International Journal of Pure and Applied Mathematics
Volume 77 No. 1 2012, 119-147
ISSN: 1311-8080 (printed version)
url: http://www.ijpam.eu

ijpam.eu

A DIVISION-DEPENDENT COMPARTMENTAL MODEL WITH CYTON AND INTRACELLULAR LABEL DYNAMICS

H.T. Banks[1] §, W. Clayton Thompson[2]

[1]Center for Research in Scientific Computation
Center for Quantitative Sciences in Biomedicine
N.C. State University
Raleigh, NC, USA

[2]ICREA Infection Biology Lab
Department of Experimental and Health Sciences
Universitat Pompeu Fabra
Barcelona, SPAIN

Abstract: Intracellular dyes such as carboxyfluorescein succinimidyl ester (CFSE) have become an important tool for the flow cytometric analysis of a proliferating population of cells. The quantitative analysis of flow cytometry data involves two components. First, the generation structure (cell counts in terms of the number of divisions undergone) of the population must be related to a description of the manner in which cells divide and die. To this end, the cyton model [11, 20, 21], which relates generation structure directly to probability distributions over times to divide and die, is both experimentally accurate and intuitive. Second, the manner in which the intracellular dye is processed by growing, dividing, and dying cells must be related to the measured fluorescence profiles of the population of cells. Several models of label-structured populations [4, 5, 7, 19, 24, 25, 36] have been used to accurately describe these fluorescence profiles. Here, we rely upon the compartmental formulations [7, 19, 36, 39] along with the recent computational framework of [19, 36] to combine these two components to derive a new single model. This new model is capable of simultaneously describing generation dynamics and label structure dynamics in a proliferating population of cells in an extremely

Received: May 8, 2012

§Correspondence author

fast computational model. The new model is shown to be at least as accurate as those of previous efforts, while also providing a more biologically intuitive analysis of the behavior of the population of cells.

Key Words: cell proliferation, cell division number, CFSE, label structured population dynamics, cyton models, partial differential equations, inverse problems

1. Introduction

While mathematical descriptions of cell division dynamics have long been considered [9, 38] (see also the review [6]), such models have been hard to validate because of the difficulty in obtaining suitable experimental data. However, in the past two decades, the use of intracellular dyes such as carboxyfluorescein succinimidyl ester (CFSE) in proliferation assays has been shown to provide an indication of the generation structure (cells per division number) of a population of cells. While generation structure in itself is insufficient for the inference of biologically meaningful parameters such as rates of proliferation and death, assay measurements taken serially in time can be combined with a mathematical model of cell division to provide insights into the dynamic parameters of a cell population. Thus there is significant interest in mathematical models which accurately *describe CFSE-based proliferation assay data, can be solved and fit to data quickly, and provide estimates of biologically relevant information.*

The mathematical description of CFSE data has generally followed one of two distinct approaches. The first approach [14, 15, 16, 17, 18, 22, 23] is to use data taken from proliferation assays to approximate the numbers of cells in terms of the numbers of divisions undergone. Mathematical models are then formulated which attempt to describe these numbers in terms of division and death dynamics. The advantage of such models is that they are generally simple and relate intuitive values (cells per division number) to biologically relevant parameters (division and death rates). In particular, the cyton model [20, 21], in which cell dynamics are characterized by independent probability distributions over times to division and death, has proven particularly successful in describing cellular dynamics [6, 11, 20]. However, the validation of such models requires careful analysis and deconvolution of CFSE data to determine the generation structure of the population of cells. This generally introduces some error and/or bias into the resulting cell number data, particularly if division peaks are poorly resolved or for high division numbers where there is significant overlap between adjacent peaks [6, 39]. The second approach is to describe the evolution of

the CFSE dynamics (intracellular turnover, partitioning upon mitosis) directly [4, 5, 24, 25]. The benefit of such an approach is that the CFSE histograms are described directly by a conservation law, without any need for deconvolution of the data. However, models of label dynamics can be quite complex and the parameters of such models are generally more difficult to relate to biologically relevant information [4, 7, 19].

Recent work [7, 19, 36, 39] has been directed toward combining these two approaches so that label dynamics are described directly while the generation structure of the population is also computed and related to rates of proliferation and death. In fact, it is possible to view such models, structured by both (discrete) division number and (continuous) label fluorescence, as a unifying framework relating models of generation structure to models of label dynamics [6, 19]. Here, we take advantage of the compartmental formulations [7, 19, 36, 39] along with the computational framework developed in [19, 36] to propose a new mathematical model which accounts for both the *turnover and dilution of intracellular dye* (via a conservation law) as well as the *evolution of the population generation structure* (via a generalization of the cyton model). This new model is then compared to previous work [7, 39] in terms of computational performance and accuracy in describing an actual CFSE data set. As will be shown, the new model is at least as good as previous models describing CFSE histogram data for the data set shown. The computational approach of [19, 36] permits the model to be solved much faster (by orders of magnitude) than in previous work [4, 6, 7, 39] and the incorporation of the cyton model [20, 21] provides biologically intuitive estimates of the population distribution of times to division and death.

We begin with a brief summary of mathematically relevant aspects of the data collection process. Next, the cyton model for cell dynamics is reviewed, as is a model of label dynamics. These are then combined to form a new mathematical approach to describing flow cytometry data. Finally, an inverse problem is formulated and the feasibility of the new modeling framework is demonstrated by fitting it to experimental data for a population of CD4+ cells labeled with CFSE and stimulated with PHA.

2. CFSE Data

CFSE is an intracellular dye which can be used to track division history in a population of lymphocytes [27, 29, 33, 35, 41] without any adverse affects on the functioning of the cells under study [26, 28, 29]. After being introduced to

a culture of cells, CFSE diffuses freely across the cell membrane and becomes bound to intracellular proteins [33]. The initial uptake of CFSE is mildly dependent on cell size and the distribution of intracellular proteins [30, 35] but is generally uniform across all cells in a population [29, 40], particularly when cells of a single type (e.g., CD4+ cells) are being studied. The fluorescence properties of the fluorescein groups permit the quantity of CFSE within a single cell to be assessed by a flow cytometer, which measures the fluorescence intensity (FI) resulting from intracellular CFSE, a quantity which is known to vary linearly with mass [27]. While individual cells are measured by the flow cytometer, it is not possible to follow a single cell (or its progeny) across multiple measurement times. Rather, individual fluorescence measurements are aggregated to indicate the distribution of FI (and hence, CFSE) within the sampled population. Such data is most commonly represented in histogram form, indicating the CFSE profile of the population at the measurement time (see Figure 1).

As a result of the uniform initial update of CFSE, the initial distribution of CFSE in the population of cells under study is unimodal. When a cell divides, the intracellular carboxyfluorescein groups (and the proteins to which they are bound) are partitioned approximately evenly between the resulting daughter cells. Thus the measured FI of a given daughter cell is less than the measured FI of a mother cell by a predictable amount. As a result, measured FI can be used to estimate the number of divisions a cell has undergone [27, 29, 33, 35, 40, 42]. Each generation of cells (understood to be those cells which have undergone the same number of divisions since initial labeling with CFSE) appears as a distinct mode or peak in the histogram data. As cells divide, subsequent generations of cells become hard to distinguish as the natural autofluorescence of cells becomes an important contributor to the FI measurement [4, 7, 27, 35].

After cells are labeled with CFSE, those cells are then stimulated to divide either by exposure to a specific antigen or nonspecifically by exposure to phytohaemagglutinin (PHA). The time at which the stimulating agent is introduced into the cell culture is considered to be $t = 0$ hours, and all measurement times are given relative to this point. The stimulated cells are placed into separate wells (each to be analyzed, e.g., counted, at a different time) containing a nutrient medium; separate wells for each measured sample are necessary to prevent the disruption of cellular function by the measurement process. At each measurement time, cells from a single well are harvested, stained and/or analyzed in accordance with a particular experimental protocol, and then analyzed by flow cytometry. There are numerous experimental protocols available for CFSE-based proliferation assays [27, 29, 35, 40, 42] and a detailed review of the mathematically relevant aspects of the data collection process can be found

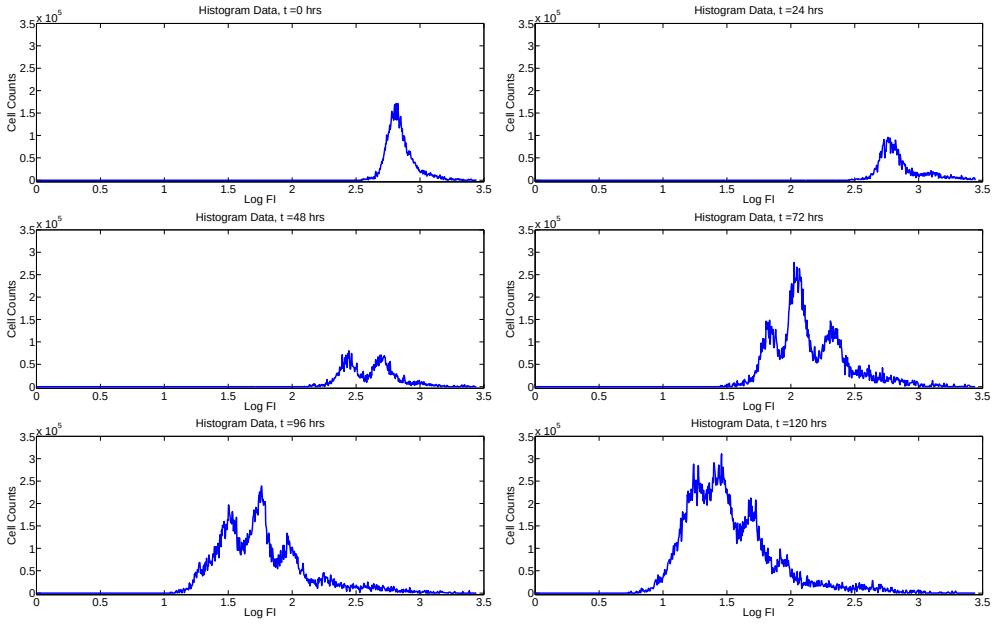

Figure 1: Cell count vs. log fluorescence intensity; originally from [25].

in [6, 39]. It will be assumed here that CFSE histogram data, such as that shown in Figure 1, represents a census of all cells in the population. Such an assumption has implications for the statistical model which describes errors in the data collection process (see [39, Ch. 4] and Sections 4 and 5 below) but is sufficiently accurate for the present purposes.

Fluorescence intensity data from a flow cytometer is most conveniently represented on a logarithmic scale. Let z represent the base-10 logarithm of fluorescence intensity (in arbitrary units of intensity, UI) and consider some partition $\{z_k^j\}$ $(1 \leq k \leq K(j))$ of the z-axis at measurement time t_j. CFSE histogram data describes the number of cells n_k^j having log-fluorescence intensity in the interval $[z_k^j, z_{k+1}^j)$ at time t_j. An example data set (originally from [25]) is shown in Figure 1. As stated above, each 'peak' in the data represents a cohort of cells having undergone the same number of divisions since initial CFSE uptake. For models of label dynamics (see Section 3.2) this data can be compared directly to the mathematical model. For models which consider only generation structure, additional effort is necessary to determine the numbers of cells having undergone a certain number of divisions at each time. Typically, this involves fitting the histogram data with a series of Gaussian-type peaks and then using

these peaks to compute cell numbers (see, e.g., [6] and the references therein). In this report, we focus exclusively on models which can be fit directly to CFSE histograms.

3. Mathematical Model

As mentioned above, our new mathematical model is built upon two existing frameworks which we summarize here. First, the generation structure (cells per division number) of the population is described by the cyton model [20, 21]. We remark that the cyton model is not the only possible model of the evolution of generation structure in a population of dividing cells. For a recent review of alternative models see [6, 32]. However, the cyton model has been shown to perform better than (or at least comparably to) most other models in several practical applications [6, 20, 32]. Additionally, the quantitative modeling of cell division and death is much more intuitive in the cyton model. We review the cyton model and then in the second framework we summarize a hyperbolic conservation law model which has been shown to accurately describe the natural turnover of CFSE and the dilution of the dye by mitosis [4, 5, 7, 24, 25, 36]. Finally, these two approaches are combined using the computational framework of [19, 36] to form a new model that directly describes flow cytometry data.

3.1. The Cyton Model

In simplest terms, the cyton model arises from the simple observation that any given cell must eventually divide or die. Based upon experimental evidence [20] it has been proposed that the processes of cell division and death operate independently of one another. Thus, a given cell can be considered to have (fixed) times to divide and die, with the eventual fate of the cell determined by whichever of the two values is the smallest. The 'cyton' refers to the underlying regulatory machinery of the cell which governs this hypothetical process [11, 20, 21]. (Some work, e.g., [32], considers the additional possibility that a subset of cells may neither divide nor die over the duration of an experiment. This can ultimately be viewed as a special case of the cyton in which the times to division and death for a given cell exceed the time remaining in an experiment.) In a population of cells, it is reasonably hypothesized that times to division and death may be different for every cell in the population, but that these times are drawn independently from probability distributions which are common to all cells having undergone the same number of divisions. (The experimental

evidence for such an assumption is considered in detail in [20, 21].) Thus, it is these probability distributions which define the dynamic behavior of the population, and the cyton model establishes a simple, intuitive relationship between generation structure and these probability distributions in a population of dividing cells.

Let $N_i(t)$, $0 \leq i \leq i_{\max}$ represent the number of cells having undergone i divisions at time t. Assume there are N_0 cells in the population at $t = 0$. In its simplest form, the cyton model relates the number of cells in the population to the number of cells which divide and die in a unit of time,

$$\begin{aligned} N_0(t) &= N_0 - \int_0^t \left(n_0^{div}(s) - n_0^{die}(s) \right) ds \\ N_i(t) &= \int_0^t \left(2n_{i-1}^{div}(s) - n_i^{div}(s) - n_i^{die}(s) \right) ds, \end{aligned} \tag{1}$$

where $n_i^{div}(t)$ and $n_i^{die}(t)$ indicate the numbers per unit time of cells having undergone i divisions that divide and die, respectively, at time t. Let $\phi_0(t)$ and $\psi_0(t)$ be probability density rate functions (in units 1/time) for the times to division and death, respectively, for an undivided cell. Let F_0, the initial precursor fraction, be the fraction of undivided cells which would hypothetically divide in the absence of any cell death. It follows that

$$\begin{aligned} n_0^{div}(t) &= F_0 N_0 \left(1 - \int_0^t \psi_0(s) ds \right) \phi_0(t) \\ n_0^{die}(t) &= N_0 \left(1 - F_0 \int_0^t \phi_0(s) ds \right) \psi_0(t). \end{aligned} \tag{2}$$

Similarly, one can define probability density rate functions $\phi_i(t)$ and $\psi_i(t)$ for times to division and death, respectively, for cells having undergone i divisions, as well as the progressor fractions F_i of cells which would complete the i^{th} division in the absence of cell death. Then the numbers per unit time of dividing and dying cells are computed as

$$\begin{aligned} n_i^{div}(t) &= 2F_i \int_0^t n_{i-1}^{div}(s) \left(1 - \int_0^{t-s} \psi_i(\xi) d\xi \right) \phi_i(t-s) ds \\ n_i^{die}(t) &= 2 \int_0^t n_{i-1}^{div}(s) \left(1 - F_i \int_0^{t-s} \phi_i(\xi) d\xi \right) \psi_i(t-s) ds. \end{aligned} \tag{3}$$

It follows from the forms of equations (1) - (3) that the evolution of the generation structure for a population of cells described by the cyton model is

entirely determined by the density functions $\{\phi_i(t)\}_{i=0}^{i_{\max}}$ and $\{\psi_i(t)\}_{i=0}^{i_{\max}}$ and the progressor fractions $\{F_i\}_{i=0}^{i_{\max}}$. Ultimately then, the cyton model is not a single model but rather a family of models which can be identified by the underlying forms and assumptions on the functions $\phi_i(t)$ and $\psi_i(t)$ and the parameters F_i. A detailed motivation of the cyton model, as well as experimental evidence for the form of the model, can be found in [20, 21].

Following the experimental evidence initially described in support of the cyton model [20, 21], it is most common to assume that the functions $\phi_i(t)$ and $\psi_i(t)$ are lognormal probability density functions. Define the (lognormally distributed) random variable T_i^{div} to be the time required for a cell to complete the i^{th} division (with the clock starting from the completion of the $(i-1)^{th}$ division), and let the parameters μ_i^{div} and σ_i^{div} represent the mean and standard deviation of the natural logarithm of this random variable. The analogous definitions hold for T_i^{die}, μ_i^{die}, and σ_i^{die}. It follows that the distributions $\phi_i(t)$ and $\psi_i(t)$ are given by

$$\begin{aligned}\phi_i(t) &= \frac{1}{t\sigma_i^{div}\sqrt{2\pi}}\exp\left(-\frac{(\log t-\mu_i^{div})^2}{2(\sigma_i^{div})^2}\right)\\ \psi_i(t) &= \frac{1}{t\sigma_i^{die}\sqrt{2\pi}}\exp\left(-\frac{(\log t-\mu_i^{die})^2}{2(\sigma_i^{die})^2}\right),\end{aligned} \tag{4}$$

for each i, $0 \le i \le i_{\max}$, and $t > 0$. Alternatively, it may be more intuitive to characterize the density functions $\phi_i(t)$ and $\psi_i(t)$ directly in terms of the means and standard deviations of T_i^{div} and T_i^{die} (as opposed to the means and standard deviations of $\log(T_i^{div})$ and $\log(T_i^{die})$). These quantities are related by the formulas

$$\begin{aligned}E[T_i^{div}] &= \exp\left(\mu_i^{div}+\frac{(\sigma_i^{div})^2}{2}\right)\\ E[T_i^{die}] &= \exp\left(\mu_i^{die}+\frac{(\sigma_i^{die})^2}{2}\right)\\ Var[T_i^{div}] &= \left(\exp\left((\sigma_i^{div})^2\right)-1\right)\exp\left(2\mu_i^{div}+(\sigma_i^{div})^2\right)\\ Var[T_i^{die}] &= \left(\exp\left((\sigma_i^{die})^2\right)-1\right)\exp\left(2\mu_i^{die}+(\sigma_i^{die})^2\right).\end{aligned}$$

For the most basic cyton model, it is assumed that the random variables T_i^{div} are identically distributed for all $i \ge 1$ and that T_i^{die} are identically distributed for all $i \ge 1$. These distributions need not be identical to each other, and may be different from the corresponding random variables for undivided cells

$(i = 0)$. Thus we have

$$\mu_i^{div} = \mu^{div}, \quad i \geq 1 \tag{5}$$
$$\sigma_i^{div} = \sigma^{div}, \quad i \geq 1 \tag{6}$$
$$\mu_i^{die} = \mu^{die}, \quad i \geq 1 \tag{7}$$
$$\sigma_i^{die} = \sigma^{die}, \quad i \geq 1. \tag{8}$$

It is also assumed that all cells which are activated to divide will continue to do so (in the absence of cell death),

$$F_i = 1, \quad i \geq 1. \tag{9}$$

Thus under these assumptions the most basic cyton model we consider has 9 parameters

$$\{\mu_0^{div}, \sigma_0^{div}, \mu_0^{die}, \sigma_0^{die}, \mu^{div}, \sigma^{div}, \mu^{die}, \sigma^{die}, F_0\}.$$

Given these parameters, the probabilities of division and death (4) can be computed for all time, which can then be used to determine the rates (2)-(3) for the numbers of cells which divide and die and then the total numbers of cells in each generation (1).

Of course, many of the assumptions made in obtaining this simplified model might be relaxed to obtain a more generalized cyton model. For instance, it is reasonable to assume that non-precursors (those cells which are not activated to divide) may die at different rates than cells which are activated to divide (but which die before dividing). For instance, it might be assumed that cells which will not progress through the first division may die with some exponential rate, β. In such a case, it follows that

$$\psi_0(t) = \frac{F_0}{t\sigma_0^{die}\sqrt{2\pi}} \exp\left(-\frac{(\log t - \mu_0^{die})^2}{2(\sigma_0^{die})^2}\right) + (1 - F_0)\beta e^{-\beta t}. \tag{10}$$

An additional possibility is that a fraction of the cells which do not progress through the first division will remain dormant throughout the experiment, without dying according to either a lognormal or exponential probability. Let p_{idle} represent the fraction of non-progressing cells which remain idle through the course of the experiment. If non-progressing, non-idle cells die with the same probability as progressing cells, then the new probability of death is

$$\psi_0(t) = \frac{1 - p_{idle}(1 - F_0)}{t\sigma_0^{die}\sqrt{2\pi}} \exp\left(-\frac{(\log t - \mu_0^{die})^2}{2(\sigma_0^{die})^2}\right). \tag{11}$$

If non-progressing, non-idle cells are assumed to die at an exponential rate, then the new probability of death is

$$\psi_0(t) = \frac{F_0}{t\sigma_0^{die}\sqrt{2\pi}}\exp\left(-\frac{(\log t - \mu_0^{die})^2}{2(\sigma_0^{die})^2}\right) + (1-p_{idle})(1-F_0)\beta e^{-\beta t}. \quad (12)$$

As before for the basic cyton model, the function $\psi_0(t)$ (regardless of form) is used to compute the rates of numbers of dying and dividing cells (2) which are then used to compute the numbers of cells per generation (1).

Another obvious generalization is to suppose that F_i (for $i \geq 1$) may not be equal to 1. In other words, at each generation there is some fraction of cells which will not progress any further. (These cells may still die according to the lognormal probability of death for that particular generation.) Following [20], the fractions F_i can be most easily defined in terms of a *division destiny*. Among the cells which will progress through at least one division (F_0N_0 of them; these cells are the *precursors* from which all subsequent cells in the population will arise). Let p_i be the probability that a cell ceases to divide after completing i divisions. Obviously,

$$\sum_{i=1}^{i_{\max}} p_i = 1.$$

Define the cumulative probabilities

$$c_i = \sum_{j=1}^{i} p_j.$$

It follows that the progressor fractions (for $i \geq 1$) are

$$F_i = \begin{cases} \frac{1-c_i}{1-c_{i-1}}, & c_{i-1} < 1 \\ 0, & c_{i-1} = 0. \end{cases} \quad (13)$$

We can now define the division destiny, d_i, to be the fraction of cells (out of those cells in the original population) which would have proceeded through exactly i divisions in the absence of any cell death. Given the definitions above, the division destinies are computed as

$$d_i = \begin{cases} 1-F_0, & i = 0 \\ F_0 p_i, & i \geq 1. \end{cases} \quad (14)$$

One possibility is to allow probabilities p_i (or the progressor fractions F_i) to vary independently of one another (subject only to the constraint that the

probabilities must sum to one). An alternative approach considered in [20] is to assume a discrete normal form for the probabilities p_i. Hence the probabilities p_i as well as the progressor fractions F_i are entirely determined by two parameters–the mean and the variance of the discrete normal distribution. This assumption is convenient not only in reducing the total number of parameters in the underlying mathematical model, but has also been shown to accurately describe experimental data [20].

The generalizations of the cyton model considered in this paper are summarized in Table 1. Of course, numerous additional generalizations are possible. For instance, cells which do not proceed through the first generation may possibly be described by a separate lognormal distribution rather than by an exponential distribution. If one considers non-progressors at every generation (according to Equation (13)), then one may define a separate probability of death for the non-progressors in each generation (say, exponential with rates β_i). In the interest of brevity, such a large number of possible models are not considered here. As will be shown below, the generalizations of the cyton model already proposed above are sufficient to accurately describe the data set shown in Figure 1.

3.2. The Label-Structured Model

While the cyton model provides a method to compute the generation structure for a population of dividing cells, it is not possible to directly measure the division number of a given cell. As discussed in Section 2, intracellular dyes such as CFSE can be used to provide some indication of division structure. However, some work is required before the CFSE histogram profile of a population can be used to infer the generation structure of the measured population. In order to accurately and efficiently determine this generation structure, it is necessary to account for the manner in which cells process the intracellular dye as they grow, divide, and die. Because flow cytometry data is typically presented as a histogram showing the numbers of cells having given values of fluorescence intensity, structured population models [31] using fluorescence intensity as the structure variable seem particularly suitable [25] to describe these dynamics. Because the FI measurement corresponds approximately to the mass of fluorescence label within the cell being measured, such models have been referred to as 'label structured models'.

It is known that cells naturally lose FI over time as a result of the natural turnover of the intracellular proteins to which CFSE is bound [27, 33]. Thus an accurate model of label dynamics must account for both this intracellular

Model	Descriptions	Equations	Parameters
Cyton1	Basic cyton model	(1)-(8),(9)	9
Cyton2	Cyton1 with exponential death for undivided non-progressors	(1)-(8),(9),(10)	10
Cyton3	Cyton1 with division destiny curve	(1)-(8),(13)	11
Cyton4	Cyton1 with exponential death for undivided non-progressors and division destiny curve	(1)-(8),(10),(13)	12
Cyton5	Cyton2 with an additional fraction of idle undivided non-progressors	(1)-(8),(9),(12)	11
Cyton6	Cyton3 with an additional fraction of idle undivided non-progressors	(1)-(8),(11),(13)	12
Cyton7	Cyton4 with an additional fraction of idle undivided non-progressors	(1)-(8),(12),(13),	13

Table 1: Summary of cyton model generalizations to be considered.

turnover of CFSE and the dilution of CFSE by division. Upon mitosis, it is assumed that the mass of CFSE within the mother cell is partitioned evenly among the new daughter cells. All cells are known to have a low level of autofluorescence which is measurable even for cells which have not been labeled with CFSE. Autofluorescence and fluorescence resulting from intracellular CFSE are assumed to be additive.

Let $n(t,x)$ be the structured density (cell per unit of fluorescence intensity) of a population of cells at time t and with measured FI x. Let $\alpha(t,x)$ represent the rate (in units 1/hr) at which cells with measured fluorescence intensity x at time t divide. Similarly, let $\beta(t,x)$ be the rate at which cells die. It should be noted that the dependence of the functions α and β on measured fluorescence intensity is not intended to be causative; because measured FI decreases approximately 2-fold with each division, the structure variable x can be used as a surrogate for the dependence of the proliferation and death rates on division number, which has been well-attested in other studies [4, 5, 7, 14,

15, 22, 23, 24, 25, 39]. Then the CFSE label dynamics can be described by the fragmentation equation

$$\begin{aligned}\frac{\partial n(t,x)}{\partial t} - ce^{-kt}\frac{\partial[(x-x_a)n(t,x)]}{\partial x} &= -(\alpha(t,x)+\beta(t,x))n(t,x)\chi_{[x_a,x^*]} \\ &\quad 4\alpha(t,2x-x_a)n(t,2x-x_a) \\ n(0,x) &= \Phi(x) \\ n(t,x_{\max}) &= 0 \\ v(t,x_a)n(t,x_a) &= 0. \end{aligned} \tag{15}$$

The advection term (with parameters c and k) represents the Gompertz decay process for decrease in FI resulting from intracellular turnover of CFSE. The parameter x_a describes the autofluorescence of activated cells. A careful derivation of this model can be found in [4, 39]. Given rates of proliferation, death, and FI decay, the structured density $n(t,x)$ provides cell counts in terms of measured FI at a given time, which can be compared directly to histogram data such as that shown in Figure 1. The model (15) has been shown to accurately describe label dynamics for CD4+ cells [4, 5, 24, 25, 39] and CD8+ cells [25].

While the model (15) provides an accurate description of the evolution of the CFSE histogram profile for a population of cells, it cannot be directly related to the generation structure of the population. Rather, the model can be reformulated with distinct compartments for each generation. The resulting model is a system of partial differential equations,

$$\begin{aligned}\frac{\partial n_0}{\partial t} - ce^{-kt}(x-x_a)\frac{\partial n_0}{\partial x} &= -(\alpha_0(t)+\beta_0(t)-ce^{-kt})n_0(t,x) \\ \frac{\partial n_1}{\partial t} - ce^{-kt}(x-x_a)\frac{\partial n_1}{\partial x} &= -(\alpha_1(t)+\beta_1(t)-ce^{-kt})n_1(t,x) + R_1(t,x) \\ &\vdots \\ \frac{\partial n_{i_{\max}}}{\partial t} - ce^{-kt}(x-x_a)\frac{\partial n_{i_{\max}}}{\partial x} &= -(\beta_{i_{\max}}(t)-ce^{-kt})n_{i_{\max}}(t,x) + R_{i_{\max}}(t,x), \end{aligned} \tag{16}$$

where now $n_i(t,x)$ is the structured population density for cells having undergone i divisions. The recruitment terms describe the symmetric division of CFSE upon mitosis and are given by $R_i(t,x) = 4\alpha_{i-1}(t)n_{i-1}(t,2x-x_a)$. Boundary and initial conditions are given as in (15). Observe that, because the number of divisions undergone has now been explicitly identified by the subscripted generation number, it is no longer necessary for the division and

death rates to depend upon the measured fluorescence intensity. This makes the compartmental model (16) both simpler and more intuitive [7, 36, 39]. As with the fragmentation model, the compartmental model has been shown to accurately describe CFSE histogram data for CD4+ cells [7, 39].

3.3. The New Model

There are two major problems with the using (16) to describe CFSE histogram data. First, computational solutions (when computed in a naive manner [7, 39]) can be quite expensive, requiring several minutes of computational time [19, 36, 39]. Second, it is not immediately clear how the time-dependent exponential rates $\alpha_i(t)$ and $\beta_i(t)$ might be related to biologically relevant quantities such as mean time to division.

Rather than considering a compartmental model (16) with measured fluorescence intensity as the structure variable, Allgöwer, et al., [19, 36] have proposed a model which is structured by the fluorescence intensity *resulting from the mass of CFSE within the cell.* Let $\tilde{x}$ represent this fluorescence intensity. Then the authors propose that the population density $\tilde{n}(t, \tilde{x})$ with respect to this quantity is described by the system of equations

$$\begin{aligned}
\frac{\partial \tilde{n}_0}{\partial t} - ce^{-kt}\frac{\partial[\tilde{x}\tilde{n}_0]}{\partial \tilde{x}} &= -(\alpha_0(t) + \beta_0(t))\tilde{n}_0(t, \tilde{x}) \\
\frac{\partial \tilde{n}_1}{\partial t} - ce^{-kt}\frac{\partial[\tilde{x}\tilde{n}_1]}{\partial \tilde{x}} &= -(\alpha_1(t) + \beta_1(t))\tilde{n}_1(t, \tilde{x}) + 4\alpha_{i-1}(t)\tilde{n}_{i-1}(t, 2\tilde{x}) \qquad (17) \\
&\vdots
\end{aligned}$$

with boundary and initial conditions as in (15) and (16). The major advantage of formulating the model in terms of $\tilde{x}$ is the very simple form of the model solution. It can be proven [19, 36] that the solution to the model (17) can be written as

$$\tilde{n}_i(t, \tilde{x}) = N_i(t)\bar{n}_i(t, \tilde{x}) \qquad (18)$$

for all i. In this representation the functions $N_i(t)$ satisfy the weakly coupled system of ordinary differential equations

$$\begin{aligned}
\frac{dN_0}{dt} &= -(\alpha_0(t) + \beta_0(t))N_0(t) \\
\frac{dN_1}{dt} &= -(\alpha_1(t) + \beta_1(t))N_1(t) + 2\alpha_{i-1}(t)N_{i-1}(t) \qquad (19) \\
&\vdots
\end{aligned}$$

with initial conditions $N_0(0) = N_0$, $N_i(0) = 0$ for all $i \geq 1$. The functions $\bar{n}_i(t, \tilde{x})$ each satisfy the partial differential equation

$$\frac{\partial \bar{n}_i(t, \tilde{x})}{\partial t} - ce^{-kt} \frac{\partial [\tilde{x} \bar{n}_i(t, \tilde{x})]}{\partial \tilde{x}} = 0 \tag{20}$$

with initial condition

$$\bar{n}_i(0, \tilde{x}) = \frac{2^i \Phi(2^i \tilde{x})}{N_0}.$$

Note that, by definition,

$$N_0 = \int_0^\infty \Phi(\tilde{x}) d\tilde{x}.$$

The structured density $\tilde{n}(t, x)$ described by (17) is related to the density $n(t, x)$ (which is in terms of the actual observed fluorescence intensity) by adding the autofluorescence. Because autofluorescence may vary from cell to cell in the population [7, 19, 34, 36, 39], this is most accurately done by computing the convolution integral

$$n(t, x) = \int_0^\infty \tilde{n}(t, \tilde{x}) p(x - \tilde{x}) d\tilde{x}, \tag{21}$$

where $p(x)$ is a probability density function describing the distribution of autofluorescence in the population. Typically, it is assumed that $p(x)$ is a lognormal distribution [7, 39], and thus characterized by its mean and variance. While the computation of the convolution can be quite formidable, such an assumption on the autofluorescence density function has been incorporated into an approximation scheme which has been shown to be both accurate and quite fast. Details and associated proofs can be found in [19] and the references therein.

In separating the model into component parts and demonstrating fast approximation techniques, Allgöwer, et al., have produced an intuitive model for a population of cells structured by fluorescence intensity and division number as well as a remarkably efficient computational strategy. However, there is still the problem of how to relate the time-dependent exponential rates of cell division and death, $\alpha_i(t)$ and $\beta_i(t)$, to biologically descriptive information. To this end, we turn to a reformulation of the model above which is, in fact, made possible exactly by the model formulation and method of solution proposed by Allgöwer, et al.

When the model (17) is separated into two components (18), it is clear that the form of the system of differential equations to account for cell numbers (19) arises naturally from the mechanisms chosen to describe cell division and death in the original partial differential equation formulation (17). However,

neither the partial differential equation (17) nor the separated solution (18) are premised upon these particular forms. To that end, we consider the system of equations

$$\begin{aligned}\frac{\partial \tilde{n}_0}{\partial t} - ce^{-kt}\frac{\partial[\tilde{x}\tilde{n}_0]}{\partial \tilde{x}} &= n_0^{div}(t) - n_0^{die}(t)\\ \frac{\partial \tilde{n}_1}{\partial t} - ce^{-kt}\frac{\partial[\tilde{x}\tilde{n}_1]}{\partial \tilde{x}} &= 2n_{i-1}^{div}(t) - n_i^{div}(t) - n_i^{die}(t)\\ &\vdots\end{aligned} \tag{22}$$

where the definitions of $n_0^{div}(t)$, $n_0^{die}(t)$, n_i^{div} and n_i^{die} are given in Equations (2) - (3). This model, which is based on simple mass balance, can be solved by factorization (18) as before; the label densities $\tilde{n}(t, \tilde{x})$ are computed according to Equation (20), and the cell numbers are now provided by the cyton model as discussed in Section 3.1. Thus the new model (22) is capable of accurately describing the evolving generation structure of a population of cells while also accounting for the manner in which the CFSE profile of the population changes in time. The model is easily relatable to biologically meaningful parameters (times to division and death) and can be solved efficiently so that it is eminently suitable for use in an inverse problem.

4. Model Validation

We now consider fitting the new model (22) to the data set shown in Figure 1. It will be assumed that the autofluorescence density $p(x_a)$ is lognormal and thus can be described by its mean and variance. Let θ represent all the parameters necessary to obtain a solution to the model (22). Thus θ contains the autofluorescence mean and variance, the label decay parameters, and the parameters associated with the given cyton model (see Table 1). For a given θ, the solution for the model (22) is computed according to the technique demonstrated in [19] to produce the population density $n(t, x)$, which can be written in expanded notation as $n(t, x; \theta)$. Of course, the data in Figure 1 is reported on a $\log_{10}$ scale. The label-structured population density in the $\log_{10}$ scale is

$$\hat{n}(t, z; \theta) = 10^z \ln(10) n(t, x(z); \theta) = 10^z \ln(10) n(t, 10^z; \theta), \tag{23}$$

with the factor $10^z \ln(10)$ needed to conserve the quantity of fluorescence intensity after the change of variables. As discussed in Section 2, the data is given as a set of ordered pairs (z_k^j, n_k^j) indicating the number n_k^j of cells having

log-fluorescence intensity in the interval $[z_k, z_{k+1})$ at time t_j. Because the transformed model solution $\hat{n}(t,z)$ is a density, numbers of cells in a given region are computed by integration,

$$I[\hat{n}](t_j, z_k^j; \theta) \equiv \int_{z_k^j}^{z_k^{j+1}} \hat{n}(t_j, z; \theta) dz \approx \left[\frac{\hat{n}(t_j, z_{k+1}^j; \theta) + \hat{n}(t_j, z_k^j; \theta)}{2} \right] \left(z_{k+1}^j - z_k^j \right).$$

Following standard inverse problem procedure [1, 8, 12, 13, 37], we can define the random variables

$$N_k^j = I[\hat{n}](t_j, z_k^j; \theta_0) + \mathcal{E}_{kj}, \tag{24}$$

and consider the data n_k^j to be realizations of these random variables. The $\mathcal{E}_{kj}$ are random variables representing noise, errors, or otherwise unmodeled phenomena which cause the data to deviate from the predictions of the model. It is standard [1, 8, 13, 37] to assume that these random variables have zero expectation, $E[\mathcal{E}_{kj}] = 0$, reflecting a belief that the data arise from the mathematical model, given some nominal truth parameter θ_0. For the results presented here, it will further be assumed that $Var(\mathcal{E}_{kj}) = \sigma_0^2$, and that the errors are independent and identically distributed (*iid*). Then we may define the ordinary least squares (OLS) estimator

$$\begin{aligned} \theta_{OLS} &= \arg\min_{\theta \in \Theta} J(\theta | \{N_k^j\}) \\ &= \sum_{k,j} \mathcal{R}_{kj}^2 = \arg\min_{\theta \in \Theta} \sum_{k,j} \left(I[\hat{n}](t_j, z_k^j; \theta) - N_k^j \right)^2. \end{aligned} \tag{25}$$

The goal of the model calibration problem is to determine the estimate $\hat{\theta}$ which minimizes the OLS cost functional $J(\theta)$ given a particular data set,

$$\hat{\theta}_{OLS}(\{n_k^j\}) = \arg\min_{\theta \in \Theta} J(\theta | \{n_k^j\}) = \sum_{k,j} \nabla_{kj}^2 = \arg\min_{\theta \in \Theta} \sum_{k,j} \left(I[\hat{n}](t_j, z_k^j; \theta) - n_k^j \right)^2.$$

This minimization was carried out using the MATLAB optimization routine `fmincon`, which implements the BFGS algorithm. It should be noted that the consistency of the least squares estimator (25) is premised upon the accuracy of the assumptions that the errors are *iid* with constant variance, $Var(\mathcal{E}_{kj}) = \sigma_0^2$. While a statistical model for the error random variables $\mathcal{E}_{kj}$ is not known, recent work [39, Ch. 4] suggests that these assumptions may be only a rough approximation to reality for the CFSE proliferation data. In the absence of any additional information, we use the OLS framework here to demonstrate that the new model proposed is capable of accurately fitting data collected from

Model	Number of parameters	Cost
Cyton1	4+9	7.5175×10^{11}
Cyton2	4+10	7.5246×10^{11}
Cyton3	4+11	4.4598×10^{11}
Cyton4	4+12	4.0423×10^{11}
Cyton5	4+11	5.7226×10^{11}
Cyton6	4+12	4.0549×10^{11}
Cyton7	4+13	3.6228×10^{11}
(17)	4+29	3.8332×10^{11}

Table 2: Summary of inverse problem results fitting the model (22) containing the cyton dynamics described in Table 1. The Cyton7 model clearly has the lowest cost of the models tested, although it also contains more parameters than any of the other cyton-based models. Significantly, the Cyton7 model outperforms the model (17) presented in [7, 39], while also having fewer total parameters. We remark that the cost of the model (17) is slightly higher than that reported in [7, 39] as a result of the different approximation and computational approaches used.

a CFSE proliferation assay (see Figure 2). Further work to characterize an accurate statistical model (and hence an unbiased estimator) is ongoing. More details about this inverse problem in general can be found in [39, Ch. 3].

Following the work presented in Section 3.1, multiple variations of the model (following from different implementations of the cyton model) were fit to the data shown in Figure 1. As a means of comparison, the model (17) was also tested, recreating the results from [7, 39] but with the new computational framework of [19, 36]. For this model, the proliferation rates $\alpha_i(t)$ for each generation are considered to be piecewise linear functions of time and the death rates $\beta_i(t)$ are assumed to be constant. Such a model has been shown to accurately fit the given data set (see [7, 39] for details).

Table 2 summarizes the results from the various inverse problems. Not surprisingly, the most complex cyton model (Cyton7) is also the most accurate. Significantly, this model is seen to outperform the old model (17) for the data set studied, even though it uses fewer parameters. As discussed above, the new model (with cyton dynamics) is more biologically intuitive than the model (17) as well. Of course, this comparison is only valid for the single data set examined in this report; further validation with additional data will be required before

any broad conclusions can be reached, and this work is ongoing. The fit of each of these two models to the data can be seen in Figure 2. Visually, there are only a few differences between the two models. While both fit the data quite well, the Cyton7 model is more accurate at $t = 24$ hours. This is most likely the result of the form of Equation (12), which permits a rapid (exponential) initial decrease in the number of undivided cells, followed by a much slower subsequent decline. Meanwhile the model (17), with $\beta_0(t)$ a constant function, permits only an exponential rate of death for the duration of the experiment; the estimated parameter $\beta_0(t) = \beta_0$ is thus biased downward in order to avoid losing too many undivided cells at subsequent measurement times. The only major discrepancies between the Cyton7 model and the data occur for $z \geq 3$ at $t = 24$ hours and for $z \approx 0.5$ at $t = 120$ hours. Both cases are explained by irregularities in the particular data set–aggregates and/or monocytes in the former case, and irregular bin spacing in the latter case. See [4, 7, 39] for details. We remark that the cost of the model (17) is slightly higher than that reported in [7, 39] as a result of the different approximation and computational approaches used.

It is more difficult to use Table 2 to reach similar conclusions regarding the performances of the various cyton models considered. While Cyton7 clearly has the lowest cost, it also contains the largest number of parameters. Model comparison tests based upon refinement [1, 2] or information criteria such as the AIC [10] exist, but each of these requires an accurate statistical model for the error processes in the data. While model comparison statistics could be computed from the information in Table 2, the results would be heavily biased by the limited accuracy of the statistical model, and perhaps misleading. Rather, we assess the fits of the various models qualitatively.

Aside from the model (17), the best models are Cyton3, Cyton4, Cyton6, and Cyton7, each of which contains parameters for a division destiny curve. Because other model generalizations (see Section 3.1) such as exponential death for undivided cells with (Cyton5) or without (Cyton2) a fraction of idle cells do not result in nearly as much improvement in the fit of the model, it seems reasonable to conclude that *division destiny* is an important feature of an accurate mathematical model of cell division. The fit of the model to the data for Cyton5, the best model lacking a division destiny curve, is shown in Figure 3. While this model does again fit the data fairly well, there are systematic errors in the population generation structure as predicted by the model, particularly at $t = 96$ hours.

Aside from Cyton7, the most accurate cyton-based model considered is Cyton4. The best-fit model solution for Cyton4 is shown in comparison to

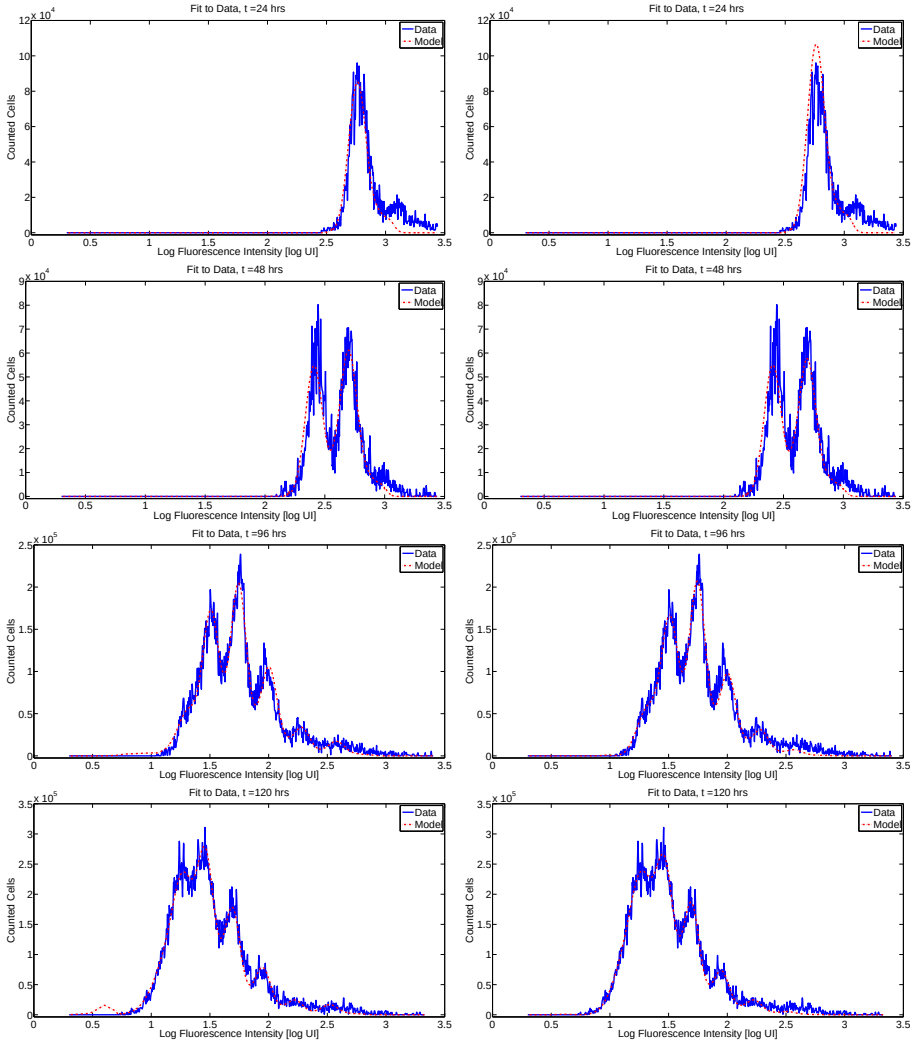

Figure 2: Best-fit results for models Cyton7 (Left) and (17) (Right). The two models have very similar solutions, but the Cyton7 model has a lower cost and fewer parameters (see Table 2). The cyton model is much more accurate at $t = 24$ hours (top), probably because the form of Equation (12) permits a rapid (exponential) initial decrease in the number of undivided cells, followed by a much slower subsequent decline. Meanwhile, the model (17), with $\beta_0(t)$ a constant function, permits only an exponential rate of death for the duration of the experiment. The small bump in the Cyton7 solution near $z = 0.5$ at $t = 120$ hours is actually the result of irregular bin spacing for this particular data set (see [4, 7, 39]).

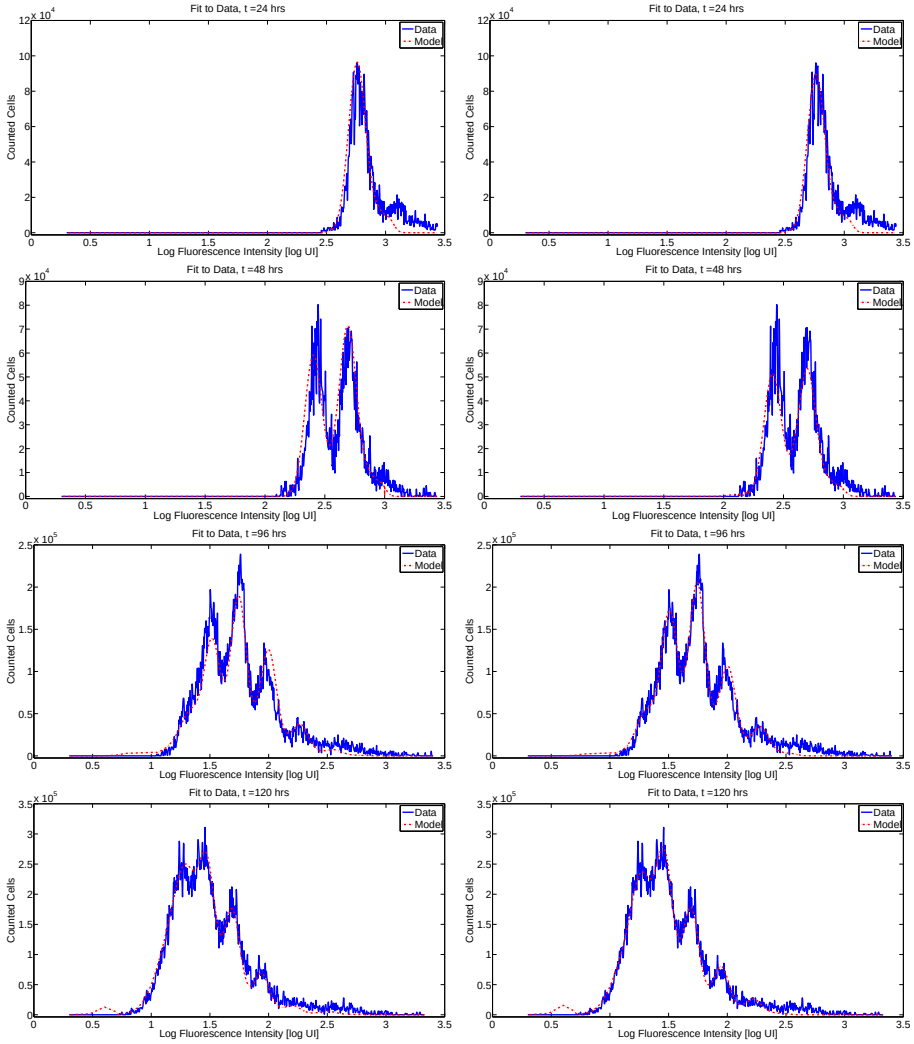

Figure 3: Left: Best-fit results for model Cyton5. While the model is fairly accurate, there are systematic errors in the population generation structure predicted by the model, particularly at $t = 96$ hours. The Cyton7 model (Figure 2, Left) provides a more accurate prediction of the CFSE profile at all time points and thus has a lower cost (Table 2). Right: Best-fit results for model Cyton4. Again, we see that the model is fairly accurate, though the absence of any non-progressing, idle cells in the undivided population results in an under-prediction for $t \geq 96$ hours.

the data in Figure 3. The only difference between Cyton7 and Cyton4 is that Cyton7 considers a fraction p_{idle} of non-progressing cells which are assumed to neither die nor divide. While the difference in both OLS cost as well as model solution are minimal between Cyton4 and Cyton7, careful observation of the Cyton4 solution at $t = 120$ hours reveals that the model under-predicts the number of undivided cells remaining in the population. It appears then, that the *fraction of non-progressing, idle cells* is also an important consideration for an accurate mathematical model.

5. Summary and Future Work

In this report, two existing frameworks (*models for label structure dynamics* and *models for generation structure*) have been combined to form an accurate, biologically meaningful quantitative description of a dividing population of cells labeled with an intracellular dye. The cyton model, which relates the generation structure of the population to probability distributions over times to divide and die, has been shown to describe the behavior of a proliferating population of cells in a manner that is both accurate and consistent with a range of experimental evidence [11, 21, 21]. Meanwhile, models of population label structure, which account for the manner in which cells process an intracellular label as they grow, divide, and die, have been shown to accurately predict the evolution of histogram profiles indicating the label-structured density of the population [4, 5, 7, 19, 24, 25, 36, 39]. Using a recently developed solution and computational technique [19, 36], we have shown how these two bodies of work can be combined into a model which can simultaneously describe both label dynamics and population generation structure.

The primary advantage of such a technique is that one is no longer dependent on any deconvolution of histogram data in order to determine the underlying generation structure of a population of cells measured by flow cytometry. The resulting model, once fit to data, provides an estimate not only on the histogram profiles themselves, but also of the number of cells having completed a given number of divisions. For instance, one can use the validated or calibrated model to determine the numbers of cells in the population in terms of division number (Figure 4) or the population generation structure (Figure 5) at any given time. While this was also possible with previous models (e.g., (17), [7, 19, 36, 39]), these models are not amenable to describing such biologically relevant features as times to division and death. Yet the new model, which contains the cyton model, is derived exactly in terms of these quantities. For

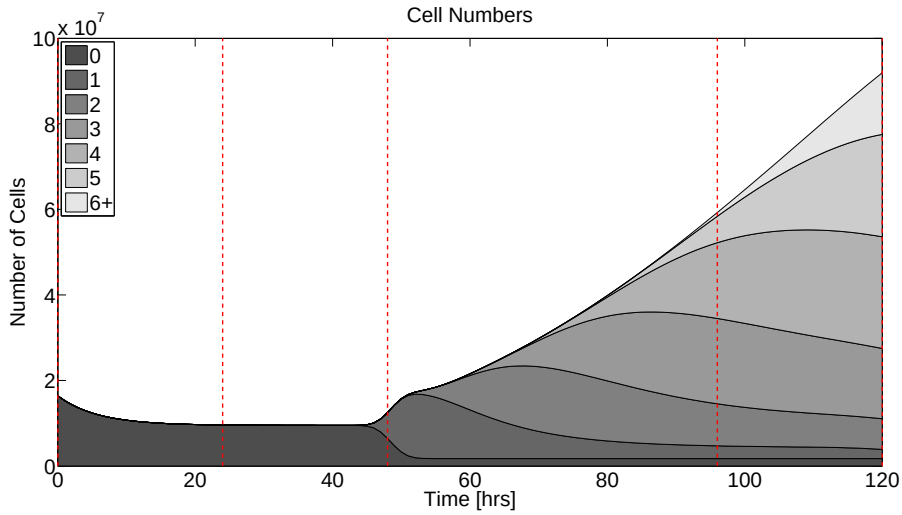

Figure 4: Total numbers of cells, and generational structure, as computed with the model Cyton7. Vertical dashed lines at $t = 0$, 24, 48, 96, and 120 hours indicate times at which data is available for the particular data set used to calibrate the model.

instance, the estimated probability densities for times to division and death are shown for undivided precursors and divided cells in Figure 6.

As discussed in Section 3.1, additional generalizations of the new model are possible. For instance, one could consider a density function other than lognormal to describe times to division and death. Given the success of the model Cyton7, which features a combined lognormal and exponential probability of death (as well as a fraction of idle cells), it is natural to wonder whether the true density function might even be multimodal, possibly reflecting multiple subpopulations within the data. Similarly, the time to division curve for divided cells (Figure 6, right) admits the possibility that a non-negligible fraction of cells divides in less than two hours. This is almost certainly non-biological. It is possible that this problem can be also explained as the result of subpopulations which are characterized by different distributions over times to divide and die; when a single parametric distribution (e.g., lognormal) is used to describe such data, the statistical moments of the estimated distribution may not accurately reflect the multiple underlying subpopulation distributions. While some experimental evidence exists to support the lognormal parametric forms for the cyton model for some types of cells [20], a rigorous analysis of the effects of the

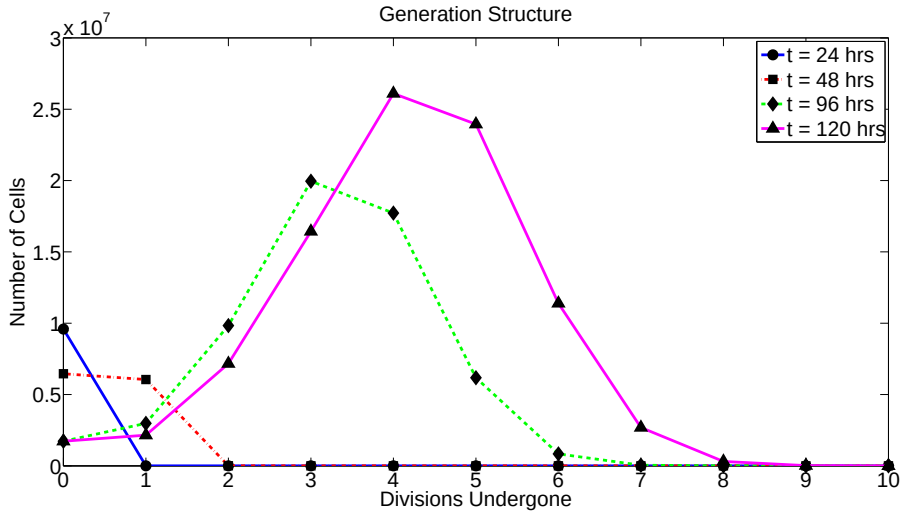

Figure 5: Similar to Figure 4, the generation structure of the population of cells at each measurement time.

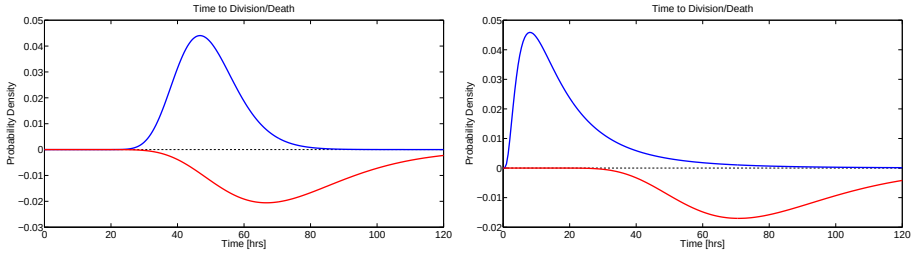

Figure 6: Lognormal distributions of division (blue) and death (red; inverted) times for undivided (Left) and divided (Right) cells. The density functions shown for division apply only to progressing cells. The density function shown for undivided cell death applies only to progressors (i.e., it does not include the exponential death of non-progressors, nor the fraction of idle cells).

parametric form on the performance of the model has not been performed to our knowledge. This problem may also be amenable to nonparametric estimation [3] of the underlying probability distributions.

The accuracy and computational efficiency of the new model (22), as well as its intuitive biological interpretation, make it an important tool for use in analyzing data collected from flow cytometry proliferation assays. Regardless of the potential improvements and/or model generalizations just discussed, the primary limitation of the current framework lies not in the mathematical model itself but rather in the statistical model (24) which links the mathematical model to the data. An accurate statistical model is of vital importance for the consistent estimation of model parameters, as well as for uncertainty quantification for estimated parameters, e.g., the unbiased estimation of confidence intervals around those parameters [1, 8, 13, 37]. An accurate statistical model is also necessary for the rigorous comparison of different model parameterizations and generalizations [2, 10]. To that end, an array of data has been collected and efforts to demonstrate an accurate statistical model of the data are underway. In the mean time, we believe the new model presented here will provide an important contribution for the quantitative analysis of proliferation assay data.

Acknowledgments

This research was supported in part by the National Institute of Allergy and Infectious Disease under grant NIAID 9R01AI071915 and in part by the U.S. Air Force Office of Scientific Research under grant AFOSR-FA9550-12-1-0188. The authors are grateful to their Barcelona collaborators (Jordi Argilaguet, Cristina Peligero and Andreas Meyerhans) for numerous discussions to help them understand CFSE labeled cell proliferation data and the experimental procedures for collecting this data.

References

[1] H.T. Banks, M. Davidian, J. Samuels, and K.L. Sutton, An inverse problem statistical methodology summary, CRSC-TR08-01, North Carolina State University, January 2008; Chapter 11 in *Mathematical and Statistical Estimation Approaches in Epidemiology*, G. Chowell, et al., eds., Berlin Heidelberg New York, 2009, pp. 249–302.

[2] H.T. Banks and B.G. Fitzpatrick, Inverse problems for distributed systems: statistical tests and ANOVA, LCDS/CSS Report 88-16, Brown University, July 1988; *Proc. International Symposium on Math. Approaches to Envir. and Ecol. Problems*, Springer Lecture Notes in Biomath., **81** (1989), 262–273.

[3] H.T. Banks, Zackary R. Kenz, and W. Clayton Thompson, A review of selected techniques in inverse problem nonparametric probability distribution estimation, to appear.

[4] H.T. Banks, Karyn L. Sutton, W. Clayton Thompson, G. Bocharov, Marie Doumic, Tim Schenkel, Jordi Argilaguet, Sandra Giest, Cristina Peligero, and Andreas Meyerhans, A new model for the estimation of cell proliferation dynamics using CFSE data, CRSC-TR11-05, North Carolina State University, Revised July 2011; *J. Immunological Methods,* **373** (2011), 143–160; DOI:10.1016/j.jim.2011.08.014.

[5] H.T. Banks, Karyn L. Sutton, W. Clayton Thompson, Gennady Bocharov, Dirk Roose, Tim Schenkel, and Andreas Meyerhans, Estimation of cell proliferation dynamics using CFSE data, CRSC-TR09-17, North Carolina State University, August 2009; *Bull. Math. Biol.*, **70** (2011), 116–150.

[6] H.T. Banks and W. Clayton Thompson, Mathematical models of dividing cell populations: Application to CFSE data, CRSC-TR12-10, N. C. State University, Raleigh, NC, April, 2012; *J. Math. Modeling of Natural Phenomena*, submitted.

[7] H.T. Banks, W. Clayton Thompson, Cristina Peligero, Sandra Giest, Jordi Argilaguet, and Andreas Meyerhans, A division-dependent compartmental model for computing cell numbers in CFSE-based lymphocyte proliferation assays, CRSC-TR12-03, North Carolina State University, January 2012; *Math Biosci. Eng.*, submitted.

[8] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes*, CRC Press, Boca Raton London New York, 2009.

[9] G. Bell and E. Anderson, Cell growth and division I. A mathematical model with applications to cell volume distributions in mammalian suspension cultures, *Biophysical Journal*, **7** (1967), 329–351.

[10] K.P. Burnham and D.R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd Edition), Springer, New York, 2002.

[11] R. Callard and P.D. Hodgkin, Modeling T- and B-cell growth and differentiation, *Immunological Reviews*, **216** (2007), 119–129.

[12] R.J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*, Chapman Hall, London, 2000.

[13] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London, 2000.

[14] R.J. DeBoer, V.V. Ganusov, D. Milutinovic, P.D. Hodgkin, and A.S. Perelson, Estimating lymphocyte division and death rates from CFSE data, *Bull. Math. Biol.*, **68** (2006), 1011–1031.

[15] R.J. DeBoer and Alan S. Perelson, Estimating division and death rates from CFSE data, *J. Comp. and Appl. Mathematics*, **184** (2005), 140–164.

[16] E.K. Deenick, A.V. Gett, and P.D. Hodgkin, Stochastic model of T cell proliferation: a calculus revealing IL-2 regulation of precursor frequencies, cell cycle time, and survival, *J. Immunology*, **170** (2003), 4963–4972.

[17] V.V. Ganusov, S.S. Pilyugin, R.J. De Boer, K. Murali-Krishna, R. Ahmed, and R. Antia, Quantifying cell turnover using CFSE data, *J. Immunological Methods*, **298** (2005), 183–200.

[18] A.V. Gett and P.D. Hodgkin, A cellular calculus for signal integration by T cells, *Nature Immunology*, **1** (2000), 239–244.

[19] J. Hasenauer, D. Schittler, and F. Allgöwer, A computational model for proliferation dynamics of division- and label-structured populations, `arXive.org`, arXiv:1202.4923v1,22Feb,2012.

[20] E.D. Hawkins, Mirja Hommel, M.L Turner, Francis Battye, J Markham and P.D Hodgkin, Measuring lymphocyte proliferation, survival and differentiation using CFSE time-series data, *Nature Protocols*, **2** (2007), 2057–2067.

[21] E.D. Hawkins, M.L. Turner, M.R. Dowling, C. van Gend, and P.D. Hodgkin, A model of immune regulation as a consequence of randomized lymphocyte division and death times, *Proc. Natl. Acad. Sci.*, **104** (2007), 5032–5037.

[22] H.Y. Lee, E.D. Hawkins, M.S. Zand, T. Mosmann, H. Wu, P.D. Hodgkin, and A.S. Perelson, Interpreting CFSE obtained division histories of B cells in vitro with Smith-Martin and Cyton type models, *Bull. Math. Biol.*, **71** (2009), 1649–1670.

[23] H.Y. Lee and A.S. Perelson, Modeling T cell proliferation and death in vitro based on labeling data: generalizations of the Smith-Martin cell cycle model, *Bull. Math. Biol.*, **70** (2008), 21–44.

[24] T. Luzyanina, D. Roose, and G. Bocharov, Distributed parameter identification for a label-structured cell population dynamics model using CFSE histogram time-series data, *J. Math. Biol.*, **59** (2009), 581–603.

[25] T. Luzyanina, D. Roose, T. Schenkel, M. Sester, S. Ehl, A. Meyerhans, and G. Bocharov, Numerical modelling of label-structured cell population growth using CFSE distribution data, *Theoretical Biology and Medical Modelling*, **4** (2007), Published Online.

[26] A. B. Lyons, Divided we stand: tracking cell proliferation with carboxyfluorescein diacetate succinimidyl ester, *Immunology and Cell Biology*, **77** (1999), 509–515.

[27] A. B. Lyons, J. Hasbold and P.D. Hodgkin, Flow cytometric analysis of cell division history using diluation of carboxyfluorescein diacetate succinimidyl ester, a stably integrated fluorescent probe, *Methods in Cell Biology*, **63** (2001), 375–398.

[28] A. B. Lyons and K. V. Doherty, Flow cytometric analysis of cell division by dye dilution, *Current Protocols in Cytometry*, (2004), 9.11.1-9.11.10.

[29] A.B. Lyons and C.R. Parish, Determination of lymphocyte division by flow cytometry, *J. Immunol. Methods*, **171** (1994), 131–137.

[30] G. Matera, M. Lupi and P. Ubezio, Heterogeneous cell response to topotecan in a CFSE-based proliferative test, *Cytometry A*, **62** (2004), 118–128.

[31] J.A. Metz and O. Diekmann, *The Dynamics of Physiologically Structured Populations*, Springer Lecture Notes in Biomathematics **68**, Heidelberg, 1986.

[32] H. Miao, X. Jin, A. Perelson and H. Wu, Evaluation of multitype mathemathematical modelsfor CFSE-labeling experimental data, *Bull. Math. Biol.*, **74** (2012), 300–326; DOI 10.1007/s11538-011-9668-y.

[33] C. Parish, Fluorescent dyes for lymphocyte migration and proliferation studies, *Immunology and Cell Biol.*, **77** (1999), 499–508.

[34] B.J.C. Quah and C.R. Parish, New and improved methods for measuring lymphocyte proliferation *in vitro* and *in vivo* using CFSE-like fluorescent dyes, *J. Immunological Methods*, (2012), to appear.

[35] B. Quah, H. Warren, and C. Parish, Monitoring lymphocyte proliferation in vitro and in vivo with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester, *Nature Protocols*, **2** (2007), 2049–2056.

[36] D. Schittler, J. Hasenauer, and F. Allgöwer, A generalized model for cell proliferation: Integrating division numbers and label dynamics, *Proc. Eighth International Workshop on Computational Systems Biology (WCSB 2011)*, June 2011, Zurich, Switzerland, p. 165–168.

[37] G.A. Seber and C.J. Wild, *Nonlinear Regression*, Wiley, Hoboken, 2003.

[38] J.A. Smith and L. Martin, Do cells cycle?, *Proc. Natl. Acad. Sci.*, **70** (1973), 1263–1267.

[39] W. Clayton Thompson, *Partial Differential Equation Modeling of Flow Cytometry Data from CFSE-based Proliferation Assays,* Ph.D. Dissertation, Dept. of Mathematics, North Carolina State University, Raleigh, December, 2011.

[40] P.K. Wallace, J.D. Tario, Jr., J.L. Fisher, S.S. Wallace, M.S. Ernstoff, and K.A. Muirhead, Tracking antigen-driven responses by flow cytometry: monitoring proliferation by dye dilution, *Cytometry A*, **73** (2008), 1019–1034.

[41] Hilary S. Warren, Using carboxyfluorescein diacetate succinimidyl ester to monitor human NK cell division: Analysis of the effect of activating and inhibitory class I MHC receptors, *Immunology and Cell Biology*, **77** (1999), 544–551.

[42] J.M. Witkowski, Advanced application of CFSE for cellular tracking, *Current Protocols in Cytometry*, (2008), 9.25.1–9.25.8.