicated. To handle the symmetry of $(r_{x,1}, \mu_{x,1}, \sigma_{x,1})$ and $(r_{x,2} = 1 - r_{x,1}, \mu_{x,2}, \sigma_{x,2})$, the estimation results are transformed such that $r_{x,1} < 0.5$ for all parameter vectors.



Supplement Figure 11: Correlation of parameter estimates for model $\mathcal{M}_2$. The scatterplot matrix for the MCMC samples ($\cdot$) are depicted. The maximum likelihood estimate ($\circ$) and the profile likelihoods with respect to the parameter in the x-axis (light red line, —) and y-axis (dark red line, —) are indicated. To handle the symmetry of $(r_{x,1}, \mu_{x,1}, \sigma_{x,1})$ and $(r_{x,2} = 1 - r_{x,1}, \mu_{x,2}, \sigma_{x,2})$, the estimation results are transformed such that $r_{x,1} < 0.5$ for all parameter vectors.



Supplement Figure 12: Correlation of parameter estimates for model $\mathcal{M}_{16}$. The scatterplot matrix for the MCMC samples ($\cdot$) are depicted. The maximum likelihood estimate ($\circ$) and the profile likelihoods with respect to the parameter in the x-axis (light red line, —) and y-axis (dark red line, —) are indicated. To handle the symmetry of $(r_{x,1}, \mu_{x,1}, \sigma_{x,1})$ and $(r_{x,2} = 1 - r_{x,1}, \mu_{x,2}, \sigma_{x,2})$, the estimation results are transformed such that $r_{x,1} < 0.5$ for all parameter vectors.



Supplement Figure 13: Comparison of (**A**, left) measured CFSE distributions and (**A**, right) measured overall cell counts with the best fits of model (**B**) $\mathcal{M}_2$, (**C**) $\mathcal{M}_{2/4}$, and (**D**) $\mathcal{M}_4$. The region in between the fine black lines (═) indicates the 90% confidence interval (5-th to 95-th percentile) of the bin counts for the particular number of measured cells for the optimal parameters. For a plausible model most data should be contained in the confidence intervals.