OPTIMAL CONTROL MODELING OF CELL DIVISION

M. A. MAMMADOV, C. WELLARD, AND R. J. EVANS

Abstract This paper investigates the population dynamics of a system of identically prepared B cells whose proliferation trajectories have been individually tracked using live-cell imaging techniques. The main goal is to investigate whether the system behavior can be determined using an optimality criterion. In order to achieve this goal we assume the existence of an intracellular mechanism that controls all cell divisions after initial stimulation with the aim of optimizing some objectives within predefined constraints that include the initial number of cells in the system, as well as the total number of cell divisions observed in the response. Numerical experiments are carried out using recent data that combines cells in 14 different wells. For each well the curve representing the number of live cells over time, is accurately approximated by the optimal solution to the control problem proposed.

1. Introduction

The population dynamics of *in vitro* stimulated lymphocyte responses have become the subject of intense study in recent years with increasing numbers of researchers attempting to describe the behaviour with the use of mathematical models. The models used vary in complexity with most being significantly more sophisticated than the classic Smith-Martin model of cell division [18] originally proposed over 40 years ago. A number of models allow for random division and death times, which can be division dependent [21, 11, 19, 20], with some taking into account the effects of observed correlations in division times between related cells [8, 16, 14]. A number of different mathematical approaches to the solution of the model have been considered, from numerical solution of a set of ordinary differential equations for the mean behaviour [21, 11] to discrete time [17] as well as continuous time [15, 19] branching processes which allow calculation of higher order moments.

The main purpose of such descriptions in the literature to date is the extraction of model parameter values via numerical fitting of the model to experimental data. In this article we take a control system approach to investigate whether the proliferation dynamics of an observed B cell response is, in some sense, optimal. With this goal in mind we use a simplified mathematical model to describe the number of live cells produced following *in vitro* CpG stimulation of a population of B cells. The experimental data used is that presented in [12], which gives further details of the experimental setup. In summary we have used time-lapse microscopy to follow the evolution of mutliple clones resulting from the CpG stimulation of a population of small resting B cells, the founder cells. We use results from two separate experiments, in each of which the founder cells are contained in multiple wells on a terasaki plate. The number of founder cells within each well varies, but all wells

Key words and phrases. Modeling and optimal control; Cell population dynamics; Immune system; B-cell proliferation.

have been identically prepared. The proliferation of each clone within these wells has been tracked individually, with special care taken to maintain cell identity, such that a complete lineage tree is available for each clone, with times to the division or death of each progeny. To describe this data we use a simplified model which assumes that following initial division of each founder cell within a well, subsequent progeny divide after a uniform period, with a division dependent probability, otherwise they die. Although these subsequent division and death times are considered uniform for cells within each well, they are free to vary between wells, as are the division dependent division probabilites.

2. Optimal Control Model

We use the following notations: X^0 denotes the number of activated founder cells, τ^1_{div} and τ_{div} are the average times to divide for the first generation and for the consequent generations, respectively. τ_{death} stands for the average time to die.

Consider the time interval $[0, T]$ that contains the life span of all descendant cells. Denote by m the maximum number of divisions that one founder cell may undergo and let $t_0 = 0$, $t_1 = \tau^1_{div}$, and $t_i = t_1 + (i-1) \cdot \tau_{div}$, $i = 2, 3, ..., m$. For the B cells in our experimental data, the number m is less than 8.

By assumption the interval $[0, T]$ is large enough to accommodate all points t_i as well as $t_n + \tau_{death}$; that is,

$$0 = t_0 < t_1 < ... < t_m < t_m + \tau_{death} < T, \forall i = 0, 1, ..., m. \tag{1}$$

Let u_i, $i = 1, ..., m$ denote the proportion of cells at time t_{i-1} that are going to divide at t_i. Clearly $u_i \in [0, 1]$, $i = 1, ..., m$.

2.1. **Control and Population dynamics.**

The proportions u_i, $i = 1, ..., m$, will be considered as control parameters that define the growth of cell population. Our data suggests that these proportions decrease with division and so we make the assumption that the sequence $(u_1,\ u_2, ...,\ u_m)$ is not increasing.

Thus, a *control* $\mathbf{u}$ is defined as a sequence of m non-increasing numbers with values in the interval [0,1]:

$$\mathbf{u} = (u_1,\ u_2, ...,\ u_m), \text{ where } 0 \le u_m \le ... \le u_2 \le u_1 \le 1. \tag{2}$$

Given any control sequence $\mathbf{u}$, we can directly calculate the state function $x(t)$ - the population dynamics.

Finding the number of live cells at any given time t, is a straightforward calculation thanks to the fixed division and death times - t_i, τ_{div} and τ_{death}. Clearly, according to this model, the number of live cells is discontinuous at points where cell-division or cell-death occurs. In order to have a continuous state function $x(t)$ we use the following scheme.

First we calculate the following linear approximation $\widehat{x}(t)$:

$$\widehat{x}(t) = \sum_{k=1}^{n} c_k(t); \tag{3}$$

where n is the total number of cells in the interval $[0, T]$ and function $c_k(t)$ is defined differently for cells undergoing division and death as follows.

- For a dividing cell k in the interval $[t_i, t_{i+1}]$ we set

$$c_k(t) = 1 + \frac{t - t_i}{t_{i+1} - t_i},\ \forall t \in [t_i, t_{i+1}],\ \text{and } c_k(t) = 0 \text{ otherwise.} \tag{4}$$

- For a dying cell k in the interval $[t_i, t_{i+1}]$ we set

$$c_k(t) = 1,\ \forall t \in [t_i, t_{i+1}],\ \text{and } c_k(t) = 0 \text{ otherwise.} \tag{5}$$

Note 2.1. *Function $\widehat{x}(t)$ defined in this way, could be interpreted as being proportional to the total mass of live cells. It is known that (see for example* [12]*), the size of a particular cell fated to die remains relatively stable until the time of its death, meanwhile the size of a dividing cell increases roughly linearly. The above formula assumes the same size at division for all cells. In addition, formula (4) assumes that the size of a dividing cell increases linearly achieving the double size just before dividing.*

The total number of cells n used in (3) is directly related to the total number of divisions over the whole interval. It will be one of the main external parameters of the control model developed in this section. Denoting the total number of divisions by D^{div}, we obtain the relation

$$n = X^0 + 2D^{div}, \tag{6}$$

where X^0 is the number of founder cells. The following property is true:

Proposition 2.2. *If $\tau_{div}^1 = \tau_{div}$ then*

$$\int_0^T \widehat{x}(t)dt = 1.5\, D^{div}\, \tau_{div} + (X^0 + D^{div})\, \tau_{death}. \tag{7}$$

The proof of this Proposition is trivial. In this integral, each divided cell contributes $1.5 \cdot \tau_{div}$ due to formula (4) and each dying cell contributes $1 \cdot \tau_{death}$ due to formula (5). Moreover, the number of divided cells is D^{div} while the number of cells to die is $X^0 + D^{div}$. Therefore, (7) is true.

Note 2.3. *The relation (7) expresses a very important property. It shows that all the control sequences* $\mathbf{u}$*, realizing the same number of divisions -D^{div}, produce cell population curves $\widehat{x}(t)$ having the same integral over $[0, T]$.*

The amplitude of function $\widehat{x}(t)$ that is, $\widehat{x}^{max} = \max_{t\in[0,T]} \widehat{x}(t)$ will play an important role in our future analysis. In this sense, the following proposition will prove useful.

Proposition 2.4. *If τ_{death} is sufficiently large so that the first cell death occurs after the last cell division (e.g. if $\tau_{death} > \tau_{div}^1 + (m-1)\tau_{div}$ assuming that the number of cell generations is m), then*

$$\widehat{x}^{max} = D^{div} + X^0. \tag{8}$$

The proof of this proposition is again obvious. X^0 is the starting point and each division adds +1 to the amplitude. Therefore, if no cell death occurs while cells continue dividing, then relation (8) will be satisfied.

However, in reality some cells die before the cell population reaches its amplitude. In this case one would expect to have the above relation in the form

$$\widehat{x}^{max} = \lambda D^{div} + X^0. \tag{9}$$

It turns out that, the coefficient λ is relatively stable for different cell populations (wells). This is an interesting fact and will be discussed in Section 3.1.

Clearly, function $\widehat{x}(t)$ constructed above is discontinuous at points $t_i + \tau_{death}$ related to the cell death. Next we "smooth" it by defining a new function $x(t)$ that will be used to represent the live cell population:

$$x(t) = \frac{1}{2\delta(t)} \int_{t-\delta(t)}^{t+\delta(t)} \widehat{x}(t)dt. \tag{10}$$

Here $\delta(t)$ is a continuous function satisfying $\delta(0) = 0$ and $\delta(T) = 0$. We set

$$\delta(t) = \begin{cases} t, & \text{if } t \in [0, \Delta) \\ \Delta, & \text{if } t \in [\Delta, T - \Delta] \\ T - t, & \text{if } t \in (T - \Delta, T] \end{cases} \tag{11}$$

where $\Delta > 0$ determines the length of the interval to average the values of $\widehat{x}(t)$.

Function $x(t)$ defined in this way could be interpreted as a cell population corresponding to some "continuous" control function $u(t)$ defined on the interval $[0, T]$, where the value $u(t)$ gives the proportion of normal B cells at time t that are going to divide. In other words, formula (10) simulates the case when the times to divide and to die are not fixed but are allowed to vary slightly from their uniform values. Nevertheless, such an averaged method still keeps an essential peculiarity of this optimal control process where the desision/control parameters are determined at certain discrete times.

Given control sequence $\mathbf{u}$ in (2), we use the notation $\mathcal{F}$ to denote the operator defined by (3)-(10) that produces the state function $\mathbf{x} = x(t)$:

$$\mathbf{x} = \mathcal{F}(\mathbf{u}). \tag{12}$$

We note that the "smoothing" function $\widehat{x}(t)$ is important in terms of eliminating discontinuity in population curves/functions, especially when measuring the distance between such curves.

2.2. **External parameters.**

To formulate an optimal control problem, we need to define the external parameters of our model which will be used as constraints for the state function. The following parameters will be used in our future investigations:

- the number of total divisions - D^{div};
- the maximum number of live cells which can occur at some particular time, denoted by N^{max}.

Number D^{div} determines the area under the live cell population curve (Proposition 2.4). Number N^{max} determines the peak value of this curve. Here we will not discuss biological meanings of these parameters, they could be influenced by many factors including the strength of activation of the founder cells and the capacity of the environment to support continued cell division.

Therefore, there are 6 parameters in our model: $X^0, D^{div}, N^{max}, \tau^1_{div}, \tau_{div}$ and τ_{death}. The first 3 could be considered as external parameters. As mentioned before, there is a relation between these parameters in the form of $N^{max} = \lambda D^{div} + X^0$ that is discussed in Section 3.1.

2.3. **Possible objectives of cell division.**

The aim of this section is to discuss some possible objectives that the cell growth may follow. The primary function of B cells in the mammalian immune response is the production of antibodies, which are used to clear pathogens. It therefore seems logical that one objective of the B cell response would be to produce the maximum quantity of antibody in the shortest time possible to best clear an offending pathogen. In this current study we do not have measurements of antibody titre and so we restrict ourselves to objectives that have to do with live cell numbers, however we note that it is likely that antibody production capability would be proportional to the number of live B cells in the system. With this in mind we have identified the following three objectives which we will use in our optimal control analysis:

Objective 1. Maximize the amplitude of live cell population, denoted by

$$x^{max} = \max\{x(t) : \ t \in [0, T]\}.$$

Objective 2. Reach this maximum amplitude x^{max} in the shortest time.

Objective 3. Maximize the life span of cell population.

Before going any further, a brief discussion about these objectives would be useful. Firstly, it is quite clear that considering just one of these objectives can not work well for different reasons. For example, Objectives 1 and 2 are active until trajectory reaches its peak, so they do not cover the entire duration of the response $[0, T]$. Conversely, objective 3 mainly influences the tail of the trajectory.

Objectives 1 and 2 may look similar, but in the presence of an upper bound N^{max} to the amplitude x^{max} they express quite different targets. In order to realize all the "required" divisions (D^{div}), it might be preferable to reach the amplitude a little bit later allowing all divisions to happen.

It is not difficult to observe that Objectives 2 and 3 are "almost" contradictory in the presence of limitations to the number of cell divisions. This creates a trade-off situation that is part of any multi-objective optimization process. Thus, these two objectives can produce

"reasonable" trajectories; however, the numerical experiments below clearly indicate that trajectories generated are much "better" (closer to the actual cell population) when the first objective is also involved.

Objectives 1 and 3 are also contradictory in the presence of limitations to the number of cell divisions. In the majority of cases these two objectives produce quite good trajectories, sometimes even better than if all 3 objectives are used. However, in general, the presence of the second objective is important especially for large populations.

To express Objectives 2 and 3 numerically, we introduce the following two notations.

1. $T^{max} = T^{max}(\mathbf{x})$ denotes the time that state function $\mathbf{x} = x(t)$ reaches its maximal value; that is:

$$x(T^{max}) = x^{max}, \text{ and } x(t) < x^{max}, \forall t < T^{max}.$$

2. $i^{max} = i^{max}(\mathbf{u})$ denotes the maximal positive index of $\mathbf{u} = (u_1, \ldots, u_m)$; that is,

$$u_{i^{max}} > 0, \text{ and } u_i = 0, \forall i > i^{max}.$$

Number i^{max} denotes the number of generations - the latest division. Since D^{death} is fixed, The objective 3 is equivalent to maximizing i^{max}.

2.4. **Formulation of optimal control problem.**

By using the objectives defined above, we can generate quite different functionals that could be used to formulate different optimal control problems. In this paper we take the approach of using a weighted sum of these three objectives as the objective functional to be maximised.

$$\begin{aligned} f(\mathbf{u}) = \quad & A \cdot x^{max} && \text{(Objective 1)} \\ - \quad & B \cdot T^{max}(\mathbf{x}) && \text{(Objective 2)} \\ + \quad & C \cdot i^{max}(\mathbf{u}). && \text{(Objective 3)} \end{aligned} \tag{13}$$

Here $\mathbf{x}$ is a state function corresponding to a given control $\mathbf{u}$.

Given a control sequence $\mathbf{u} = (u_1, \ldots, u_m)$, the number of total divisions, denoted by $TD(\mathbf{u})$, is given by the formula:

$$TD(\mathbf{u}) = X^0 \cdot \sum_{k=1}^{m} 2^{k-1} \prod_{j=1}^{k} u_j, \tag{14}$$

where X^0 is the number of founder cells. In this sum, the first term $X^0 u_1$ represents the first generation of dividing cells, $2X^0 u_1 u_2$ represents the second generation of dividing cells,

etc. These numbers should be integers; so it is natural to require that

$$2^{k-1}X^0\prod_{j=1}^{k}u_j\in\mathbb{Z}\doteq\{0,1,2,\cdots\},\ \forall k.$$

The aim is to maximise functional (13) subject to the constraints outlined above. The optimal control problem is formulated below.

Problem: Given an initial number of (founder) cells X^0 and two (external) parameters: the number of total divisions (D^{div}) and the maximum population level (N^{max}) :

$$\text{Maximize}: f(\mathbf{u}); \tag{15}$$

$$\text{Subject to}: \quad \mathbf{x}=\mathcal{F}(\mathbf{u}), \qquad \mathbf{u}=(u_1,\dots,u_m); \tag{16}$$

$$0\le u_m\le ...\le u_1\le 1; \tag{17}$$

$$2^{k-1}X^0\textstyle\prod_{j=1}^{k}u_j\in\mathbb{Z}, \quad \forall k\in\{1,...,m\}; \tag{18}$$

$$TD(\mathbf{u})=D^{div}; \tag{19}$$

$$x(t)\in[0,N^{max}], \qquad \forall t\in[0,T]. \tag{20}$$

In this formulation, the relation (20) represents inequality constraints for the state function. (19) represents an equality constraint applied to the control parameters. Recall that the total number of divisions $TD(u)$ is defined by (14).

3. Results on Fam2 and Fam3

In this section we provide some numerical experiments involving time sequence data from two different experiments Fam2 and Fam3, that contain 107 and 89 founders cells respectively (see [12] for details). Filming started 24 hours after initial stimulation at a frame rate of one image every two minutes. The total interval, in units of frames, is $[0,T]=[0,3500]$, which represents 7000 min $\approx$ 117 hours. The maximum number of divisions that one founder cell was observed to undergo was eight, and so in this study we set $m=8$.

Table 1 summarizes the main parameters of different wells in Fam2 and Fam3 used in our numerical experiments below.

3.1. **Dynamics of cell population in different wells.**

Denote by $\tilde{N}(t)$ the total number of cells at time $t\in[0,3500]$. Clearly, $\tilde{N}(t)$ is a discontinuous function with jumps at the points with cell division/death. A piecewise linear function, that will be denoted by $N(t)$, can naturally be generated similar to the above described method for generating $x(t)$ in section 2.1.

TABLE 1. Data - Fam2 and Fam3. X^0 is the number of founder cells, D^{div} is the total number of divisions, N^{max} is the maximal level of population size, τ^1_{div} is the average time to divide for the first generation (after 24 hours of simulation), τ_{div} is the average time to divide for the consequent generations, τ_{death} is the average time to die. Each unit corresponds to 2 minutes.

Fam	Well	X^0	D^{div}	N^{max}	τ^1_{div}	τ_{div}	τ_{death}
2	1	9	58	53.45	360.89	319.65	828.10
	2	8	96	79.76	356.75	325.66	849.16
	3	14	106	110.18	439.07	285.82	1064.60
	4	8	23	25.84	427.25	357.87	920.35
	5	13	89	90.38	374.08	329.41	1110.66
	6	30	212	204.46	451.17	328.12	961.69
	7	26	170	165.56	448.85	318.08	944.20
3	1	12	68	67.83	415.92	301.70	814.32
	2	9	56	55.64	622.22	324.06	801.16
	3	18	149	145.26	625.33	320.70	796.51
	4	6	38	40.01	488.83	289.66	941.90
	5	18	129	128.76	561.94	348.57	964.44
	6	8	110	91.69	571.25	319.47	690.82
	7	17	103	96.20	497.88	332.90	785.57

First we set

$$\tilde{N}(t) = \sum_{i=1}^{n} c_i(t); \tag{21}$$

where n is the total number of cells in the whole time interval and $c_i(t)$ is defined in (4) and (5) for dividing and dying cells respectively. After, we "smooth" further by applying formula (10); that is,

$$N(t) = \frac{1}{2k_t + 1} \sum_{i=t-k_t}^{t+k_t} \tilde{N}(i). \tag{22}$$

Here we set $k_t \in \mathbb{Z}$ as the largest integer in the interval $[0, \delta(t)]$, where $\delta(t)$ is defined in (11). Note that, for the following we have used $\Delta = 300$ although our results are robust to changes in this value. We choose to keep Δ slightly under τ_{div} (see Table 1) so that at the time of cell division the formula (11) uses information about the next and previous generations and, at the same time, provides quite a large spectrum to vary cell death times from their averaged value τ_{death}.

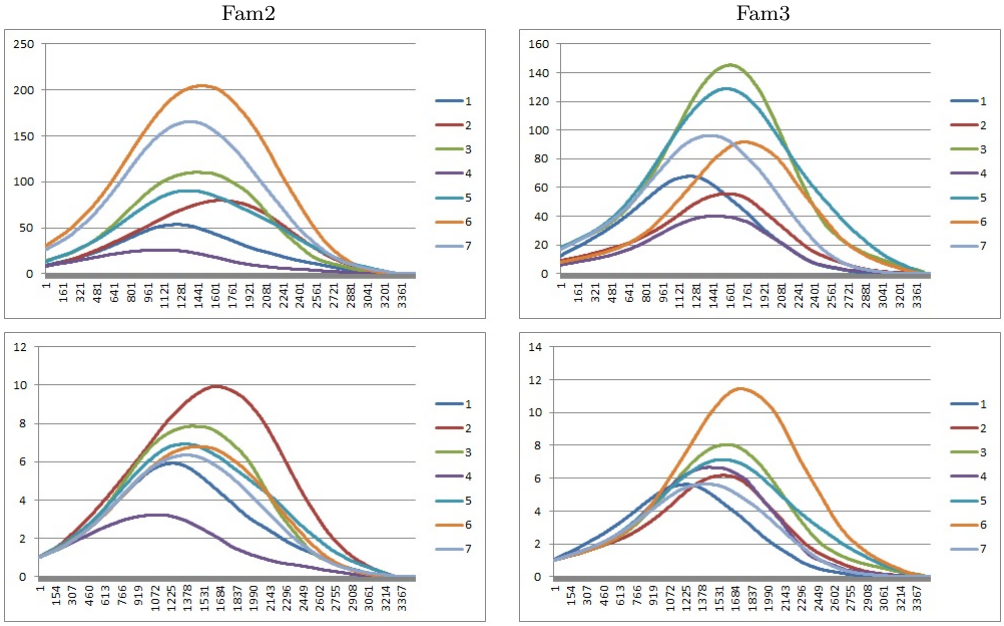
Table 2 shows the number of cells $N(t)$ for each well. Normalized graphs; that is, the graph of the function $N(t)/N(0)$, where $N(0) = X^0$ is the number of founder cells, are also plotted in order to have some idea about the rate of increase per one founder cell.

It can be observed that in Well 2 (Fam2) and Well 6 (Fam3) the increase (from 1 cell to 10 at the peak) is much higher than the others; in contrast, in Well 4 (Fam2) the growth is slow (from 1 cell to 3 at the peak).

We will use the notation $\mathbf{N}$ for the function $N(t),\ t = 1, ..., 3500$.

As mentioned before there is a tight relation between parameters X^0, D^{div} and $N^{max} = \max\{N(t),\ t = 1, ..., 3500\}$ in the form (9); that is, $N^{max} = \lambda D^{div} + X^0$.

TABLE 2. Plots of live cell population $N(t)$ and normalized population $N(t)/N(0)$ for wells 1-7 in Fam2 and Fam3.



We estimate the coefficient λ as best fit for all 14 wells in Table 1:

$$\lambda^* = \arg\min_{\lambda} \sum_{i=1}^{14} (N_i^{max} - \lambda D_i^{div} - X_i^0)^2 \tag{23}$$

This provides $\lambda^* = 0.83$. This means that the relation (9) has the form

$$N^{max} = \lambda^* D^{div} + X^0, \text{ with } \lambda^* = 0.83. \tag{24}$$

Note that according to Table 1, the value $\lambda = (N^{max} - X^0)/D^{div}$ over all 14 wells is relatively stable always being in the interval [0.75, 0.91].

3.2. **Data Fitting Problem.**

In section 2.1, by taking fixed times to divide and to death (τ_{div}^1, τ_{div} and τ_{death}), a procedure is described to determine cell population $\mathbf{x} = x(t)$ given any control sequence $\mathbf{u}$. Since this procedure applies fixed values for all parameters, it is interesting to check if the solutions $\mathbf{x}$ produced in this way could fit accurately to the actual cell population $\mathbf{N}$.

Both functions $x(t)$ and $N(t)$ are "smoothed" in the same way, so we can consider the distance between them in order to measure their closeness. The distance will be defined as

$$dist(\mathbf{x}, \mathbf{N}) = \sum_{i=1,,\ldots,n_{grid}} \rho((i, x(i)), \mathbf{N}); \tag{25}$$

where $\rho((i,x(i)),\mathbf{N})$ is the distance from point $(i,x(i))$ to the closest point on the graph of function $N(t)$ defined by

$$\rho((i,x(i)),\mathbf{N}) = \min\{\left[(i-j)^2 + (x(i_1) - N(j_1))^2\right],\ j = i-\tilde{n},\cdots,i+\tilde{n}\}.$$

Here $i_1 = (i-1)t_{step}+1,\ \ j_1 = (j-1)t_{step}+1,\ \ t_{step} = 3500/n_{grid}$ and

$$\tilde{n} = \min\{\max\{i-k_t,1\},\max\{i+k_t,n_{grid}\}\}.$$

In the calculations below we set $n_{grid} = 100$ and $k_t = 5$. In this case $t_{step} = 35$.

The aim is to find a sequence $\mathbf{u} = (u_1,\ldots,u_8)$, such that the corresponding state function $x(t)$ provides the best approximation to the actual population $N(t)$ in terms of distance (25) defined above.

Since $m = 8$, this is an optimization problem for minimizing objective function (25) over the 8 variables $\mathbf{u} = (u_1,\ldots,u_8)$ satisfying (2).

Data Fitting Problem (DF): Given an actual cell dynamics $\mathbf{N} = N(t)$ and a total number of divisions - D^{div} :

$$\text{Minimize : } dist(\mathbf{x},\mathbf{N}) \tag{26}$$

$$\text{Subject to : } \quad \mathbf{x} = \mathcal{F}(\mathbf{u}) \qquad \mathbf{u} = (u_1,\ldots,u_7); \tag{27}$$

$$0 \le u_8 \le ... \le u_1 \le 1; \tag{28}$$

$$2^{k-1}N(0)\prod_{j=1}^{k} u_j \in \mathbb{Z}, \quad \forall k; \tag{29}$$

$$TD(u) = D^{div}. \tag{30}$$

$$x(t) \in [0, N^{max}(1+\varepsilon)], \quad \forall t \in [0,T]. \tag{31}$$

Here $N^{max} = \max\{N(t),\ t = 1,...,3500\}$. The optimal control and the corresponding solution to the data fitting problem will be denoted by $\mathbf{u}^f$ and $\mathbf{x}^f$, respectively.

Note that, in (31) we use the constraint $x(t) \in [0, N^{max}(1+\varepsilon)]$, instead of $x(t) \in [0, N^{max}]$ in (20), where the small number $\varepsilon > 0$ is a relaxation parameter.

Numerical experiments are provided below (Tables 3, 4 and 5) along with the results of the optimal control problem that is described next. The values τ^1_{div}, τ_{div} and τ_{death} for each well in Fam2 and Fam3 are taken from Table 1. Note that in all cases data fitting provides quite "accurate" approximation to the actual curves.

3.3. **Optimal Control Problem.**

Considering a particular data/well, we can formulate an optimal control problem in the form of (15)-(20) by using average values for τ^1_{div}, τ_{div}, τ_{death}, N^{max} and the total number of divisions D^{div}.

Optimal Control Problem (OC): Given the initial number of (founder) cells $N(0)$ and two (external) parameters: the number of total divisions (D^{div}) and the maximum population level (N^{max}) :

$$\text{Maximize : } f(\mathbf{u}) = A \cdot x^{max}(\mathbf{x}) - B \cdot T^{max}(\mathbf{x}) + C \cdot i^{max}(\mathbf{u}); \tag{32}$$

$$\text{Subject to : } \quad \mathbf{x} = \mathcal{F}(\mathbf{u}), \qquad \mathbf{u} = (u_1, \ldots, u_8); \tag{33}$$

$$0 \le u_8 \le ... \le u_2 \le u_1 \le 1; \tag{34}$$

$$2^{k-1} N(0) \prod_{j=1}^{k} u_j \in \mathbb{Z}, \qquad \forall k; \tag{35}$$

$$TD(\mathbf{u}) = D^{div}; \tag{36}$$

$$x(t) \in [0, N^{max}(1+\varepsilon)], \qquad \forall t \in [0, T]. \tag{37}$$

Here the total number of divisions $TD(u)$ is defined by (14) and and $i^{max}(\mathbf{u}) \in \{1, 2, ..., 8\}$ is the largest index with positive value $u_{i^{max}} > 0$. Moreover

$$x^{max}(\mathbf{x}) = \max\{x(t),\ t = 1, ..., 3500\},$$

and

$$x(t) = x^{max}(\mathbf{x}) \text{ for } t = T^{max}(\mathbf{x}), \text{ and } x(t) < x^{max}(\mathbf{x}) \text{ for } t < T^{max}(\mathbf{x}).$$

As in the data fitting problem, the values of the parameters for each well will be taken from Table 1. Parameters A, B and C are chosen taking into account the values for Objectives 1-3. In all numerical experiments we set: $A = 10^2$ and $B = 1$. Moreover, we set $C = 10^4$ for Fam2 and $C = 10^2$ for Fam3.

The optimal control and corresponding solution will be denoted by $\mathbf{u}^c$ and $\mathbf{x}^c$, respectively.

Now the main question is weather the solution $\mathbf{x}^c$ to the above optimal control problem can approximate the actual curve $\mathbf{N}$ "accurately" enough. Note that the theoretically best result is when the optimal control solution coincides with the best data fit; that is, when $\mathbf{x}^c = \mathbf{x}^f$.

3.4. **Solving problems (26)-(31) and (32)-(37).**

Problems (26)-(31) and (32)-(37) are hugely difficult global optimization problems. The search space for each variable u_i, $i \in \{1, ..., 8\}$, is discrete, where the feasible (discrete) values depend on the previous components - u_j, $j = 1, ..., i-1$, as well as on the initial number of cells - $N(0)$.

On the other hand the constraints imposed on variables u_i, shrinks the feasible set. Eventually, it turns out that, for a small number of founder cells (say $N(0) \le 30$) and a relatively small number of total divisions (say $D^{div} \le 300$) all the feasible combinations of variables can be considered in a reasonable time frame. Table 6 in Section 3.6 lists the total number of feasible solutions in each well.

Therefore, problems OC and DF can be solved by considering all possible combinations of variables, not involving any optimization methods. This allows us to have global optimal solutions in all experiments below.

TABLE 3. Fam2: Global optimal solutions obtained for Data Fitting problem and Optimal Control problem, denoted by $\mathbf{u}^f$ and $\mathbf{u}^c$, respectively. In each generation, corresponding number of divisions, denoted by $\mathbf{k}^f$ and $\mathbf{k}^c$, are also provided for both problems. In calculations ε is set to be 0.01 for all wells except well 4 (with the extremely small number of divisions) for which $\varepsilon = 0.05$. In all cases the 8-th component of $\mathbf{u}^f$ and $\mathbf{u}^c$ is zero; so they are not listed in the table. Vectors $\mathbf{k}^f$ and $\mathbf{k}^c$ stand for the number of actual divisions in each generation for both solutions.

Well	Solution							
1	$\mathbf{u}^f$	1	0.94	0.38	0.35	0.33	0.25	0.17
	$\mathbf{u}^c$	1	0.94	0.38	0.35	0.33	0.25	0.17
	$\mathbf{k}^f$	9	17	13	9	6	3	1
	$\mathbf{k}^c$	9	17	13	9	6	3	1
2	$\mathbf{u}^f$	1	1	0.63	0.56	0.39	0.25	0.11
	$\mathbf{u}^c$	1	1	0.75	0.40	0.39	0.37	0.14
	$\mathbf{k}^f$	8	16	20	23	18	9	2
	$\mathbf{k}^c$	8	16	24	19	15	11	3
3	$\mathbf{u}^f$	1	0.89	0.74	0.24	0.22	0.19	0.17
	$\mathbf{u}^c$	1	1	0.57	0.28	0.25	0.22	0.13
	$\mathbf{k}^f$	14	25	37	18	8	3	1
	$\mathbf{k}^c$	14	28	32	18	9	4	1
4	$\mathbf{u}^f$	1	0.5	0.25	0.25	0.25	0	0
	$\mathbf{u}^c$	1	0.5	0.25	0.25	0.25	0	0
	$\mathbf{k}^f$	8	8	4	2	1	0	0
	$\mathbf{k}^c$	8	8	4	2	1	0	0
5	$\mathbf{u}^f$	1	1	0.42	0.32	0.29	0.25	0.25
	$\mathbf{u}^c$	1	0.88	0.48	0.43	0.21	0.19	0.17
	$\mathbf{k}^f$	13	26	22	14	8	4	2
	$\mathbf{k}^c$	13	23	22	19	8	3	1
6	$\mathbf{u}^f$	1	0.97	0.44	0.40	0.34	0.07	0
	$\mathbf{u}^c$	1	0.95	0.46	0.40	0.27	0.13	0.08
	$\mathbf{k}^f$	30	58	51	41	28	4	0
	$\mathbf{k}^c$	30	57	53	42	23	6	1
7	$\mathbf{u}^f$	1	0.94	0.42	0.35	0.29	0.21	0.07
	$\mathbf{u}^c$	1	0.96	0.39	0.38	0.27	0.22	0.14
	$\mathbf{k}^f$	26	49	41	29	17	7	1
	$\mathbf{k}^c$	26	50	39	30	16	7	2

Clearly, this method will not work for a large number of founder cells. For example, for $N(0) = 50$ and $D^{div} = 300$ the number of all combinations satisfying constraints (28)-(30) is already too large to perform all the required calculations within a reasonable time.

3.5. **Data fitting and optimal control solutions.**

The results of data fitting define the best possible approximations (to the actual curves $\mathbf{N}$) that could be achieved by the class of functions $\mathbf{x}$ generated by operator $\mathcal{F}$ in (12).

In Tables 3 and 4 the results are summarized for all wells in Fam2 and Fam3, providing global optimal solutions obtained for both the data fitting problem (26)-(31), denoted by $\mathbf{u}^f$, and the optimal control problem (32)-(37), denoted by $\mathbf{u}^c$. The number of actual divisions in each generation is also provided for both solutions; they denoted by $\mathbf{k}^f$ and $\mathbf{k}^c$.

TABLE 4. Fam3: Global optimal solutions obtained for Problem 1 (Data Fitting) and Problem 2 (Optimal Control), denoted by $\mathbf{u}^f$ and $\mathbf{u}^c$, respectively. In each generation, corresponding number of divisions, denoted by $\mathbf{k}^f$ and $\mathbf{k}^c$, are also provided for both problems. In calculations ε is set to be 0.01 for all wells. Vectors $\mathbf{k}^f$ and $\mathbf{k}^c$ stand for the number of actual divisions in each generation for both solutions.

Well	Solution							
1	$\mathbf{u}^f$	1	0.92	0.39	0.29	0.25	0.20	0
	$\mathbf{u}^c$	1	0.88	0.40	0.38	0.15	0.13	0
	$\mathbf{k}^f$	12	22	17	10	5	2	0
	$\mathbf{k}^c$	12	21	17	13	4	1	0
2	$\mathbf{u}^f$	1	0.89	0.56	0.25	0.17	0.17	0
	$\mathbf{u}^c$	1	0.94	0.53	0.22	0.19	0.17	0
	$\mathbf{k}^f$	9	16	18	9	3	1	0
	$\mathbf{k}^c$	9	17	18	8	3	1	0
3	$\mathbf{u}^f$	1	1	0.71	0.25	0.24	0.21	0.20
	$\mathbf{u}^c$	1	0.89	0.86	0.25	0.22	0.17	0.13
	$\mathbf{k}^f$	18	36	51	25	12	5	2
	$\mathbf{k}^c$	18	32	55	27	12	4	1
4	$\mathbf{u}^f$	1	0.83	0.60	0.29	0.21	0	0
	$\mathbf{u}^c$	1	1	0.42	0.30	0.25	0.17	0
	$\mathbf{k}^f$	6	10	12	7	3	0	0
	$\mathbf{k}^c$	6	12	10	6	3	1	0
5	$\mathbf{u}^f$	1	1	0.54	0.26	0.25	0.25	0.10
	$\mathbf{u}^c$	1	1	0.57	0.26	0.21	0.17	0.17
	$\mathbf{k}^f$	18	36	39	20	10	5	1
	$\mathbf{k}^c$	18	36	41	21	9	3	1
6	$\mathbf{u}^f$	1	0.94	0.93	0.55	0.26	0.25	0.25
	$\mathbf{u}^c$	1	1	1	0.41	0.33	0.29	0.05
	$\mathbf{k}^f$	8	15	28	31	16	8	4
	$\mathbf{k}^c$	8	16	32	26	17	10	1
7	$\mathbf{u}^f$	1	0.82	0.52	0.34	0.18	0.14	0
	$\mathbf{u}^c$	1	0.97	0.36	0.33	0.28	0.17	0.17
	$\mathbf{k}^f$	17	28	29	20	7	2	0
	$\mathbf{k}^c$	17	33	24	16	9	3	1

We observe that in Wells 1, 4 of Fam2, the optimal control $\mathbf{u}^c$ coincides with the best data fitting solution $\mathbf{u}^f$; that is, the optimal control problem provides the best possible approximation to the actual curve $\mathbf{N}$ in the class of functions defined by the operator $\mathcal{F}$ in (12). For the remaining wells we observe quite good approximations.

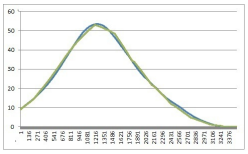
It should be noted that to measure how these approximations are accurate is not a simple task. Some related issues will be discussed in the next sections. We present in Table 5 the curves $\mathbf{N}$ (actual), $\mathbf{x}^f$ (best data fit) and $\mathbf{x}^c$ (optimal control solution) so that the quality of approximations can be visualized.

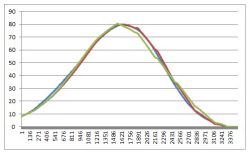
The main conclusion we can draw from these results is that for all wells in Fam2 and Fam3, the optimal trajectory to the control problem (32)-(37) (and its shape) provides very close approximation to the actually observed data. This is the case in the presence of quite different shapes and rates of growth per one founder cell (see Table 2). This might hint at an underlying optimization principle in operation for cell dynamics.

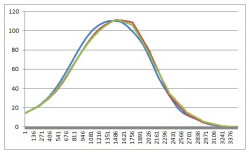
3.6. **Quality of optimal solutions in terms of data fitting.**

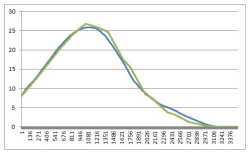
The results for wells 1 and 4 in Table 3 show that optimal control problem OC provides the best possible solution in terms of data fitting. For other wells there is a gap between the optimal control and data fitting solutions. However, as it can be seen from figures in Table 5, in all cases both optimal control and best-fit solutions are quite "good" in approximating the actual curves of cell population. In Table 6 we present distances between the actual curves, and the optimal control and data fitting solutions; that is, the values $\Delta^f = dist(\mathbf{x}^f, \mathbf{N})$ and $\Delta^c = dist(\mathbf{x}^c, \mathbf{N})$, which show that in most cases the best fit curve does not give a significantly better approximation to the true result than the optimal control curve does.

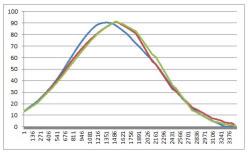
TABLE 5. Graphs for: $N(t)$ - actual curve (blue), $\mathbf{x}^f$- best data fit (red) and $\mathbf{x}^c$- optimal control solution (green)

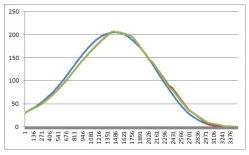

Well 1 (Fam2)

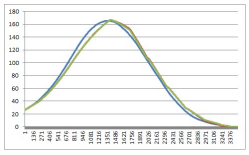

Well 2 (Fam2)

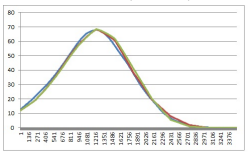

Well 3 (Fam2)

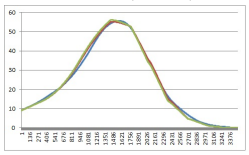

Well 4 (Fam2)

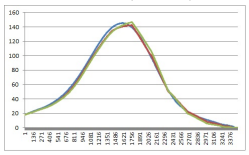

Well 5 (Fam2)

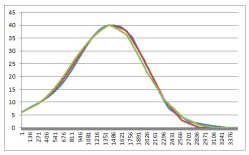

Well 6 (Fam2)

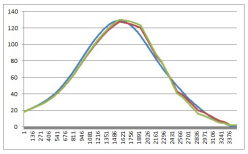

Well 7 (Fam2)

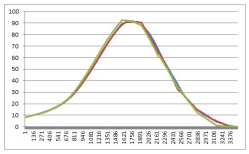

Well 1 (Fam3)

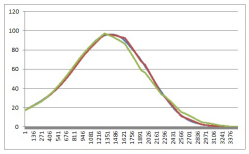

Well 2 (Fam3)


Well 3 (Fam3)


Well 4 (Fam3)


Well 5 (Fam3)


Well 6 (Fam3)


Well 7 (Fam3)

There is also another interesting issue worthy of discussion. Perhaps it is unreasonable to expect very "accurate" optimal control solutions in such a living system. Particularly considering how difficult it is to find the best (optimal) solution. If there is only a small number of possible/feasible scenarios or there is a large subset of "good" solutions, then there is more chance that the system will be able to realize one of these good solutions.

The results presented in Table 6 below addresses these issues. Together with $\Delta^f = dist(\mathbf{x}^f, \mathbf{N})$ and $\Delta^c = dist(\mathbf{x}^c, \mathbf{N})$ mentioned above, the following numbers are provided.

(1) M^{tot} - total number of feasible solutions satisfying (33)-(37);

TABLE 6. The quality of optimal control solutions in terms of actual cell population. $\mathbf{N}$ is the actual (averaged) curve of cell population, $\mathbf{x}^c$ and $\mathbf{x}^f$ are the optimal control and best data fitting solutions and $\Delta^f = dist(\mathbf{x}^f, \mathbf{N})$, $\Delta^c = dist(\mathbf{x}^c, \mathbf{N})$. The notation M^{tot} stands for the total number of feasible solutions satisfying (33)-(37); Top/per - Top is the rank of the optimal control solution $\mathbf{x}^c$ in terms of data fitting out of all possible feasible solutions, per is its percent equivalent; $Av(M^*)$ - Av is the average value of $dist(\mathbf{x}, \mathbf{N})$ out of top M^* best (closest to the actual curve) solutions $\mathbf{x}$.

Well (Fam)	M^{tot}	Top/per	Δ^f	Δ^c	$Av(M^*)$
1 (2)	974	1/0.00	61.34	61.34	1482.59 (900)
2 (2)	5262	382/7.26	74.00	190.52	1602.41 (5000)
3 (2)	25067	5/0.02	238.03	264.20	1390.99 (10000)
4 (2)	36	1/0.00	53.83	53.83	344.97 (10)
5 (2)	9362	50/0.65	468.81	584.99	1916.93 (5000)
6 (2)	1192598	1510/0.13	358.97	464.23	1135.80 (20000)
7 (2)	332469	33/0.00	369.06	396.09	780.94 (20000)
1 (3)	2777	9/0.32	92.88	117.22	1949.66 (2200)
2 (3)	1153	23/2.00	51.86	83.40	1527.53 (1000)
3 (3)	162138	80/0.05	244.15	342.31	948.08 (10000)
4 (3)	183	20/11	43.43	71.31	761.84 (170)
5 (3)	78691	141/0.18	252.8	373.58	1055.11 (10000)
6 (3)	13972	472/3.38	44.25	149.09	1330.66 (10000)
7 (3)	22254	1636/7.35	45.01	286.09	1476.54 (10000)

(2) Top/per - Top is the rank of the optimal control solution $\mathbf{x}^c$ in terms of data fitting out of all possible feasible solutions, per is its percent equivalent;
(3) $Av(M^*)$ - Av is the average value of $dist(\mathbf{x}, \mathbf{N})$ out of top M^* best (closest to the actual curve) solutions $\mathbf{x}$.

The results show that in majority of wells optimal control solutions are "good" solutions. Note that these results are obtained for fixed values of parameters A, B and C. In fact, some of results would be even better if these parameters were tuned appropriately (see next section). This appears to further support the idea that live cell population dynamics is not "arbitrary" and might obey some sort of optimality criteria.

3.7. **Some results on artificial wells.**

The results of previous sections show that each particular cell population (well) has its own growth pattern. We can call them "normal" patterns that could be quite accurately predicted by the optimal control model introduced. In this section we consider different artificially generated cell populations (artificial wells) and check if the dynamics of cell population in these wells show similar patterns.

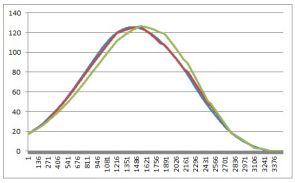
We generate artificial wells in three different ways. First we combine several "small" wells in one "larger" well. It could be expected that such an artificially created well may "display" a similar growth pattern applicable to a particular population. Secondly, we generate wells where the "sense" of a whole population is no longer true. For this aim, we divide a particular well (Well 6 in Table 7) into two subsets by separating highly "productive"

founder cells. At last, we generate a new artificial well by selecting several founder cells from different wells. A well generated in this way will have no sense of a "united population".

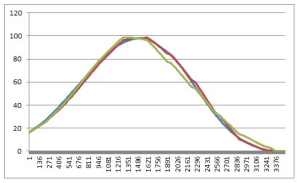
Several selected results are presented in Table 7. We can observe that in the case of a combination of different wells, cell dynamics can be well predicted by our optimal control problem. However, when considering a subset of wells or a set of some founder cells from different wells, optimal curves do not approximate related curves accurately enough.

The conclusion we can make is that, each "united cell population" has it own growth dynamics that is "normal" in terms of the optimal control model introduced. An artificial well generated by violating the sense of "unity", may not produce such a "normal" dynamics.

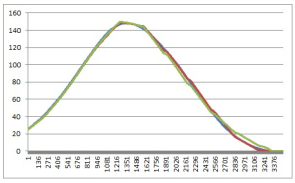
TABLE 7. Results for several combined and artificial wells: $N(t)$ - actual curve (blue), $\mathbf{x}^f$- best data fit (red) and $\mathbf{x}^c$- optimal control solution (green). *Subset 1* includes 5 most divided founder cells (ID's: 45, 54, 57, 64, 69) in Well 6 Fam2, *Subset 2* include the remaining founder cells in that well, *Mixed well* is generated from 6 arbitrarily chosen founder cells (ID's: 1,6,7,9,17,41) from different wells in Fam2.



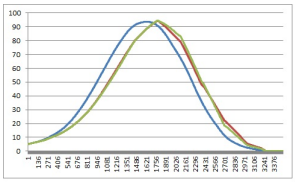
Combined: wells 1,2 (Fam2)



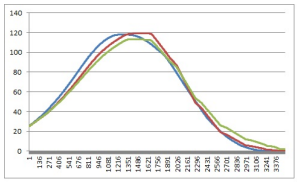
Combined: wells 1,4 (Fam2)



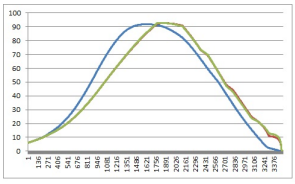
Combined: wells 1,2,4 (Fam2)



Subset 1: Well 6 (Fam2)



Subset 2: Well 6 (Fam2)



Mixed well from Fam2

4. The impact of the 3 objectives defined in Section 2.3 on optimal control solutions

This section examines the influence of parameters A, B and C on optimal control solutions. Note that in this paper the problem of searching for the best values for these parameters, in order to obtain the best possible approximation to the actual curve in each particular case (well), is not considered. In the previous section the first two parameters are fixed at the level $A = 10^2$, $B = 1$ and parameter C is set to be 10^4 for $Fam2$ and 10^2 for $Fam3$.

However, for example, if we set $A = 10^4$ (instead of $A = 10^2$) for Well 2 Fam2 (Table 6), the solution (in terms of number of divisions $\mathbf{k}$) would be $\mathbf{k}^c = (8, 16, 20, 22, 21, 7, 2)$. This is quite close to the best data fitting result $\mathbf{k}^f$ in Table 3. Corresponding value for Top in

TABLE 8. The impact of parameters A, B and C on optimal control solutions. $\Delta^c_{ABC} = dist(\mathbf{x}^c_{ABC}, \mathbf{N})$ where $\mathbf{N}$ is the actual (averaged) curve of cell population and $\mathbf{x}^c_{ABC}$ is the optimal control solution. *Change* stands for the difference in distances from $\mathbf{N}$; that is, $Change = dist(\mathbf{x}^c_{ABC}, \mathbf{N}) - dist(\mathbf{x}^c, \mathbf{N})$, where $\mathbf{x}^c$ is the optimal control solution obtained for $A = 10^2$, $B = 1$; and $C = 10^4$ for Fam2 and $C = 10^2$ for Fam3 in Table 6.

Well(Fam)	A	B	C	Δ^c_{ABC}	*Change*	Well(Fam)	A	B	C	Δ^c_{ABC}	*Change*
1 (2)	0	1	10^4	4173.29	4111.95	1 (3)	0	1	10^2	5701.82	5584.60
1 (2)	10^2	0	10^4	61.34	0.00	1 (3)	10^2	0	10^2	117.22	0.00
1 (2)	10^2	1	0	418.47	357.13	1 (3)	10^2	1	0	127.71	10.49
2 (2)	0	1	10^4	3715.53	3525.01	2 (3)	0	1	10^2	7016.25	6932.85
2 (2)	10^2	0	10^4	93.06	-97.46	2 (3)	10^2	0	10^2	83.40	0.00
2 (2)	10^2	1	0	158.87	-31.65	2 (3)	10^2	1	0	83.40	0.00
3 (2)	0	1	10^4	10597.50	10333.30	3 (3)	0	1	10^2	18732.83	18390.52
3 (2)	10^2	0	10^4	264.20	0.00	3 (3)	10^2	0	10^2	342.31	0.00
3 (2)	10^2	1	0	559.71	295.51	3 (3)	10^2	1	0	1052.01	709.7
4 (2)	0	1	10^4	53.83	0.00	4 (3)	0	1	10^2	2876.83	2805.52
4 (2)	10^2	0	10^4	53.83	0.00	4 (3)	10^2	0	10^2	71.31	0.00
4 (2)	10^2	1	0	53.83	0.00	4 (3)	10^2	1	0	58.32	-12.99
5 (2)	0	1	10^4	11342.25	10757.26	5 (3)	0	1	10^2	42072.51	41698.93
5 (2)	10^2	0	10^4	584.99	0.00	5 (3)	10^2	0	10^2	373.58	0.00
5 (2)	10^2	1	0	819.04	234.05	5 (3)	10^2	1	0	503.4	129.82
6 (2)	0	1	10^4	90864.51	90400.28	6 (3)	0	1	10^2	1822.66	1673.57
6 (2)	10^2	0	10^4	534.19	69.96	6 (3)	10^2	0	10^2	159.26	10.17
6 (2)	10^2	1	0	464.30	0.07	6 (3)	10^2	1	0	175.77	26.68
7 (2)	0	1	10^4	91367.64	90971.55	7 (3)	0	1	10^2	27143.95	26857.86
7 (2)	10^2	0	10^4	864.15	468.06	7 (3)	10^2	0	10^2	286.09	0.00
7 (2)	10^2	1	0	396.09	0.00	7 (3)	10^2	1	0	52.07	-234.02

Table 6 would be $Top = 23$ which is equivalent to having this solution in the top of 0.44 percent of all feasible solutions in terms of closeness to the actual cell population.

There are two main questions that we are interested in this section.

Q1. The importance and necessity of each of these objectives.

Q2. The "stability" of optimal control solutions with respect to perturbations of A, B and C.

These issues are directly relevant to the main conclusion of this paper stating that the cell division process follows some optimality criteria over the period of existence. While the first question justifies the role of each objective, the second issue ensures some simplicity in finding optimal control solution that would be quite "difficult" be realized otherwise.

First we discuss the question **Q1.** To illustrate the importance of each objective some numerical experiments are provided in Table 8 by elimination one objective in each case. It can be observed that the role of Parameter A and therefore the role of the first objective - *Maximize the amplitude,* is highly significant; in all cases without this objective, the violation from the original curve is quite high.

The role of parameter C; that is, the third objective (*Maximize the life span of cell population*) is also important in the majority of cases. For Well 2 (Fam2) and Well 4 (Fam3) the optimal control solution (almost) does not change if we set $C = 0$; however there is a noticeable improvement for Well 7 (Fam3) in this case.

The role of parameter B; that is, the second objective (*Reach the maximal amplitude in a shortest time*) is the most interesting one. In majority of cases it did not have any effect on the results. In one case (Well 2 Fam2) setting $B = 0$ even provided better results, however in 3 cases setting a positive value for B, that is, involving the second objective was helpful in terms of approximating actual curves. The effect of this objective was specially important for large populations (Well 6,7 Fam2) where we can assume that the competition between cells is tight. For small populations, like Well 1, 4 (Fam2) and Wells 2,4 (Fam3), we observe that optimal control solutions obtained for $B = 1$ and $B = 0$ coincide. This is partly due to the fact that the number of feasible solutions/choices is relatively small and therefore the optimal solution is expected to be "robust".

The results in Table 8 show that, the first objective, that is *maximizing the amplitude of cell population* within allowed boundaries, is the most dominant criteria in the cell division process. Compared with this, the second and third objectives seem secondary; however their presence is crucial in a majority of cases to obtain tight approximations to actual curves and to produce different shapes and rates of increase. It should also be noted that, the last two objectives may need to be adjusted to obtain the best outcome in different wells. This fact itself can be used to characterize different populations (wells), and possibly the existing environment.

Now we discuss the second question **Q2.** The "stability" of optimal control solutions with respect to perturbations of A, B and C mainly depends on the number of feasible solutions. Table 6 lists the number of feasible solutions in each well. This number is small for small populations where there are only a "few" choices. For large wells this number is large but it is still finite. This means that in all cases some small perturbations of A, B and C should not affect the optimal control solution. This has been confirmed with many numerical experiments that are not provided here. We provide only the results for a "small" Well 1(Fam2) where the number of feasible solutions is 974.

The optimal control $\mathbf{u}^c = (1, 0.94, 0.38, 0.35, 0.33, 0.25, 0.17, 0)$ obtained for Well 1(Fam2) by setting $A = 10^2$, $B = 1$ and $C = 10^4$ coincides with the best data fitting solution $\mathbf{u}^c = \mathbf{u}^f$ (Table 3). Let $\mathbf{x}^c$ be the corresponding cell population dynamics. The parameters in the definition of functional (32) are:

- the amplitude of $\mathbf{x}^c$ is - $x^{max}(\mathbf{x}^c) = 52.995$;
- the time this amplitude is reached - $T^{max}(\mathbf{x}^c) = 1207$;
- the last division (generation) - $i^{max}(\mathbf{u}^c) = 7$.

We fix two out of three parameters A, B, C and change (decrease or increase) the third one until the optimal control changes. The aim is to find the range of that parameter for which the optimal control is stable.

For parameter A this interval is $[10^{-12}, 10^4]$; for $A < 10^{-13}$ or $A > 10^5$ the optimal controls obtained are different from $\mathbf{u}^c$. For example, for $A = 10^5$ we have the above parameters for corresponding optimal solution as: $x^{max} = 53.975$; $T^{max} = 1526$; $i^{max} = 6$. Simply saying, a small increase in the amplitude (0.98 = 53.975 - 52.995) is achieved one generation (319 = 1526-1207) later with the shorter life span (6 generations instead of 7).

This means that, given $B = 1$ and $C = 10^4$, almost any positive number (not grater than 10^4) would be acceptable for parameter A to obtain the above control $\mathbf{u}^c$ by solving problem (32)-(37).

For parameter B this interval is $[0, 10^6]$; when B is greater than 10^7 the optimal control changes. Again this means that, given $A = 10^2$, $C = 10^4$, almost any (less that 10^6) number including 0 would suit to obtain $\mathbf{u}^c$ as a solution to problem (32)-(37).

An analogous interval for parameter C is the range $C \geq 71$. The optimal control obtained for $C = 70$ is different from $\mathbf{u}^c$; in this case, we have the above parameters for corresponding optimal solution as: $x^{max} = 53.701$; $T^{max} = 1207$; $i^{max} = 6$. Taking any value $C \geq 71$ is enough to increase the life span to 7 generations (meanwhile decreasing slightly the amplitude) and to achieve the best possible approximation.

The above analysis for Well 1(Fam2) shows that the best fitting solution $\mathbf{u}^f$ can be obtained as an optimal control in problem (32)-(37) with a quite wide range of values for parameters A, B and C. Note that this control is a unique optimal solution out of 974 feasible solutions. Clearly, for large populations one can expect smaller intervals for these parameters are needed to keep the optimal control unchanged.

5. Optimal control solutions for different values of D^{div} and N^{max}

In this section we consider the behavior of optimal solutions when one external parameter is fixed and the other changes. This is about examining the influence of one of these parameters on the shape of optimal curves.

In order to be able to compare these optimal solutions with the results from the previous section obtained for Fam2 and Fam3, we consider similar values for the parameters of the optimal control problem. In the calculation we set the parameters as follows:

- the time interval is $[0, 3500]$, each unit corresponds to two minutes;
- the number of initial (founder) cells is $X^0 = 10$;
- the time to divide is $\tau^1_{div} = \tau_{div} = 320$;
- the time to death is $\tau_{death} = 830$.

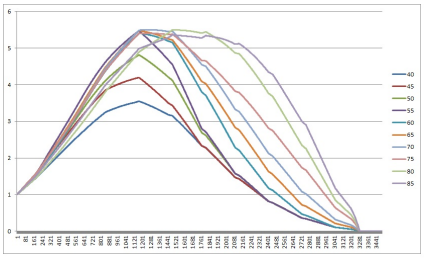
Moreover, the values of A, B and C are set as in the previous section: $A = 10^2, B = 1$ and $C = 10^4$. We will consider the following cases.

Case 1: The amplitude is fixed at $N^{max} = 55$ and the number of total divisions D^{div} changes.

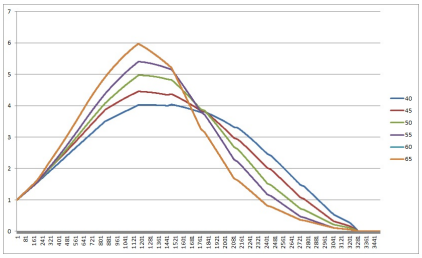
Optimal curves related to the values $D^{div} = 40, 45, ..., 80, 85$ are presented in Table 9(1). We observe that when D^{div} is less than 55 the corresponding optimal curve does not reach the maximum allowed level $N^{max} = 55$. The peak of optimal curves increase till $D^{div} = 55$ then shifts to the right on time axis. Interestingly, the optimal control values u_i starting from $i = 3$ division stabilize at a level 0.5 that corresponds to a "steady" division rate; see Table 10.

Case 2. The number of total divisions is fixed at $D^{div} = 60$ and amplitude N^{max} changes.

TABLE 9. Optimal control solutions for different values of divisions $D^{div} = 40, 45, ..., 80, 85$, with a fixed amplitude $N^{max} = 55$; and for different amplitudes $N^{max} = 40, 45, 50, 55, 60, 65$, with a fixed number of divisions $D^{div} = 60$. The other parameters are: the number of initial (founder) cells is $X^0 = 10$; time to divide is $\tau^1_{div} = \tau_{div} = 320$; time to die is $\tau_{death} = 830$.



1. $N^{max} = 55$, $D^{div} = 40, 45, ..., 80, 85$



2. $D^{div} = 60$, $N^{max} = 40, 45, 50, 55, 60, 65$

TABLE 10. Optimal controls for different vales $D^{div} = 40, 45, ..., 80, 85$ at a fixed level $N^{max} = 55$. The other parameters are $X^0 = 10$, $\tau_{div} = 320$ and $\tau_{death} = 830$.

D^{div}	u_1	u_2	u_3	u_4	u_5	u_6	u_7
60	0.90	0.89	0.47	0.33	0.30	0.25	0.17
65	1.00	0.80	0.44	0.39	0.36	0.25	0.25
70	1.00	0.75	0.50	0.40	0.38	0.33	0.25
75	1.00	0.75	0.47	0.43	0.42	0.40	0.38
80	0.80	0.75	0.58	0.50	0.46	0.42	0.36
85	0.90	0.72	0.50	0.50	0.50	0.50	0.42

The results of optimal curves related to the values $N^{max} = 40, 45, 50, 55, 60, 65$ are presented in Table 9(2). As might be expected, the peak of optimal curves moves up (and slightly to the left) together with the increase in N^{max} till $N^{max} \leq 65$. After that point, the increase in N^{max} does not affect optimal solution.

6. CONCLUSIONS

This paper examines cell population dynamics corresponding to a small number of founder cells. An optimal control problem is introduced to describe the cell division process. It enables the estimation of live cell population dynamics by optimal solutions to that problem. This approach does not involve any statistical distribution functions commonly used in the literature to estimate cell population.

Therefore, the proposed approach assumes the existence of some intracellular control mechanism for cell division that follows some optimality criteria over a certain time period. In other words, a decision to divide or to die at any time is made by considering all the further dynamics of cell population and environmental limitations. This approach gives a different insight to the cell division process and allows some interesting conclusions. For example, it helps to identify/differetiate external and internal factors of cell division processes.

We mention another issue related to the degree of freedom for possible curves describing cell dynamics. We do not include in this discussion the number of founder cells. Moreover, it can be observed that the average times to divide (τ_{div}) and to death (τ_{death}), as well as the parameters of the optimal control problem (32)-(37) (that is, A, B, C) are relatively stable.

Consequently, there are only two main parameters that define cell dynamics - the number of total divisions (D^{div}), as a strength of activation of founder cells (potential to divide); and the maximum population level (N^{max}) that may represent in some sense the capacity of the present environment. In the case of no such limitations; that is, in "free growth" environment the number of total divisions (D^{div}) remains the only major factor that defines the shape/dynamics of live cell population.

Possible improvements in generating similar optimal control models, as well as different possible conclusions that could be made out of such models are interesting topics for future investigations. Another important development would be considering times to divide (τ_{div}) and to die (τ_{death}) as control variables rather than fixed parameters derived from data.

References

[1] H.T. Banks and B.F. Fitzpatrick. Estimation of growth rate distributions in size-structured population models. *Quart. Appl. Math.,* 49 (1991), 215–235.

[2] H.T. Banks, J.L. Davis, S.L. Ernstberger, S. Hu, E. Artimovich, A.K. Dhar and C.L. Browdy. A comparison of probabilistic and stochastic differential equations in modeling growth uncertainty and variability. *Journal of Biological Dynamics,* 3 (2009), 130–148.

[3] H.T. Banks and S. Hu, Nonlinear stochastic Markov processes and modeling uncertainty in populations. *Mathematical Bioscience and Engineering,* 9 (2012), 1–25.

[4] H.T. Banks and W.C. Thompson. Mathematical Models of Dividing Cell Populations: Application to CFSE Data. *Journal on Mathematical Modelling of Natural Phenomena,* to appear.

[5] de Boer R, Perelson AS: Estimating division and death rates from CFSE data. *J Comp Appl Math* 2005, 184:140-164.

[6] De Boer RJ, Ganusov VV, Milutinovic D, Hodgkin PD, Perelson AS: Estimating lymphocyte division and death rates from CFSE data. *Bull Math Biol* 2006, 68(5): 1011–1031.

[7] Deenick EK, Gett AV, Hodgkin PD: Stochastic model of T cell proliferation: A calculus revealing IL-2 regulation of precursor frequencies, cell cycle Time, and survival. *J Immunol* 2003, 170(10): 4963–4972.

[8] K. R. Duffy, V. G. Subramanian, On the impact of correlation between collaterally consanguineous cells on lymphocyte population dynamics. *J. Math. Biol.* **59**, 255 (2009).

[9] 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.

[10] A. V. Gett, P. D. Hodgkin, A cellular calculus for signal integration by T cells. *Nat. Immunol.* **1**, 239 (2000).

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

[12] Hawkins, E.D., Markham, J.F., McGuinness, L.P. and Hodgkin, P.D. A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13457–13462 (2009).

[13] M. D. Johnston, C. M. Edwards, W. F. Bodmer, P. K. Maini, and S. Jonathan Chapman. *Proc. Natl. Acad. Sci. U.S.A.* **104**, no 10, 40084013 (2007).

[14] J. F. Markham, C. J. Wellard, E. D. Hawkins, K. R. Duffy, P. D. Hodgkin, A minimum of two distinct heritable factors are required to explain correlation structures in proliferating lymphocytes. *J. R. Soc. Interface* **7**, 1049 (2010).

[15] V. G. Subramanian, K. R. Duffy, M. L. Turner, P. D. Hodgkin, Determining the expected variability of immune responses using the cyton model. *J. Math. Biol.* **56**, 861 (2008).

[16] C. Wellard, J. Markham, E. D. Hawkins, P. D. Hodgkin, The effect of correlations on the population dynamics of lymphocytes. *J. Theor. Biol.* **264**, 443 (2010).
[17] Yates A, Chan C, Strid J, Moon S, Callard R, George A and Stark J: Reconstruction of cell population dynamics using CFSE. *BMC Bioinformatics* 2007, **8**: 196 doi:10.1186/1471-2105-8-196.
[18] J.A. Smith and L. Martin, Do Cells Cycle?, *Proc. Natl. Acad. Sci,* 70 (1973), 1263–1267.
[19] O. Hyrien and R. Chen and M.S. Zand, An age-dependent branching process model for the analysis of cfse-labelling experiments, *Biol. Direct,* 5 (2010)
[20] A. Zilman and V.V. Ganusov and A.S. Perelson, Stochastic models of lymphocyte proliferation and death, *PLoS one,* 5 (2010)
[21] K. León and J. Faro and J. Carneiro, A general framework to model generation structure in a population of asynchronously dividing cells, *J. Theor. Biol.* 2004, **229** 455–476

M. A. Mammadov
University of Ballarat, Victoria 3353, Australia
National ICT Australia, VRL, Melbourne, Victoria 3353, Australia
E-mail address: m.mammadov@ballarat.edu.au, musa.mammadov@nicta.com.au

C. Wellard
Walter and Eliza Hall Institute of Medical Research, Parkville, Australia
E-mail address: wellard@wehi.edu.au

R. J. Evans
University of Melbourne, Department of Electrical and Electronic Engineering, Parkville, Australia
National ICT Australia, VRL, Melbourne, Victoria 3353, Australia
E-mail address: rob.evans@nicta.com.au