Derivation of continuum models from an agent-based cancer model: optimization and sensitivity analysis

Dimitrios Voulgarelis,[*,†,‡,§] Ajoy Velayudhan,[¶] and Frank Smith[‡]

†*Centre for Mathematics, Physics and Engineering in the Life Sciences and Experimental Biology, UCL, Physics Building, Gower Pl, London WC1E 6BT*

‡*Department of Mathematics, Gower Street, London, WC1E 6BT , UK*

¶*Department of Biochemical Engineering, Bernard Katz Building, Gordon Street, London WC1H 0AH, United Kingdom , UK*

§*Corresponding author.*

E-mail: dimitrios.voulgarelis.14@ucl.ac.uk

Abstract

Background: Agent-based models provide a formidable tool for exploring complex and emergent behaviour of biological systems as well as accurate results but with the drawback of needing a lot of computational power and time for subsequent analysis. On the other hand, equation-based models can more easily be used for complex analysis in a much shorter timescale. **Methods & Objective**: This paper formulates an ordinary differential equations and stochastic differential equations model to capture the behaviour of an existing agent-based model of tumour cell reprogramming and applies it to optimization of possible treatment as well as dosage sensitivity analysis. **Results**: For certain values of the parameter space a close match between the equation-

based and agent-based models is achieved. The need for division of labour between the two approaches is explored.

Keywords

cancer; agent-based; ODE; ABM; SDE; optimization

1 Introduction

1.1 Motivation

The research concerns a particular cancer-related clinical setting. The basic research aim was to formulate two different types of continuum models, ordinary differential equations (ODE) and stochastic differential equations (SDE), that could match the qualitative and quantitative behaviour of an agent-based model (ABM) in that setting. Furthermore, connecting these models parametrization-wise is not usually trivial but we will show the parametrization we have done allows us to move from the agent-based to the continuum view for optimization and sensitivity analysis. When a clinically relevant setting is considered then each model has advantages and disadvantages. Specifically ABMs provide more realistic results which can be used by both the researcher and the clinician but systematic and rigorous analysis of the system simulated by the ABM is very time-consuming. Continuum models by contrast allow for complex analysis and exploration of the underlying system with the trade-off of realism and accuracy. Matching between the two types of models would mean that one could use the continuum model for analysis, such as optimization and sensitivity analysis. These procedures could be extremely time-consuming if performed directly in ABMs for two reasons. The first is that ABMs are in a way in-silico experiments which have to be run a great number of times for each set of parameter values in order to be able to draw a sensible conclusion. Moreover and specifically in the case of cell biology scenarios, realistic numbers of cells in a system could reach the order of millions. Such numbers would mean that huge computational power is needed to run simulations, in contrast to continuum equations which usually can be handled numerically very fast. On the other hand, ABMs provide a more realistic representation of the real system and can prove very valuable to clinicians. Hence the main idea here is to move from an ABM to a continuum model for analysis and understanding of the system and dynamics and then move back to the ABM in order to use the conclusions of the analysis to obtain realistic data.

Creating a rigorous connection between the two aforementioned types of models for a generic system is a difficult task, depending on the system under consideration. Therefore, we approach our main idea on a case-by-case basis. To this end we reproduce the agent-based model found in the recent paper by P.M. Biava et al.[1] based on a recently developed tumour paradigm which is supported experimentally. Attempts to connect agent-based models to equation-based models in order for one of them to act as a complementary model to the other have been done previously in a number of different papers.[2–4]

1.2 Cancer cell model

Several experimental studies exploring the interaction between a tumour and the embryonic micro-environment have shown a delay or even stopping of the proliferation of different human cancer lines when development and more specifically organogenesis is at work[5] . In 2002 the same group that developed the ABM model[1] showed that factors taken from specific developmental stages of the zebrafish embryo decreased the proliferation rate of several human cancer lines significantly.[6] In addition to in-vitro, an in-vivo study revealed a slowdown of the Lewis lung cancer carcinoma in mice when administered with developmental factors.[7] Moreover, recent experimental results have demonstrated that embryonic and cancer cells share some of the molecular signal and pathways.

The above has led to the development of a new tumour paradigm in which cancer is considered a developmental deviation of normal undifferentiated cells.[5] Cancer cells are viewed as undifferentiated cells that are stuck in the proliferative stage between two differentiation stages. As a result the same molecular factors that differentiate cells during embryonic development could help the cancer differentiate and even become a healthy cell. A less differentiated stem cell differs from a more differentiated one in the fact that a greater number of genes are expressed in the latter and in reality all the developmental factors have to work in a complete network to induce the expression of many genes.[5] In this paradigm there are five malignancy stages, from more to less malignant, in accordance with the five steps of

a stem cell, i.e. pluripotent (most malignant), multipotent, oligopotent, differentiating and differentiated (healthy).

1.3 Aims

The rest of our study is divided into four main sections and some into further subsections. **Section 2** is **material and methods**. Initially we present the ABM we developed which is very similar in nature to the model developed in.[1] We build on this model to explore dosage optimization and sensitivity analysis. We explain how the model works, its agents as well as the parameter values being derived from empirical and experimental data. Following the development of the ABM we formulate two new models, an ordinary ODE and a SDE to capture the dynamics of the ABM. We justify the form of our models and explain the fitting to the ABM procedure. Then, we extend the ODE to two more complex and realistic forms that will be used for further analysis. **Section 3** presents the **results and discussion**. Here we show how well the two continuum models match the ABM and present the rest of the analysis which is related to a hypothetical therapy based on the cancer cell theory. Namely, we conduct an optimization of the dosage of molecular factors as well as an analysis of a patient missing some dosages. **Section 4** is the **further discussion**, where we summarize our work and findings. Finally, in **section 5**, the **conclusion**, we raise possible issues and areas left unexplored as well as further work towards making a more realistic agent-based and continuum model.

2 Material and methods

2.1 Agent-based model of cancer cell differentiation

2.1.1 Form of the agent-based model

Purpose. The purpose of the ABM is to simulate the temporal evolution of the total population of five cancer-cells differentiation stages, going from most malignant to healthy cells, induced by interaction with four distinct types of molecular factors, found in embryonic development, each characteristic to the differentiation stage. This model could potentially advise a clinician on the progress of the cancer treated by administration of these factors.

Grid cells, temporal and spatial scale. The world is a rectangular grid of discrete patches and it is also toroidal, meaning that both the horizontal and vertical edges are wrapped. Spatial units are abstract. Since we do not model volume-exclusion effects there can be many cells in the same patch; that can be interpreted as patches being large compared to individual cells. The time units (time-steps) are minutes since the therapy time-frame is days and hence smaller times would require a huge simulation time. Simulation occurs through the passing of discrete time-steps.

Cancer cells. As mentioned before there are five cancer stages named after the respective stem-cell stages in order of malignancy: pluripotent, multipotent, oligopotent, differentiating and differentiated (healthy) cells. At each time step cells might grow according to a probability representing their growth rate and interact (or not), depending upon fitness, with factors in the same patch as the cells. To account for the fact that cancer cells are relatively static and diffuse much slower than factors (1000 times slower[1]), each cancer cell moves by a unit one (jumping to one of the 8 neighbour patches) every 1000 time-steps.

Molecular factors. There are four types of factors each representing a different differentiation stage and affecting the corresponding cancer differentiation stage. There are no factors for healthy cells. At every time-step factors move a distance of one (one of the eight neighbour patches) and interact with cells according to their fitness, which is characteristic

of the stage.

Dosage. Doses of factors are administered every 8 hours (480 minutes). Every dose includes 2000 factors of each type which are randomly distributed in the ABM world. This dosage strategy is chosen as a simplistic way of killing all malignant cancer cells.

Growth. At each time step each cell has a probability of dividing into two daughters. The probability is given by the growth rate per minute of that differentiation stage of cancer derived from empirical data. Daughters are cancer cells of the same stage. We are not interested in the dynamics of the healthy population and hence the growth rate is kept at zero.

Interaction. If a factor and a cell are in the same patch they might interact with some probability called fitness. Fitness is a $4x4$ matrix since factors could potentially interact with all different malignant cancer stages but both in the original ABM and here we consider a diagonal matrix where each type of factor only interacts with cancer cells of the same differentiation-stage type. If a factor interacts with a cell then that factor dies and the cell moves to the next differentiation stage.

Initialization. The model is initialized with 1000 randomly distributed pluripotent cancer cells and zero for the other four populations and 2000 factors of every type.

2.1.2 Parameters

Fitness values were taken from the original ABM paper[1] where they were derived by fitting the simulation to experimental data from.[6] We derived growth rates for each stage by fitting exponentials to the proliferation curves of kidney adenocarcinoma cells found in[6] The growth rates used in the original ABM might have been found by fitting to different cancer proliferation curves from the same paper as there were many cancer lines and there was no mention as to the specific ones used. The coefficients of the exponentials correspond to the growth rates for each cancer stage and are used for all three model types as we will see. Tables 1,2 show the values for both the fitness and growth rates. Despite not knowing some

specifics of the ABM in[1] like the world dimensions, movement profiles of agents and exact growth rates fitted and used we tried to reproduce the two plots of Fig3. shown in[1] for the two dosing strategies mentioned in that paper. The plots can be seen in the appendix.

Table 1: Fitness of molecular factors.

	Stage I factors	Stage II factors	Stage III factors	Stage IV factors
Pluripotent	4%	0	0	0
Multipotent	0	5%	0	0
Oligopotent	0	0	2.5%	0
Differentiating	0	0	0	1%

Table 2: Growth rates of cancer stages.

Stage	Pluripotent	Multipotent	Oligopotent	Differentiating	Healthy
Growth rate (min^{-1})	$3.9 * 10^{-4}$	$3.3 * 10^{-4}$	$3.1 * 10^{-4}$	$2.8 * 10^{-4}$	0

Figure 1 shows a stopped frame of a random run of the ABM after a few thousand time-steps with the parameter values given in the aforementioned tables.

Figure 1: Single frame of a random run of the ABM with the parameter values found in table 1 and 2.

2.2 Deterministic continuum model

Despite the fact that the ABM provides a more realistic representation of the real system, performing complex analysis and optimization can prove extremely time-consuming as results may need to be obtained by running the model several times with many parameter values. Formulating and solving numerically a continuum model that matches the ABM allows that analysis to be done much faster and more rigorously. We do not need the continuum model to match every aspect of the ABM, only the parameter/behaviour space that is of interest.

As mentioned before, in[1] an ABM was formulated to represent the biological system without any continuum analog. Here, we describe an ordinary differential equations system (ODE) to capture the main dynamics of our ABM. The ODE has the form of coupled equations, identical to chemical reaction kinetics with the addition of growth. This means that the interaction of cells and factors is proportional to their concentration. This choice is justified by the form of the ABM where no exclusion phenomena are considered and cells and factors are positioned randomly in the whole space, giving an almost homogeneous mix.

In this ODE system each variable represents one of the five cell and four factor populations represented by the initial letter of the type of cell or factor, e.g. P for pluripotent cells and F_p for the factors. The model equations are given below:

$$\dot{P}(t) = g_p P(t) - \beta P(t) F_p(t) f_1, \tag{1a}$$

$$\dot{M}(t) = g_m M(t) - \beta M(t) F_m(t) f_2 + \beta P(t) F_p(t) f_1, \tag{1b}$$

$$\dot{O}(t) = g_o O(t) - \beta O(t) F_o(t) f_3 + \beta M(t) F_m(t) f_2, \tag{1c}$$

$$\dot{D}(t) = g_d D(t) - \beta D(t) F_d(t) f_4 + \beta O(t) F_o(t) f_3, \tag{1d}$$

$$\dot{H}(t) = g_h H(t) + \beta D(t) F_d(t) f_4, \tag{1e}$$

$$\dot{F}_p(t) = Ds - \beta P(t) F_p(t) f_1, \tag{2a}$$

$$\dot{F_m}(t) = Ds - \beta M(t) F_m(t) f_2, \tag{2b}$$

$$\dot{F_o}(t) = Ds - \beta O(t) F_o(t) f_3, \tag{2c}$$

$$\dot{F_d}(t) = Ds - \beta D(t) F_d(t) f_4. \tag{2d}$$

Here g_i represents the growth rates of different cancer cell stages, where $i = \{p, m, o, d, h\}$, and f_j is the fitness, where $j = \{1, 2, 3, 4\}$. The values for these parameters are exactly the same as the ABM and are given by **Tables 1** and **Table 2** respectively. Ds is the dosage. In the ODE the dosage is modelled by constant value, instead of discrete steps. The value is given by the size multiplied by the frequency. Hence,

$$Ds = Size * f = 2000/480 = 4.1666 factors * min^{-1}.$$

Moreover, it should be noted that the interaction term, i.e. $\beta[CancerCells][Factors]$ is the same in both the cancer cell equations and the respective factor equations. This can be understood again by taking as an example chemical reactions. A typical reaction has the form:

$$a[A] + b[B] \rightarrow c[C]$$

Where a, b, c are called the stoichiometric coefficients. The reaction rate is given by $r = -\frac{1}{a}\frac{d[A]}{dt} = -\frac{1}{b}\frac{d[B]}{dt} = \frac{1}{a}\frac{d[C]}{dt}$. In our case since one cell interacts with one factor to produce one different type of cell we have $a = b = c = 1$ hence the reaction rate, given by the reaction term, is the same for both cells and factors.

In the ODE system there is an extra parameter, β, which is the rate of interaction per cell per minute. This parameter is the same for all interaction terms since it is dependent on external conditions such as the size of the world or the diffusive speed of the cells and factors. It is the most important parameter since it is the one that needs to be calibrated in order for the two models to match. Its calibration is conducted by the use of the NonLinearModelFit

function found in Mathematica. To produce the data set the ABM was run 100 times with the same parameter values and a total time of 8000 minutes which allowed for all malignant cells to die out. Then, the 100 runs were averaged to give mean population evolution curves for each cell population.

Following calibration, the ODE model will work as a stepping stone for more complex models that will be used for our analysis. It is worth mentioning that β is the only fittable parameter as the growth parameters for all our models are found from fitting exponential to cancer proliferation curves and hence there is no doubt as to which model parameter corresponds to which experimental parameter.

2.3 Stochastic continuum model

In addition to the deterministic ODE model used to describe the agent-based model, a stochastic differential equations (SDE) model was also formulated. Here we compare it to the deterministic model as well as the ABM.

The SDE model is of the Langevin kind and its derivation as well as justification can be found in.[8] In Appendix A we include the basic theory behind the derivation as well as the case-specific values that are used to create our model. The equations for the cancer cell populations are:

$$\dot{P}(t) = g_p P - \beta P F_p f_1 + \sqrt{g_p P}\Gamma_1 - \sqrt{\beta P F_p f_1}\Gamma_{10}, \tag{3a}$$

$$\dot{M}(t) = g_m M - \beta M F_m f_2 + \beta P F_p f_1 + \sqrt{g_m M}\Gamma_2 + \sqrt{\beta P F_p f_1}\Gamma_{10} - \sqrt{\beta M F_m f_2}\Gamma_{11}, \tag{3b}$$

$$\dot{O}(t) = g_o O - \beta O F_o f_3 + \beta M F_m f_2 + \sqrt{g_o O}\Gamma_3 + \sqrt{\beta M F_m f_2}\Gamma_{11} - \sqrt{\beta O F_o f_3}\Gamma_{12}, \tag{3c}$$

$$\dot{D}(t) = g_d D - \beta D F_d f_4 + \beta O F_o f_3 + \sqrt{g_d D}\Gamma_4 + \sqrt{\beta O F_o f_3}\Gamma_{12} - \sqrt{\beta D F_d f_4}\Gamma_{13}, \tag{3d}$$

$$\dot{H}(t) = g_h H + \beta D F_d f_4 + \sqrt{g_h H}\Gamma_5 + \sqrt{\beta D F_d f_4}\Gamma_{13}. \tag{3e}$$

Here, both population variables (P, M, O, D, H) and Gaussian white (Γ_1-$Gamma_{13}$) noise

are time-dependent see Appendix A, e.g. P means $P(t)$.

The numerical simulation of the SDE model was conducted with Mathematica, using the Euler-Maruyama method found in.[9]

2.4 Extension of the ODE model

After calibration of our initial ODE system with the ABM, it can be used as a cornerstone to build more complex models that can be used for further analysis of the system of interest. Specifically, two new models were created, both being different from the first in the dosage term and in the fact that factors die out. The latter comes from the fact that an administered drug is metabolised by the organism and after the passing of a few hours its quantity drops exponentially. This provides a much more realistic system for clinically relevant analysis since otherwise a single dose can last for many days or even weeks until all factor molecules have interacted with the cancer. So in both of the following models the factors decay with a rate of 90% decrease in 8 hours. The ODE for the factor population becomes:

$$\dot{F}_i(t) = dosage - \beta F_i(t) C_i(t) - \gamma F_i(t)$$

where C is the respective cancer population and γ is the decay rate of factors due to metabolism with a value of $4.8 * 10^{-3} min^{-1}$.

2.4.1 Step doses

In the initial model the dosage is constant and all four doses are administered from the beginning of the simulation and are also given even after the extinction of the respective cancer stage. This scenario is unsuitable for clinical analysis; hence a new model was created where the doses were given in a step-like manner, meaning that the dosing of a factor type starts after the creation of the respective cancer type and stops immediately after extinction. This is achieved by multiplying the initial dose term, Ds, by an inverse tangent function as

this gives a smooth transition from zero to Ds. The function has the following form:

$$dosage = Ds * arctan(100 * C)\frac{2}{\pi}$$

where C represents a cancer cell population. When C goes to zero the dosage goes to zero and when C goes above zero the function increases rapidly to a steady value which is given by Ds. The coefficient of 100 is used to make the increase or decrease faster.

Now this model is used in order to explore how the time of complete cancer extinction is affected by the increase of the dosing and what is the best distribution of the total dosage for the four different doses. That is to provide insight as to the importance of each of the four factors. The time of cancer extinction is specified as the time when all numerical solutions for the four malignant cancer populations go below 1. As in the ABM the initial condition for the first factor population is 2000 but zero for the rest and 1000 for the pluripotent cancer cells and zero for the other populations.

2.4.2 δ function doses

In the second additional model doses are administered in a way similar to the ABM, in the form of an injection. Here, doses are given in full every 8 hours. To achieve this we used a sum of Dirac delta functions multiplied by the size of the dose (the total factor count). The dosage function is as follows:

$$dosage = Size * \sum_{i=1}^{n-1} \alpha_i \delta(t - i * 480).$$

Here, n is the total number of dosages depending on the number of days of the treatment and α_i is the intake coefficient which takes the value of zero or one depending on whether a particular dose was taken or missed by the patient. The first dose is taken at time zero and it is given as initial conditions for the factors-population ODE. That allows for a sensitivity analysis of missed doses, concerning how the time of cancer extinction is affected by a missed

dose as well as the importance of the dose timing.

3 Results and Discussion

3.1 Fitting of ABM and ODE and comparison with SDE

Using NonLinearModelFit we found that the β which gives the best match between the ABM and the initial ODE has the value $6.857 * 10^{-6} agent^{-1} min^{-1}$. Figure 2 shows the variation of the populations for 100 runs of the ABM model. We can see that there are no unexpected behaviours and that for these parameters values a mean approximation should work well. For subsequent analysis the value was rounded to $6.9 * 10^{-6} agent^{-1} min^{-1}$. The value as well as the basic statistics for β-value fitting are:

	Estimate	Standard Error	t-Statistic	P-Value
β	$6.85734 * 10^{-6}$	$1.44617 * 10^{-9}$	4741.74	$1.63450391131 * 10^{-55019}$

We can see the variation of the 100 ABM runs in Figure 2.

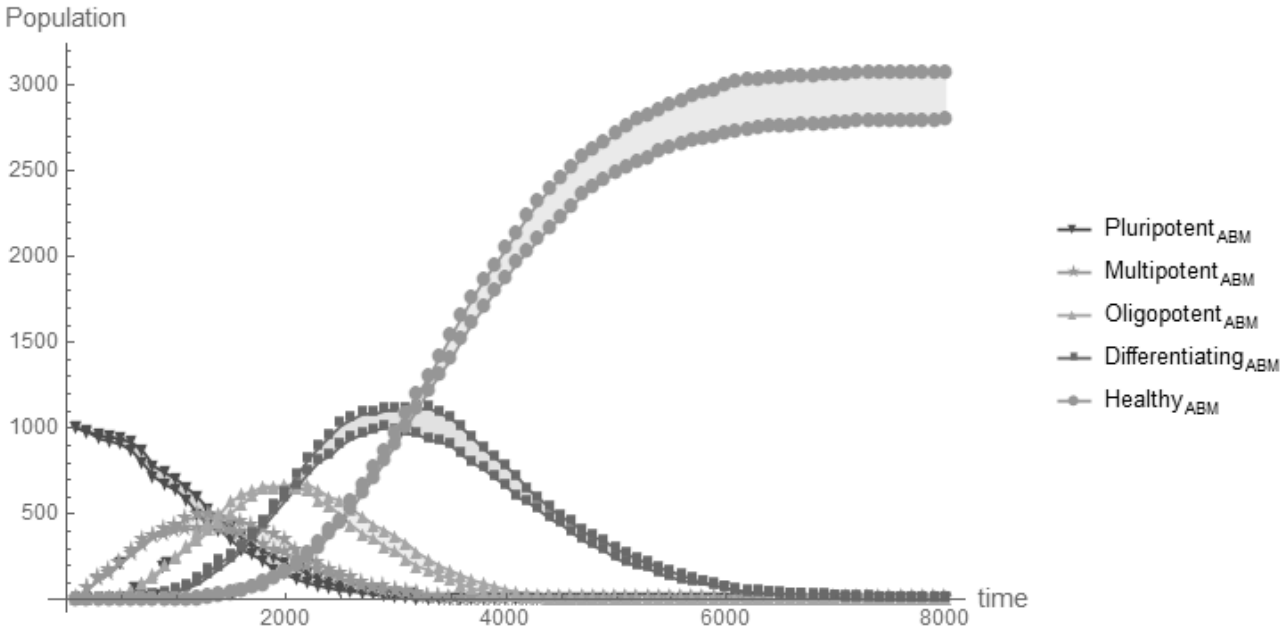

Figure 2: Variation of the 100 ABM runs for the parameter values found in Table 1.

Figure 3 demonstrates the comparison of the ODE and the mean ABM proliferation curves.

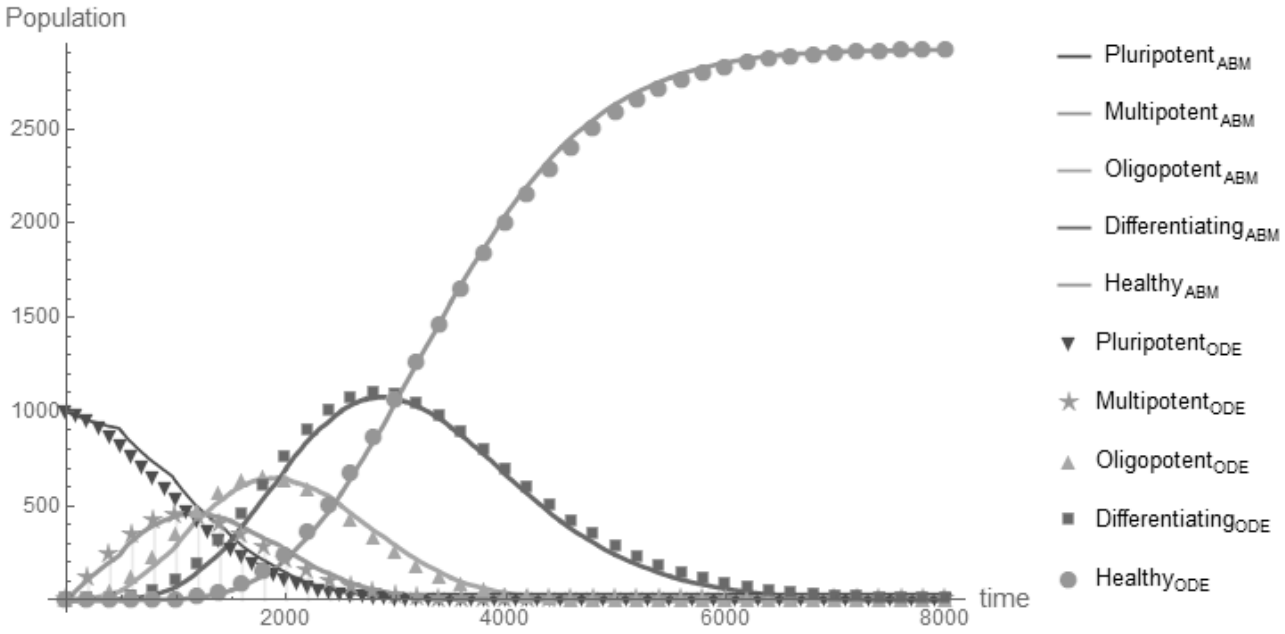

Figure 3: Comparison of the average of 100 runs of the ABM, with the parameter values found in Table 1 and 2, and of the ODE with the same values for the growth rates and dosage and a $\beta = 6.9 * 10^{-6} agent^{-1} min^{-1}$. The dashed lines are the populations of the ODE model and the full lines are for the ABM.

The match of the two models for the specific, experimentally obtained, parameter values is very close. There is both a qualitative and quantitative matching between the two which hints towards using the continuum model for further analysis by extending it as mentioned in the previous section.

Using the same β we plot the average of 50 SDE runs against the same 100-runs average of the ABM, as used in the ODE fitting; this gives us again a very good fitting between the two models. The behaviour of the SDE is almost indistinguishable from that of the ODE for these parameter values. Figure 4 shows the comparison between the SDE and ABM.

Finally, Figure 5 shows the range of variation for the five populations for these 50 runs.

The healthy cell population shows a larger variation but this can be accounted for by the variation of the differentiating cancer cell population especially close to the turning point.

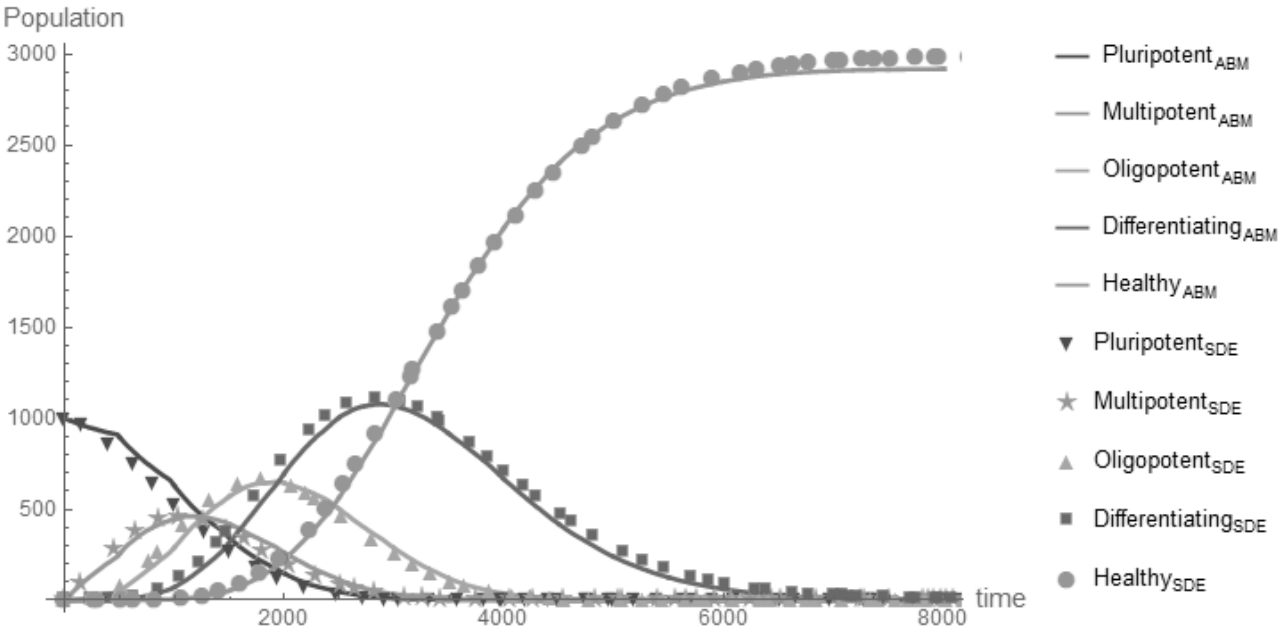

Figure 4: Comparison of the average of 100 runs of the ABM, with the parameter values found in Table 1 and 2, and of the SDE with the same values for the growth rates and dosage and a $\beta = 6.9 * 10^{-6} agent^{-1} min^{-1}$. The dashed lines are the populations of the SDE model and the full lines are for the ABM.

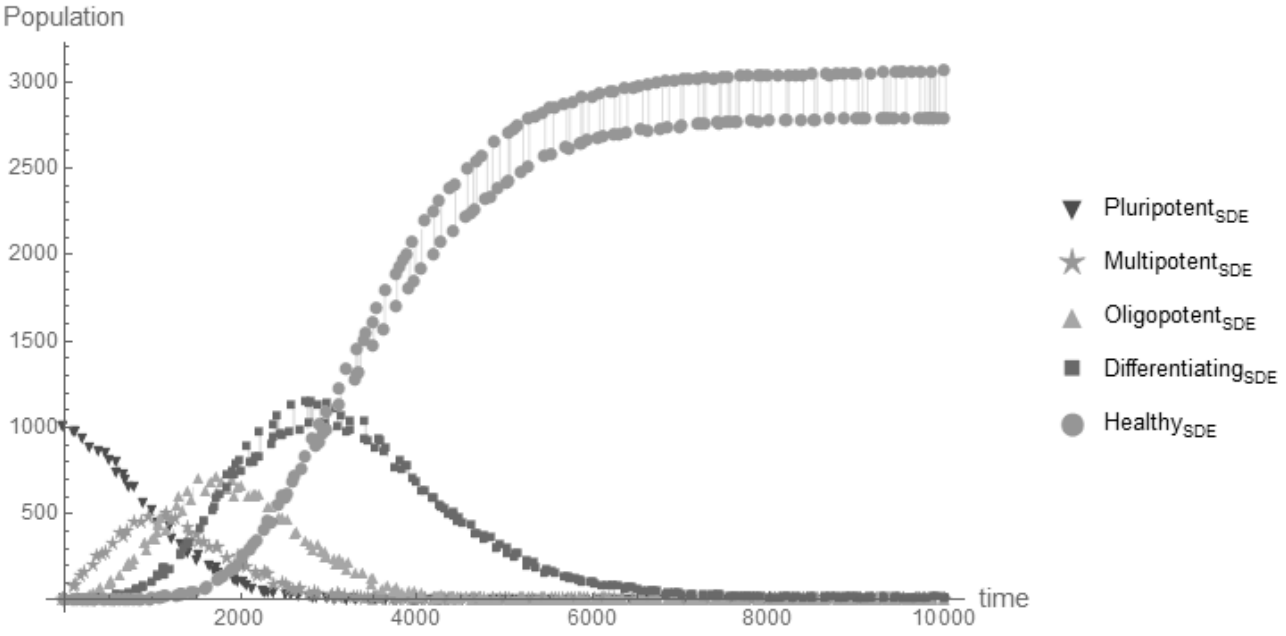

Figure 5: Range of variation of 50 runs for the five types of cells for the SDE model.

So small variation throughout the simulation can lead to significant differences in the total amount of cells (healthy population) but without any other consequences. Due to the small variations observed in both the ABM and the SDE for the parameter values used we conclude that the dynamics of the system can be well approximated by a mean field approach hence we will use the ODE for the subsequent analysis.

3.2 Dosage optimization

Using the first of the additional models (step-like dosage) we can optimize the treatment by administrating different doses for the four factors (scheme 1) and compare the results with the scheme where doses are the same (scheme 2).

3.2.1 Optimal distributions

To find the optimal distribution of the doses we first pick a total dosage, meaning the sum of the individual doses ($Ds = Ds1 + Ds2 + Ds3 + Ds4$), then we produce all the combinations of four integer numbers that add up to D and from these we pick the one that gives the shortest time of cancer extinction (ToCE). This procedure is repeated a number of times for other total dosage values. The results are summed up in Table 3 and are illustrated in Figure 5.

Table 3: Optimal distribution of doses for a fixed total dosage.

Total dosage	Stage I dose	Stage II dose	Stage III dose	Stage IV dose	ToCE (mins)
60	10	8	13	29	107280
70	11	9	16	34	64800
80	13	10	18	39	46368
90	14	12	20	44	36144
100	16	13	22	49	29520
110	18	14	24	54	25056
120	19	16	27	58	21600

Looking at the values we can see a pattern in how the dosage is distributed every time we add an extra 10. More specifically almost every time 5 goes to the last dosage, except

when going from 110 to 120 where the increase is by 4. So we have $40-50\%$ going to $Ds4$. We also have $20-30\%$ going to $Ds3$ and the rest $20-30\%$ distributed at $Ds1$ and $Ds2$. There was no case where we had 40% going to the first two doses. From Table 3 we can see that the last dose is the most important in affecting time of cancer extinction by far followed by the third dose, but then we see a turnover where the first dose is more important than the second. That order is evident in all total dosage values we selected.

To further explore this pattern of distribution we wanted to see whether it is present if we move to the second and third best distributions. Tables 4 and 5 show the second and third best combinations respectively along with the mins of cancer extinction for each.

Table 4: Second optimal distribution of doses for a fixed total dosage.

Total dosage	Stage I dose	Stage II dose	Stage III dose	Stage IV dose	ToCE (mins)
60	10	7	13	30	108144
70	12	9	15	34	64800
80	13	11	17	39	46656
90	15	12	20	43	36144
100	17	13	22	48	29664
110	18	15	24	53	25056
120	19	16	26	59	21600

Table 5: Third optimal distribution of doses for a fixed total dosage.

Total dosage	Stage I dose	Stage II dose	Stage III dose	Stage IV dose	ToCE (mins)
60	10	7	14	29	109440
70	11	9	15	35	64944
80	13	10	17	40	46656
90	15	11	20	44	36288
100	16	13	23	48	29664
110	18	14	25	53	25056
120	20	16	26	58	21600

The first point we notice is that the ToCE is very slightly different between optimal distributions, one to three. The difference is only significant in the smallest total dosage (60). Moreover, we can see again the same ordering as well as overall pattern. Again, the highest amount of dosing goes to $Ds4$ then $Ds3$ then $Ds1$ and finally $Ds2$. When extra

dosage is provided again the most of it goes to $Ds4$, this time $40-60\%$. $20-30\%$ goes to $Ds3$ and $10-20\%$ to $Ds1, Ds2$ individually. So, the overall pattern is preserved. That gives an indication indeed of how a clinician could distribute not only their initial total dosage but also any extra they administer.

To explore how changes in β might affect the pattern of optimal dosage we performed the same optimization for $\beta' = 2\beta$ and $\beta'' = \beta/2$. Other than changes in the time of cancer extinction and the total dosage needed, the patterns seemed to remain almost exactly the same with again the greatest dosage given on the last does followed by the third, first and second.

Table 6: Optimal distribution of doses for a fixed total dosage and $\beta' = 0.000014$.

Total dosage	Stage I dose	Stage II dose	Stage III dose	Stage IV dose	ToCE (mins)
60	10	8	13	29	21225
70	11	9	16	34	16708
80	12	11	18	39	13785

Table 7: Optimal distribution of doses for a fixed total dosage and $\beta'' = 0.00000345$.

Total dosage	Stage I dose	Stage II dose	Stage III dose	Stage IV dose	ToCE (mins)
140	23	18	31	68	63666
150	25	19	33	73	53496
160	12	11	18	39	13785

3.2.2 Comparison to Scheme 2

Furthermore, we compare the optimal distribution to a scheme where all four dose values are the same (total dosage divided by 4), for the same total dosage. These values are the square markers of Figure 6. There are only three points as for the total dosage values of 90, 80, 70 and 60 the cancer survived. In addition we can see a very big difference in the ToCE, pointing towards the efficiency of the optimal distribution.

Finally, in scheme 1 we can see that increasing the total dosage results in a plateau after some value, meaning that the further you increase the total dosage the less benefit you get.

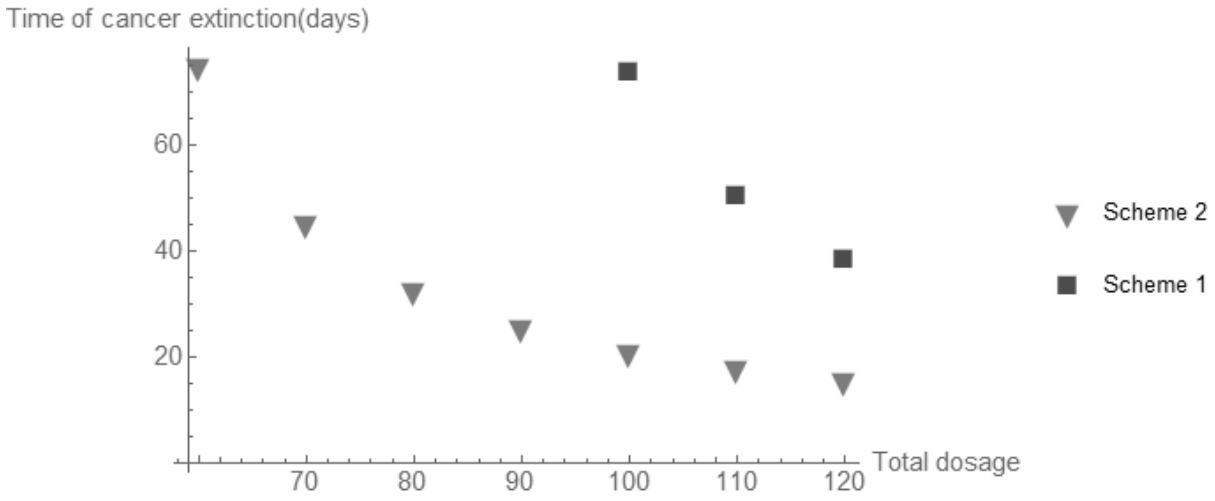

Figure 6: Total dosage versus time of cancer extinction for the optimal distribution of doses.

This needs to be taken into consideration for choosing the appropriate dosage by considering both the ToCE as well as cost of dosing or side effects.

3.3 Missing doses sensitivity analysis

Using the discrete dosing model we can explore what happens when a patient misses some doses and how the results are affected by the timing of those doses. To do that we first need to pick specific dosage size values for the four doses as well as the number of days for administration. We can make an informed decision using our previous analysis. Table 3 shows that for a total dosage/frequency ratio of 120 cancer dies out at 21613 mins or 15 days. To make sure that the cancer is dead a clinician would probably need to add one more day and not stop administration of drug at the exact time predicted here. Furthermore, this optimum result is achieved by distributing the dosage/frequency ratio in the four doses as $\{19, 16, 27, 58\}$. That gives size values of $\{9120, 7680, 12960, 27840\}$ for the four doses respectively with a fixed frequency of 8 hours. Again this number is rounded up to $\{10000, 8000, 13000, 28000\}$. The error between the rounded up and the exact values is less than 10% in all cases. That accuracy error is consistent with the measurement accuracy of the concentration of complex biologial drugs (e.g. monoclonal antibodies). Hence a clinicial

could potentialy increase the actual dosage to make sure that even with the error difference the doses remain close to the exact values.

In Figure 7 we see the results of missing one dose. Here 16 days correspond to 48 doses and we explore the effect of missing either of them. We observe that the later the missing dose is, the better it is for the patient but the differences are small and there is no danger of the cancer surviving. The continuous line is the ToCE when all doses are administered.

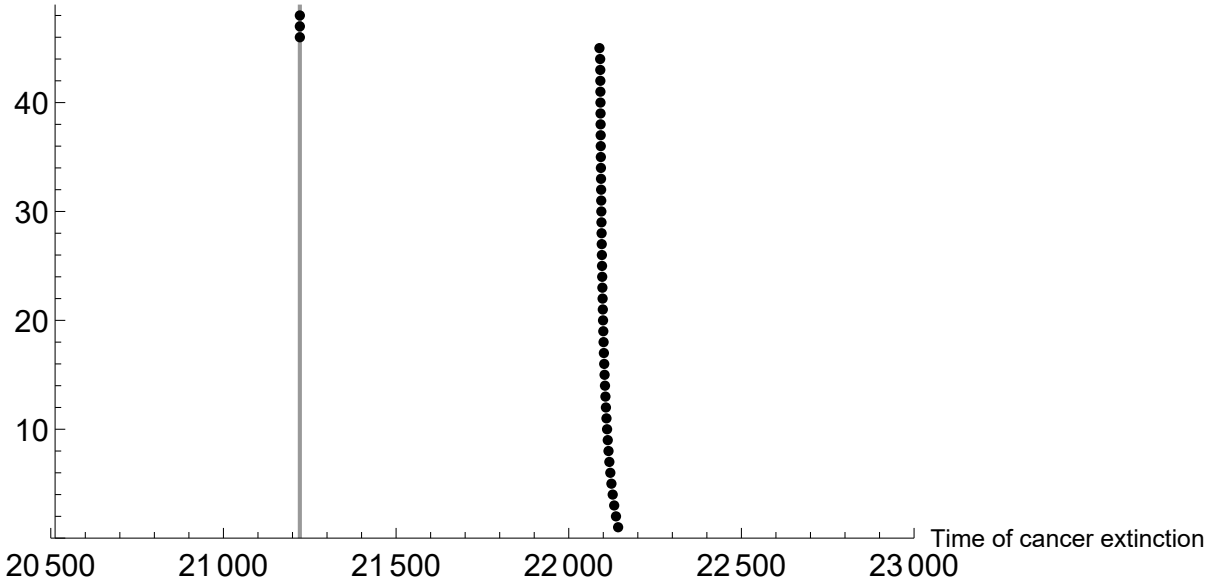

Figure 7: Time of cancer extinction with respect to the position of one missed dose. The red line is the ToCE when all 48 doses are taken normally.

Also interesting is the contour plot of two missed doses. In that case we see that there is a region where if two doses are lost then cancer survives. This is the white region of Figure 8 above the dotted line, which shows how the plot should normally be. Figure 9 is the complete plot whereas in Figure 8 we have zoomed to the interesting area. Here we still see that the best case scenario is when doses are missed close to the end of the treatment period, which is as expected. There is a large region where the position of the missed doses does not affect ToCE significantly (orange contour). But as we move missed doses one or two close to the end we can see sudden jumps to other contours of smaller ToCE. Finally,

the blank region above the dotted line shows all the possible combinations of the two missed doses. This sensitivity analysis can be potentially very useful in order to either warn the patient or possibly increase the duration of the treatment period such that there is no blank region at least for two doses. For example adding one more day (3 doses) gives us a contour plot showing no cancer-surviving region (Figure 9).

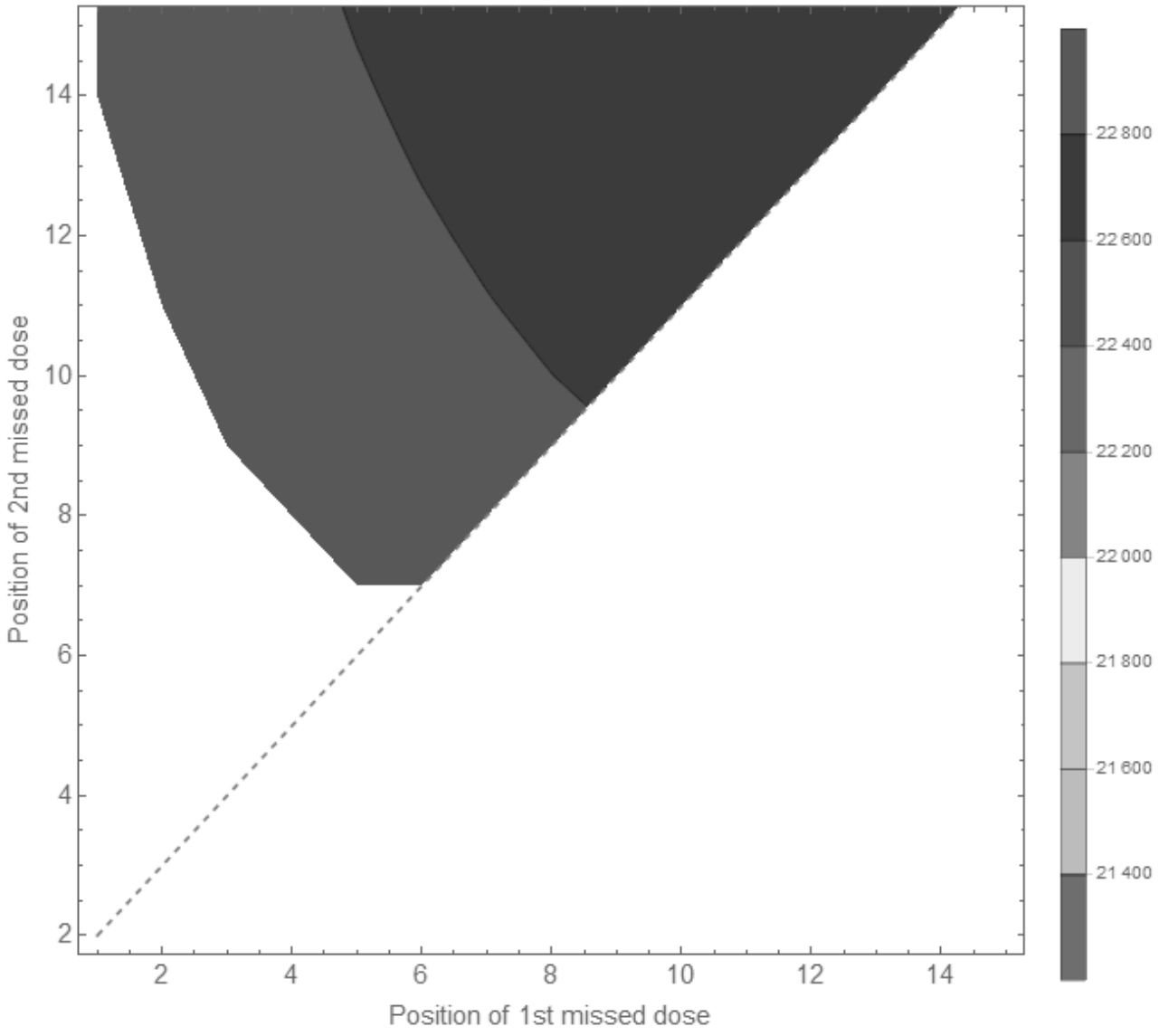

Figure 8: Contour plot of ToCE versus the position of two missed doses for 23040 mins or 16 days of treatment. White region above the dotted line indicates region where cancer survives. (Cropped)

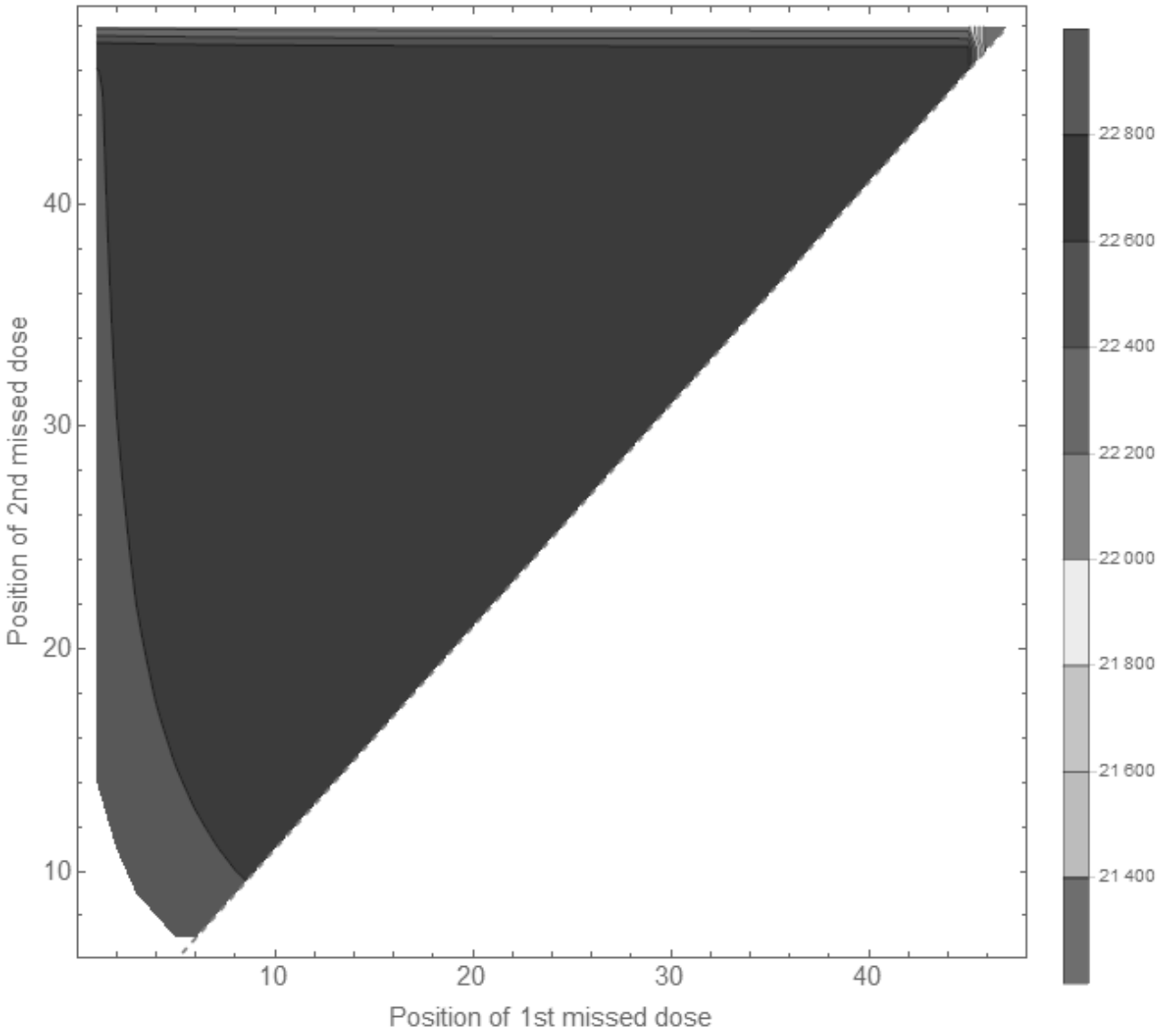

Figure 9: Contour plot of ToCE versus the position of two missed doses for 23040 mins or 16 days of treatment.(full)

4 Further discussion

A crucial parameter to the ABM is the diffusivity, i.e. how fast factors and cancer cells diffuse which in our captured in the ABM by the how often and how far agents move. According to[1] due to their size difference factors and cells must have a diffusion speed difference of the order of 1000, meaning that the factors diffuse 1000 times faster than cells. Of course that

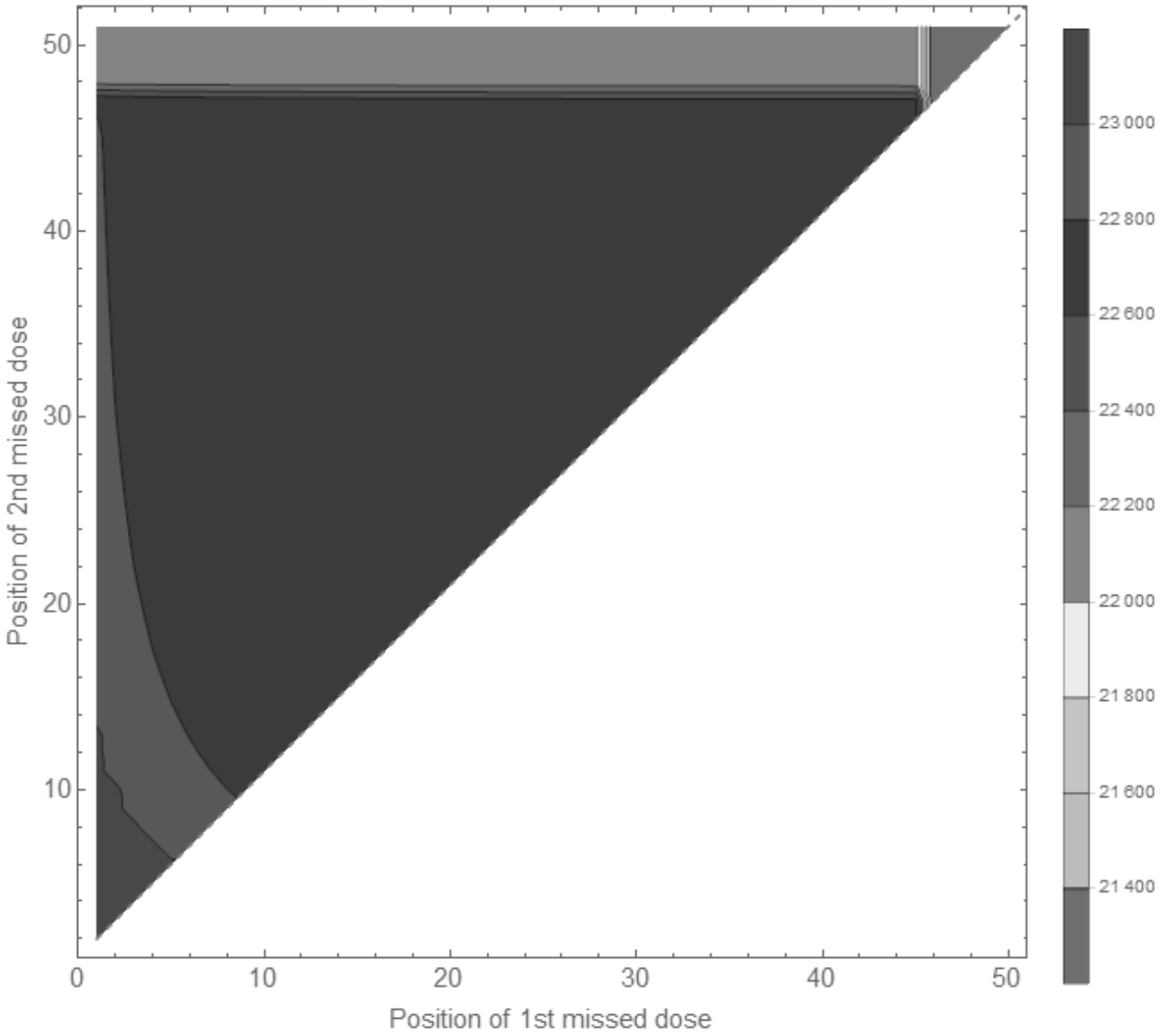

Figure 10: Contour plot of ToCE versus the position of two missed doses for 24480 mins or 17 days of treatment.

is a relative difference and the actual speed in the ABM affects the results significantly, as discussed below.

We initially tried the following scenario. Cells have unit speed, meaning that they move randomly to one of their neighbour patches at each unit of time and factors can move 1-1000 units of distance (patches). At each time step a random integer number between 1-1000 is chosen for the distance covered by factors. There are two issues with this scenario. The first

is that factors simply jump that distance without having the chance to interact with cell in between. This is not realistic but necessary as otherwise the computational time needed just for a single time step would be large. In addition this scenario appears, for our world dimensions, well-mixed (factors and cells are everywhere). Although that gives a better match with our ODE, which is expected since chemical-kinetics equations are mainly used for well-mixed cases of chemicals, it is unrealistic for a cancer case especially since cancer is more static.

The second scenario we tried, and the one we finally used, is that factors move a distance 1 every time-step and cancer cells move 1 every 1000 time steps. This has the following advantages: a) factors can now interact in every time-step, b) the cancer is much more static and grows out of its initialization positions. That scenario also gives rise to a different phenomenon. Since factors move slower and cancer cells are almost static there appears to be a critical over-density phenomenon, meaning that if the cancer cell in a single location exceeds a certain number then factors cannot kill them and that local population keeps increasing. In the cases were the over-density mentioned above appears in the ABM we wanted to see if the ODE can capture that behaviour to that end we tried changing the value for β which gave a qualitative behaviour similar to the ABM but not a close quantitative match. That effect cannot be captured correctly by the ODE model since the parameter β, although dependent on the agent speeds, cannot capture the full dynamics introduced by diffusion which would require a spatial model. Despite the mismatch in that particular case if the malignant cancer dies out completely then the two models match well, meaning that if the parameters are such that the over-density is never achieved we can see a good fit and that is the case for which we fitted β and performed the subsequent analysis. For the growth values found in Table 1 and an administration of 2000 factors every 8 hours the over-density never occurred and hence cancer never survived (Figure 2), all malignant cancer died and the two models have both a qualitative and quantitative match as we saw in Figure 3. That justifies the use of an ODE model. For different parameter values and more general exploration there is probably

the need for an alternative model of the form of a PDE in order to capture spatial effects as well. Moreover, if exclusion effects are included then some of the over-densities can be averted as cancer would be unable to grow without space.

From the above it is clear that the ODE cannot capture the ABM in its entirety, nor does it need to since we do not need a full description of the ABM via an ODE: achieving that would be very hard due to the different nature of these models. We only need them to be complementary and the continuum model to capture just the features and parameter spaces we are interested in further exploring.[4]

1000 cells were selected for the simulation of the ABM. An increase or decrease of this number would not have significant effects to the analysis performed, other than changes in the total dosage, as long as we remain in the "well-mixed" regime were the chemical kinetics-like ODE captures well. So as long as we avoid very small number of cells which in the ABM and SDE can lead to unexpected extinction or very large number which can cause the overdensity mentioned above the analysis holds.

The most important parameter in our continuous ODE model is β. It is a parameter not controlled by the experimentalist/modeller that reflects some intrinsic properties of the environment (size) and of the agents (speed, reaction type) and needs to be calibrated according to the available data. Changes in that parameter reflect how often cells and factors meet and hence have a change of interaction. As a result, an increase in β would keep the qualitative behaviour the same as the one explored and the only difference would be shorter timescales due to the fact that increased β means more factors interact with cancer cell per minute. That could potentially lead to decreased total dosage or frequency of administration. On the other hand decreasing β could even lead to cancer surviving in some cases depending on the dosage which would mean that there needs to be a higher or a more frequent dosage in order to kill cancer. We believe that as long as we are in the parameter regime where the ODE provides a sound approximation of the ABM, β does not affect the pattern of optimal dosage as evident by the distributions found for half and double the

original value of β used. That leads to the conclusion that the pattern is mainly dependent on the growth rates and hierarchy of the cancer cells.

The SDE model provides a more realistic alternative and we observed that it captures the variation observed in the ABM for the specific parameter values. Despite the fact that there was some noticeable variation in the healthy cell population, noise had little effect to the malignant populations which are the important ones for the optimization and sensitivity analysis performed. Due to that very narrow variation in these populations we are confident that a mean field approach (ODE) provides a sound approximation and would give very similar results to the SDE, on top of making numerical simulations less challenging.

Dosage distribution optimization was conducted via exhaustive search of the possible dose combinations for a particular total dosage. This brute force approach to optimizing the dosage was selected due to the fact that we only have four different cases so the number of combination for a specific total dosage is not very high. In addition to determining optimum distribution, which can potentially be used to reduce either ToCE or amount of dosage, there can be further exploration in order to find both the best timing for administration of each of the four doses as well as the best dosing scheme. A dosing scheme would mean a varying dose size which would be a function of the respective cancer size rather than a constant dosage size as in this paper. This additional exploration was not conducted here as the main aim of the paper was to show that a connection between agent-based and continuum models can be achieved in the cancer-related clinical setting.

5 Conclusion

Based on experimental evidence and a previously developed agent-based model we built a very similar model which was used as our basis for attempting a connection between discrete and continuous models. To this end we formulated two continuous models, one ODE and one SDE, in order to capture the behaviour of the ABM for specific, experimentally derived

parameters. The match between both models and the ABM was close both qualitatively and quantitatively which allowed for an extension of the ODE so that it can be used for more complex analysis, specifically optimization and sensitivity analysis. Through dosage optimization the pattern of optimum distribution for the four doses was found which shows significant gains in comparison to equal distribution of the four doses. Finally, a sensitivity analysis was conducted for a patient missing some doses which clearly demonstrated that the position of the dosage in the course of therapy is important. Future work can move in two directions, either towards the development of a new type of model (PDE) in order to capture the ABM better or towards more systematic optimization of doses using the already derived ODE. The latter is possible as we already mentioned that for a specific parameter space (when cancer dies out) the two types of models have a close fit. Further optimization would be necessary in order to find not only the right timing for administration of a certain dosage but also a varying-size scheme which could significantly reduce the overall amount and yet keep the ToCE to an acceptable range.

Disclosure

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the Engineering and Physical Sciences Research Council DTC [grant EP/F500351/1].

References

(1) Biava, P. M.; Basevi, M.; Biggiero, L.; Borgonovo, A.; Borgonovo, E.; Burigana, F. *Current Pharmaceutical Biotechnology* **2011**, *12*, 231–242.

(2) Krinner, A.; Roeder, I.; Loeffler, M.; Scholz, M. *BMC systems biology* **2013**, *7*, 117.

(3) Lugo, C. A.; McKane, A. J. *Physical Review E* **2008**, *78*, 051911.

(4) Oremland, M.; Laubenbacher, R. *Bulletin of Mathematical Biology* **2015**, *77*, 434–459.

(5) Biava, P. M.; Bonsignorio, D. *Journal of Tumor Marker Oncology* **2002**, *17*, 47–54.

(6) Biava, P. M.; Bonsignorio, D.; Hoxha, M. *Journal of Tumor Marker Oncology* **2001**, *16*, 195–202.

(7) Biava, P. M.; Bonsignorio, D.; Hoxha, M.; Impagliazzo, M.; Frosi, A.; Larese, F.; Negro, C. *Journal of Tumor Marker Oncology* **2002**, *17*, 55–58.

(8) Gillespie, D. T. *The Journal of Chemical Physics* **2000**, *113*, 297–306.

(9) Higham, D. J. *SIAM review* **2001**, *43*, 525–546.

Appendix

Comparison of original ABM to the one developed in this paper

Here we present two plots, figure 11 and figure 12, similar to the ones found in Fig.3[1] for the hypothetical case of 100% fitness. The first figure is for a constant dosage of 1000 factors for all 4 doses and a frequency of 8 hours whereas the second in for a dosage of 250 factors for all 4 doses and the same frequency. The time unit here is minutes whereas in the original plot it is days more specifically the plots range from 0 to 1.75 and 0 to 3.75 days which correspond to 0 to 1800 and 0 to 5400 respectively in our case.

Langevin representation of reaction systems

According to Gillespie,[8] in a system of chemical reactions we can use the Master equation to describe its evolution. The Master Equation is an ODE describing the probabilistic change

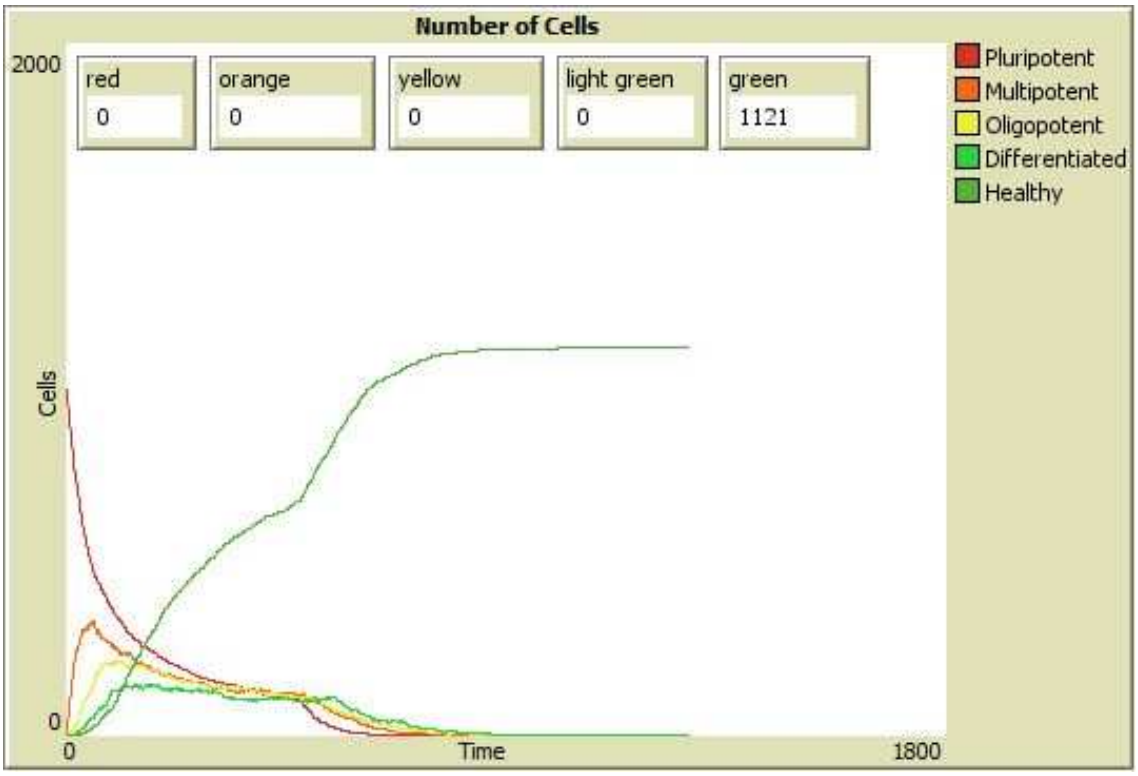

Figure 11: Simulation results for the hypothetical case of 100% factor fitness. For the case where $F_p = F_m = F_g = F_d = M(0) = 1000$.

of the state of the system. If two conditions are satisfied then according to that paper the Master Equation can be approximated by the Langevin equation, an SDE describing the evolution of the reactants population due to deterministic and stochastic events. The two conditions are:

(A) We require that there exists an infinitesimal interval dt, such that the change in the propensity. i.e,

$$\alpha_j(X_{t'}) = \alpha_j(X_t),$$

where $t' = t + dt$ and $\alpha_j(X) = c_j h_j(X)$. Here h(X) is equal to the product of the concentra-

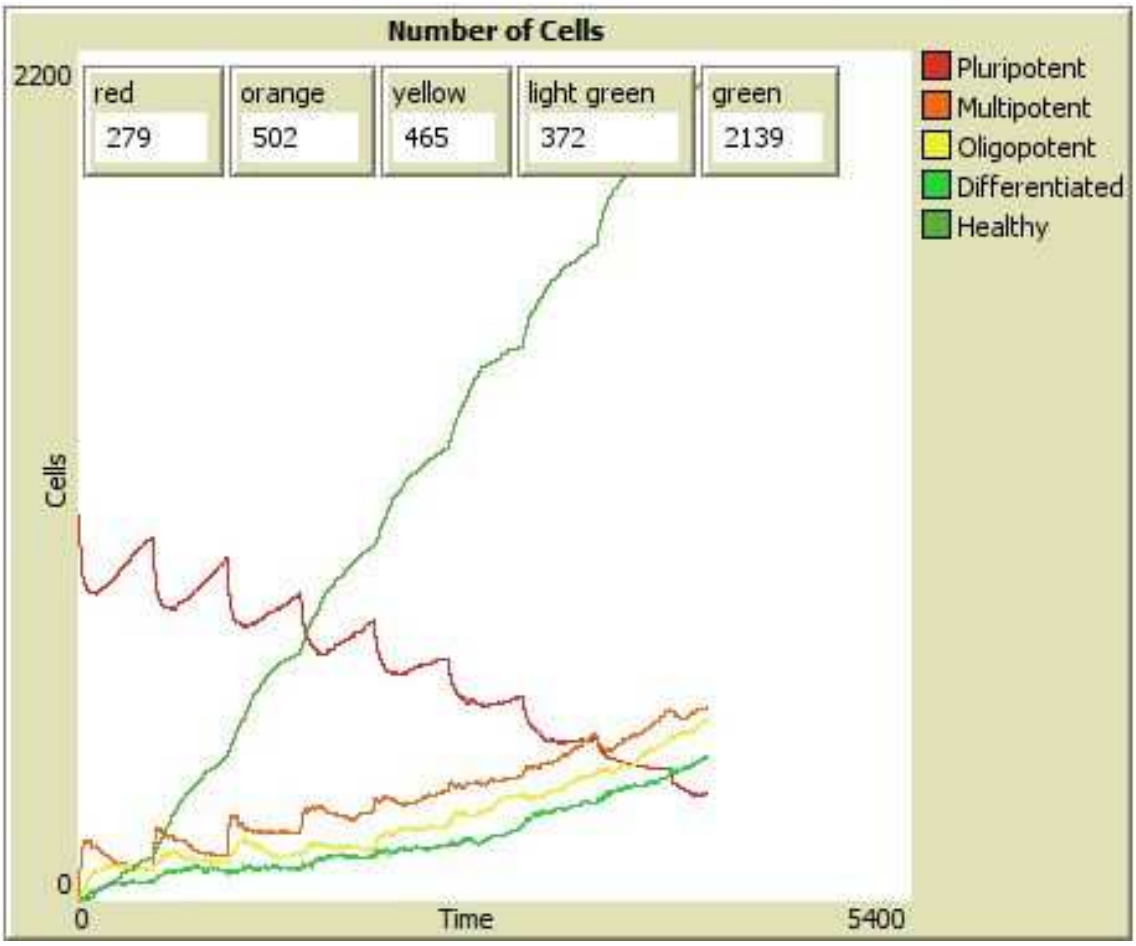

Figure 12: Simulation results for the hypothetical case of 100% factor fitness. For the case where $F_p = F_m = F_g = F_d = 250$.

tion of the reactants in bimolecular reactions or equal to the concentration of one reactant in monomolecular reactions.

(B) We require that this dt is large enough that the expected number of interactions is large.

These two conditions might seem contrasting but in the case of large populations they are both satisfied as (i) more than one reaction will occur in dt and (ii) the change of the populations in that interval will be insignificant compared to the total population and hence

the propensity will be almost constant.

If these two conditions are satisfied then it can be shown that a random variable $K(X(t),t)$ can become a statistically independent Poisson variable $P(a(X(t)),t)$ and then a normal random variable $N(a(X(t))dt, a(X(t))dt)$. Following that one can use the linear combination theorem for normal random variables,

$$N(m,\sigma^2) = m + \sigma * N(0,1),$$

to write the normal random variable in the following form:

$$a(X(t))dt + (a(X(t))dt)^{1/2}.$$

Using the procedure explained in[8] one can find the Langevin equation satisfying the system and that equation will be a good approximation of the Master equation. Furthermore, one can find the Fokker-Plank equation but here we concentrate on the Langevin. The general formulation is as follows:

$$\frac{dX_i}{dt} = \sum_{j=1} \nu_{j,i}\alpha_j(X(t)) + \sum_{j=1} \nu_{j,i}\alpha_j^{1/2}(X(t))\Gamma_j(t),$$

Here, $\nu_{j,i}$ describes the change in the population X_i due to reaction j and Γ_j is Gaussian white noise. The above formula was formulated by its discretized version which we use to simulate the model in Mathematica. The discretized version, which can be recognized as the Euler-Maruyama method, is:

$$X_i(t+dt) = X_i(t) + \sum_{j=1} \nu_{j,i}\alpha_j(X(t))dt + \sum_{j=1} \nu_{j,i}\alpha_j^{1/2}(X(t))N_j(0,1)dt^{1/2}.$$

N(0,1) is a random number of a normal distribution with mean zero and unit variance.

.1 *Application to present ABM*

Let us consider our cells and factors reaction model. We have the following reactions:

1. five reproducitons of the form X $\rightarrow$ 2X with rates g_p, g_m, g_o, g_d, g_h respectively
2. four productions $\emptyset \rightarrow F_i$ with rate D
3. four interactions of the form $X + F_x \rightarrow \emptyset$ with rate β.

So in total there are thirteen reactions. That gives the following vector and matrix for α and ν:

$$\alpha = \{g_p P, g_m M, g_o O, g_d D, g_h H, D, D, D, D, \beta P F_p, \beta M F_m, \beta O F_o, \beta D F_d\},$$

$$\nu = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

We can then formulate the Langevin equations for the populations using the formula mentioned above.