IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A mathematical model to study breast cancer growth

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1664614 since 2018-03-29T17:27:22Z

Publisher:

IEEE

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

10 October 2026

A mathematical model to study breast cancer growth

Giorgia Chiavassa
Dipartimento di Informatica
Università degli Studi di Torino
Torino, Italy
gchiavassa@di.unito.it

Chiara Fornari
Drug Safety and Metabolism
AstraZeneca
Cambrige, UK
chiafornari@gmail.com

Roberta Sirovich
Dipartimento di Matematica
Università degli Studi di Torino
Torino, Italy
roberta.sirovich@unito.it

Marzio Pennisi
Dipartimento di Matematica e Informatica
Università degli Studi di Catania
Catania, Italy
mpennisi@dmi.unict.it

Marco Beccuti
Dipartimento di Informatica
Università degli Studi di Torino
Torino, Italy
beccuti@di.unito.it

Francesca Cordero
Dipartimento di Informatica
Università degli Studi di Torino
Torino, Italy
fcordero@di.unito.it

***Abstract*—The aim of this paper is (i)to study breast cancer growth by mean of a mathematical model describing cell population dynamics during cancer growth, and (ii)to use this model to reproduce and explain experimental data.**

We started from a linear model describing cancer subpopulations evolution based on the Cancer Stem Cell (CSC) theory, and we added feedback mechanisms from the cell populations to mimic micro- environment effects in cancer growth. In details, we hypothesized two feedback mechanisms and we studied their effects both separately and combined together. In this way we obtained three new models that we tuned using data derived by TUBO Cancer cell line and describing the evolution of the total cell population and the subpopulations over time. Finally, we exploited these three models to understand which combination of feedback mechanisms better describe the experimental data.

***Index Terms*—Breast cancer growth model, Cancer Stem Cell theory and mathematical models.**

I. Introduction

Cancer Stem Cells (CSCs), a small population of cancer cells endowed with stem-like properties, play a crucial role in tumor initiation and progression, but CSCs also take part into therapy failure [2]. CSC-based tumors, like breast cancer, show a hierarchical organization with CSCs at the apex, Progenitor Cells (PCs) as the main proliferating population, and Totally differentiated Cells (TCs) forming the main tumor mass, [10]. CSCs reside in specific tumor regions called niches, which provide them nutrients and protection at the same time, [12]. This particular spatial organization in addition to the CSC phenotyoic properties may explain why current therapies are still ineffective on this small population of cancer cells, [3]. Complex dynamics characterize the mechanisms underlying the CSC hierarchy, such as cell fate decision and regulatory feedbacks. Multi-disciplinary approaches, which combine experimental studies with mathematical modeling, represent the winning strategy to address these open questions, [5]. In fact, a countless number of experiments can be performed *in silico* allowing to simulate different dosing schedules and therapies, and then testing only the most promising *in vitro* and/or *in vivo*.

Identify applicable funding agency here. If none, delete this.

From a mathematical point of view, cancer cell models range from deterministic to stochastic. Deterministic models are typically used for large population, while stochastic models are appropriate if single cell behaviours are interesting for the study [11]. Focusing on deterministic models, different choices can be done, as Partial Differential Equations (PDEs), Delay Differential Equations(DDEs) or Ordinary Differential Equations (ODEs) [11]. In this paper we chose ODEs to model breast cancer cell growth. Thus, in Sec. III three new models are introduced to extend with feedback mechanisms the one proposed in [1]. Then, in Sec.IV we describe how these new models can be tuned exploiting biological data. A discussion on the ability of these models to mimic the real cell growth is instead reported in Sec.V. Finally, in Sec. VI we conclude presenting some future works.

II. Background

First in this section we introduce the Breast cancer compartment model proposed in [1], that it is the starting point for the new models proposed in this paper. Secondly we describe the weighted least squares method used for the data fitting task. Then different techniques for estimating the goodness-of-fit of a statistical model are reported.

A. Breast cancer compartment models

In [1] we proposed a model to describe the evolution of an Erb2+ mammary cancer based on the CSC theory. A schematic representation of this mode is provided in Fig. 1 in which we considered $m+1$ stages corresponding to one CSC stage followed by m Progenitor Cell (PC) stages. We stated that CSCs can proliferate only symmetrically: a CSC generates two CSC-like daughter cells with probability p, and it generates two first generation PC (PC_1) with probability $(1-p)$. CSCs

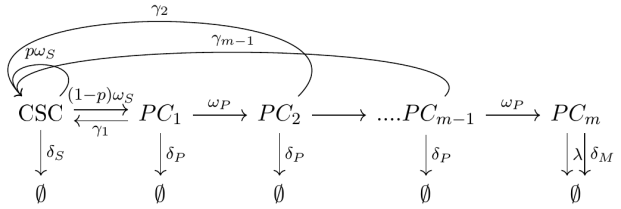

Fig. 1. Schematic representation of Breast Cancer compartment model (m+1 states).

proliferation rate is ω_s. We affirmed that a PC of generation i produces two PCs of generation $i+1$, for $i = 1, .., m-1$, with rate ω_p. PC_m are terminally differentiated cells, i.e. they can not proliferate and they will die for natural death or for mitotic failure. Each subpopulation is affected by a death rate specific for each cell type: δ_S, δ_P and $(\delta_M + \lambda)$ respectively. Finally, we considered de-differentiation from PC_i, $i = 1, .., m-1$, into CSC as proposed by Tang et al. in [2].

The obtained model in case of proliferation potential equal to 2 (i.e. m=2) was described through the following ODE system:

$$\begin{cases} \frac{dN_{CSC}}{dt} & = (2p-1)\omega_S N_{CSC} + \gamma N_{PC_1} - \delta_S N_{CSC} \\ \frac{dN_{PC_1}}{dt} & = 2(1-p)\omega_S N_{CSC} - \gamma N_{PC_1} - \omega_P N_{PC_1} \\ & -\delta_P N_{PC_1} \\ \frac{dN_{PC_2}}{dt} & = 2\omega_P N_{PC_1} - (\delta_P + \lambda) N_{PC_2} \\ N_{CSC}(0) & = N^0_{CSC} \\ N_{PC_1}(0) & = N^0_{PC_1} \\ N_{PC_2}(0) & = N^0_{PC_2} \end{cases} \tag{1}$$

where $N_{CSC}, N_{PC_1}, N_{PC_2}$ are the total number of CSCs, PCs_1 and PCs_2 respectively.

Observe in the rest of the paper we assume for simplicity $m = 2$, however all the obtained results can be easily extended for a generic m.

B. Weighted least squares method

The goal of data fitting is to find those parameter values that best describe the relation between experimental conditions and observations. The outcomes of the experimental measurements are called observations $y = (y_1, \ldots, y_N)$. They are dependent on several conditions, described by a vector $x = (x_1, .., x_N)$. Model parameters are described by a vector $a = (a_1, \ldots, a_M)$. Then, a model is given by $y = f(x|a)$. Observations are affected by random errors, and so we can reformulate the problem as $y_i = f(x_i|a) + \epsilon_i$ with $i = \{1, \ldots, N_s\}$ where a is an unknown and it is calculated through data fitting. methods.

Least-square method [4] is a data fitting technique that is based on minimizing the squared residual error S between observations y_i and the values calculated through the model function $f_i(x_i|a)$, i.e. minimizing:

$$S = \sum_{i=1}^{N} [f_i(x_i|a) - y_i]^2 \tag{2}$$

When observations are weighted, the method is called *weighted least squares method* and it is based on minimization of χ^2 , defined as:

$$\chi^2 = \sum_{i=1}^{N} w_i [f_i(x_i|a) - y_i]^2 \tag{3}$$

where w_i represents the reliability of the observations i, so that a more uncertainty associated with y_i correspond to lower w_i values. When the standard uncertainty σ_i of each y_i parent distribution is known, the optimal weight is equal to $w_i = \frac{1}{\sigma}$

C. How to evaluate the Goodness-of-fit

Goodness-of-fit of a statistical model, in general, measures how well the observed data correspond to the fitted model. Typical methods to evaluate the goodness of fit of a model consist in using some statistical coefficients, as for instance the coefficient of determination (R^2) and the Root Mean Standard Error (RMSE), or/and in the residual analysis.

The residuals are defined as the differences $e_i = y_i - \hat{y}_i$, with $i = \{1, 2, \ldots, N\}$, where y_i is an observation and $\hat{y}_i$ is the corresponding fitted value obtained by the model. Weighted residuals are instead defined as:

$$Resi = \frac{y_i - \hat{y}_i}{\sigma_i}$$

Then, the residuals contain information on how the model can fit the data [6]. Some of the most popular goodness-of-fit methods are:

1. *The Sum of Squared Residuals* (SSR) is the sum of the weighted squares of residuals:

$$SSR = \sum_{i=i}^{N} w_i (y_i - \hat{y}_i)^2$$

and it is a measure of the discrepancy between the data and the fitting model.

2. *Coefficient of determination* (R^2) is defined for linear models and it measures how well the fit explains the variation in the data. R^2 is defined as the ratio of the explained sum of squares (SSE) and the total sum of squares (SST) [7]. Indicating with $\bar{y}$ the mean of the observations $y_i, i = 1, ..., n$, SSE is defined as:

$$SSE = \sum_{i=1}^{N} w_i \cdot (\hat{y}_i - \bar{y})^2$$

and SST is defined as:

$$SST = \sum_{i=1}^{N} w_i \cdot (y_i - \bar{y})^2$$

where SST = SSR + SSE. Given these definitions, R^2 is expressed as:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

It is important to observe that the use of R^2 to evaluate and choose among nonlinear models can lead to not robust

results [8].

3. *The Root Mean Standard Error*(RMSE) is the square root of the variance of the residuals [6]. It is defined as:

$$RMSE = s = \sqrt{MSE}$$

where

$$MSE = \frac{SSR}{N-k}$$

with k indicating the number of parameters and N the number of observations. It indicates how close the observed data points are to the model predicted values. Moreover, RMSE penalizes the lack of parameter parsimony in a model.

4. *Akaike's information criterion (AIC)* [9] is especially used to compare models with different number of parameters by penalizing those that use a greater number of parameters. It is defined as:

$$AIC = N \ln\left(\frac{SSR}{N}\right) + 2p$$

5. *Residual Analysis* assumes that if the fitting model is correct, then the residuals approximate the random errors. Thus, when the residuals appear to behave randomly then the model fits the data well. Therefore, after fitting experimental data with weighted least squares method, it is necessary to check the residual plots in order to validate the model. The normality of the residuals can be checked in different ways, for example with histograms, probability plots or with the Kolmogorov-Smirnov test [6].

III. Breast cancer compartment model

In this section we introduce three new Breast cancer compartment models that extend the model presented in the previous section II to take into account micro-environment in cancer growth. Indeed recent works have underlined the importance of the micro-environment in cancer growth, indicating that it can be modeled as a feedback mechanism from the cell populations in themselves. In particular, spatial limitation can results in a slow down of CSC proliferation when the PC population increased.

Hence, to describe such an auto growth limitation process, we hypothesized two feedback mechanisms that modulate CSC proliferation rate and probability of CSC self-renewal: both separately and combined together. In details, we propose to replace CSC proliferation rate ω_S with:

$$\omega_S(PC_1 + PC_2) = \frac{\omega_S}{1 + \alpha(N_{PC_1} + N_{PC_2})} \tag{4}$$

and/or the probability of CSC self-renewal p with

$$p(PC_1 + PC_2) = \frac{p}{1 + \beta(N_{PC_1} + N_{PC_2})} \tag{5}$$

Feedback mechanism in Eq. 4 acts on CSC proliferation rate, slowing it down as subpopulations PC_1 and PC_2 grow. Consequently, CSC self renewal and CSC symmetric differentiative division are both slowed down. The parameter α at the denominator of Eq. 4 represents the feedback intensity and we imposed it to be positive and less than one. Moreover, it is reasonable to think that it cannot be too small, because otherwise it could delete feedback effects.

Feedback in Eq. 5 acts on the probability of CSC self-renewal, slowing it down as subpopulations PC_1 and PC_2 grow. Consequently there is a decrement in CSC self-renewal and an increase in CSC symmetric differentiative division. The parameter β at the denominator, as α in Eq. 4, represents the feedback intensity and we imposed it to be positive and less than one.

Starting from the equations 4 and 5 we defined the following three extensions of the ODE system in 1.

- **Model 1**: feedback only on CSC proliferation rate in which ω_s is replaced with Eq. 4:

$$\begin{cases} \frac{dN_{CSC}}{dt} &= (2p-1)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC} \\ &\quad +\gamma N_{PC_1} - \delta_S N_{CSC} \\ \frac{dN_{PC_1}}{dt} &= 2(1-p)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC} \\ &\quad -\gamma N_{PC_1} - \omega_P N_{PC_1} - \delta_P N_{PC_1} \\ \frac{dN_{PC_2}}{dt} &= 2\omega_P N_{PC_1} - \delta_{P_2} N_{PC_2} \\ N_{CSC}(0) &= N^0_{CSC} \\ N_{PC_1}(0) &= N^0_{PC_1} \\ N_{PC_2}(0) &= N^0_{PC_2} \end{cases} \tag{6}$$

- **Model 2**: feedback only on probability of CSC symmetric division rate in which p is replaced with Eq. 5:

$$\begin{cases} \frac{dN_{CSC}}{dt} &= \left(2\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})} - 1\right)\omega_S N_{CSC} \\ &\quad +\gamma N_{PC_1} - \delta_S N_{CSC} \\ \frac{dN_{PC_1}}{dt} &= 2\left(1 - \frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\omega_S N_{CSC} - \gamma N_{PC_1} \\ &\quad -\omega_P N_{PC_1} - \delta_P N_{PC_1} \\ \frac{dN_{PC_2}}{dt} &= 2\omega_P N_{PC_1} - \delta_{P_2} N_{PC_2} \\ N_{CSC}(0) &= N^0_{CSC} \\ N_{PC_1}(0) &= N^0_{PC_1} \\ N_{PC_2}(0) &= N^0_{PC_2} \end{cases} \tag{7}$$

- **Model 3**: feedback on both CSC proliferation rate and probability of CSC symmetric division rate in which ω_s and p are replaced with Eq. 4 and Eq. 5:

$$\begin{cases} \frac{dN_{CSC}}{dt} &= \left(2\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})} - 1\right)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC} \\ &\quad +\gamma N_{PC_1} - \delta_S N_{CSC} \\ \frac{dN_{PC_1}}{dt} &= 2\left(1 - \frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC} \\ &\quad -\gamma N_{PC_1} - \omega_P N_{PC_1} - \delta_P N_{PC_1} \\ \frac{dN_{PC_2}}{dt} &= 2\omega_P N_{PC_1} - \delta_{P_2} N_{PC_2} \end{cases} \tag{8}$$

It is important to observe that in all these three extensions the introduction of the auto growth mechanisms made system non linear, so that these ODE systems are more complex to be studied than the original one (see ODE system 1).

	Repl1 TUBO		Repl2 TUBO		Repl3 TUBO
T	Counts	T	Counts	T	Counts
0	100000	0	100000	0	120000
8	175000	8	125000	8	NaN
24	225000	24	200000	24	225000
32	262500	32	250000	32	300000
48	292500	48	437500	48	370000
56	300000	56	575000	56	NaN
72	350000	72	775000	72	400000
80	1338000	80	800000	80	500000
		96	1368750	96	660000
		104	1512500	104	NaN
		120	1500000	120	1400000
		144	1500000	144	1400000
				152	1375000

TABLE I
CELL GROWTH OVER TIME IN TERM OF CELL COUNTS FOR THREE INDEPENDENT REPLICATES.

	Repl1 TUBO		Repl2 TUBO		Repl3 TUBO
T	CD44+/CD24-	T	CD44+/CD24-	T	CD44+/CD24-
0	0.53	0	0.3	0	0.35
24	0.3	24	0.2	24	0.26
48	0.12	48	1	48	0.1
72	0.1	72	0.5	72	0.3
		96	0.5		

TABLE II
PERCENTAGE OF CD44+/CD24- CELLS OVER TIME, FOR EACH INDEPENDENT REPLICATE.

IV. THE TUNING OF THE THREE MODELS

In this section we describe how to find those parameter values that best describe the relation between experimental condition and observations.

In particular the experimental data were generated via in vitro experiments on TUBO cancer cells. TUBO cells are cancer epithelial cells derived from BALB-neuT mice. These mice are transgenic for ErbB2 gene and spontaneously develop breast cancer few weeks after birth.

	TUBO	
	MEAN	ST. DEV
T		
0	106666.67	9428.09
8	150000	25000
24	216666.67	11785.11
32	270833.33	21245.92
48	366666.67	59242.91
56	437500	137500
72	508333.33	189663.44
80	879333.33	346680.77
96	1014375	354375
104	1512500	346680.77
120	1450000	50000
144	1450000	50000
152	1375000	346680.77

	CD44+/CD24-	
T	mean	st dev
0	0.3933333	0.0987702
24	0.2533333	0.0410961
48	0.4066667	0.4196295
72	0.3	0.1632993
96	0.5	0.4196295

TABLE III
MEAN AND STANDARD DEVIATION OF CELL COUNTS AND CD44+/CD24- PERCENTAGE OVER TIME, RESPECTIVELY.

Cell growth was monitored over time and table I shows the results from three independent replicates of the experiment. In addition to cell counts, we collected also flow cytometry data to infer the behavior and the class/type of cells. Specifically CD44+/CD24- was used as a marker for CSCs. Table II shows the percentage over time of CD44+/CD24- cells.

Then, means and standard deviations of the data reported in the three replicates of tables I and II were calculated and reported in tables III(left) and III(right). Observe that for all the cases where only one replicate was available we fixed the standard deviation equal to the maximum value between all the other standard deviations.

Afterward, the weighted least squares method is used to fit experimental data of III(left) and III(right) with our models.

Data fitting for Model 1.
For ODE system 6, in which only the feedback on CSC proliferation rate is considered, we define the following parameter vector $a = (p, \omega_S, \omega_P, \delta_S, \delta_P, \delta_{P_2}, \alpha, \gamma)$. Then, a qualitative study on the non trivial steady state of ODE system 6 is derived to widely restrict the search parameter space for the data fitting problem. Leaving behind the term with γ, because de-differentiation is a biological almost neglictable process, the ODE system 6 becomes:

$$\begin{cases} (2p-1)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC} - \delta_S N_{CSC} = 0 \\ 2(1-p)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC} - \omega_P N_{PC_1} - \delta_P N_{PC_1} = 0 \\ 2\omega_P N_{PC_1} - (\delta_P + \lambda)N_{PC_2} = 0 \end{cases} \tag{9}$$

Moreover, we know from the biological experiments that all the parameters are non-negative and that variables N_{CSC}, NP_{C1} and NP_{C2} are different from zero at this steady state. From the first equation of ODE system (9) we obtained

$$(2p-1)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})} = \delta_S$$

$$(2p-1)\omega_S = [1+\alpha(N_{PC_1}+N_{PC_2})]\delta_S$$

As the term at the right of the equal is positive, we deduced that

$$2p-1 > 0$$

$$p > 0.5$$

From the second equation of ODE system (9) we obtained

$$2(1-p)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC} = (\omega_P+\delta_P)N_{PC_1}$$

$$2(1-p)\frac{\omega_S}{\omega_P+\delta_P} = \frac{N_{PC_1}}{N_{CSC}}(1+\alpha(N_{PC_1}+N_{PC_2})) > \frac{N_{PC_1}}{N_{CSC}}$$

and $\frac{N_{PC_1}}{N_{CSC}} > 1$ because CSC form is the smallest population inside the tumor. Therefore:

$$2(1-p)\omega_S > \frac{N_{PC_1}}{N_{CSC}}(\omega_P+\delta_P)$$

Knowing that $p > 0.5$ (from the first equation) we deduced that $2(1-p) < 1$ and so

$$\omega_S > 2(1-p)\omega_S > M(\omega_P+\delta_P)$$

This means that ω_S must be larger than the sum of ω_P and δ_P. In this way we widely restricted the search parameter space for the data fitting problem speeding up the computation of the optimal vector a. Finally, the optimal vector derived through the weighted least squares method is $a = (p = 0.5204, \omega_S = 8.7529, \omega_P = 0.0105, \delta_S = 0.3256, \delta_P = 0.0030, \delta_{P_2} = 0.0157, \alpha = 9.2551 \cdot 10^{-8}, \gamma = 1.0035 \cdot 10^{-6})$.

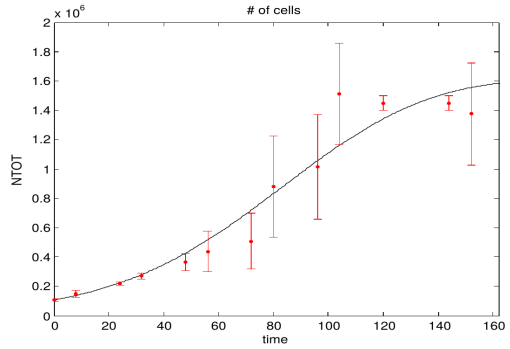

Fig. 2. **Model 1**: results of fitting data w.r.t. to cell count.

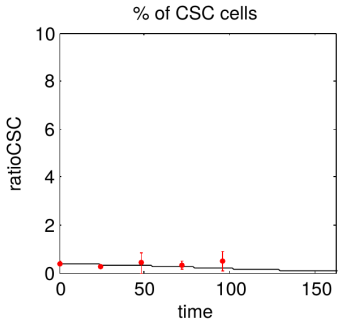

Fig. 3. **Model 1**: results of fitting data w.r.t. to CSC subpopulation percentage.

Figs. 2 and 3 represent the results of fitting data of tables III(left) and III(right) respectively. The red points, with the corresponding error bars, represent experimental data with the corresponding standard deviations. Black curves correspond to total population evolution and subpopulations percentage evolution obtained from **Model 1** considering as input the computed optimal parameters vector a.

To evaluate the goodness of fit of **Model 1** we analyzed the residuals. In Fig. 4 the normal probability plot of the residuals is an approximate straight line, indicating the normal behavior of the residuals. Moreover the Kolmogorov-Smirnov test was applied on the residuals. We obtained a p-value equal to 0.5446 and, consequently, we did not reject the null hypothesis: the residuals follow a normal distribution.

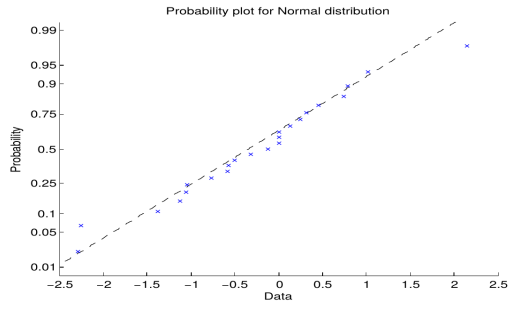

Fig. 4. **Model 1**: normal probability plot of the residuals.

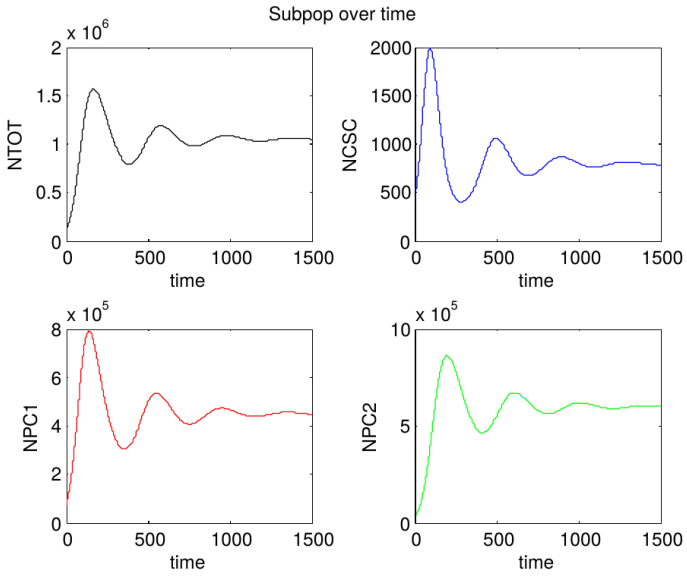

Fig. 5. **Model 1**: the evolution over the time interval T=[0,1500] of total population, CSC, PC_1 and PC_2 subpopulations,respectively.

Figs. 5 shows total population and subpopulations evolution over the time interval [0,1500] for **Model 1**. We noted that all the subpopulations oscillate , until reaching a stabilization around time $t = 1000$, with a final total population value around $1 \cdot 10^6$. CSC final population percentage is very little, while PC_1 and PC_2 final percentages are quite similar: PC_1 subpopulation fraction is around the 45%, while PC_2 subpopulation fraction is around 55%.

Data fitting for Model 2.

For ODE System 7, the parameter vector a was defined by: $a = (p, \omega_S, \omega_P, \delta_S, \delta_P, \delta_{P_2}, \beta, \gamma)$. A qualitative study on the non trivial steady state of ODE system 7 is performed to widely restrict the search parameter space for the data fitting problem again.

We experimentally know that all the parameters are non-negative and that variables N_{CSC}, NP_{C1} and NP_{C2} are different from zero at this steady state.

From the first equation of System 7 neglecting the term with γ we obtained:

$$\left(2\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}-1\right)\omega_S = \delta_S$$

As the term at the right of the equal is positive, we deduced that

$$\left(2\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}-1\right) > 0$$

$$\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})} > 0.5$$

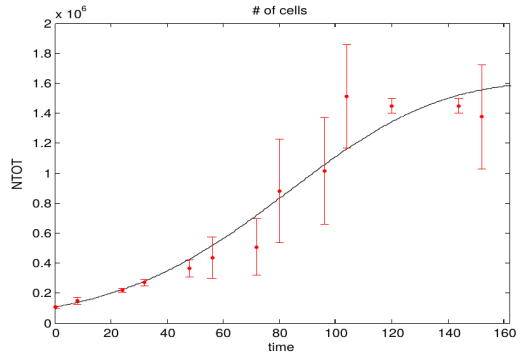

Fig. 6. **Model 2**: results of fitting data w.r.t. to cell count.

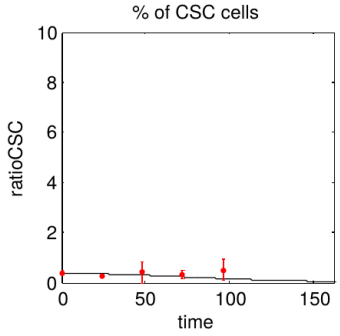

Fig. 7. **Model 2**: results of fitting data w.r.t. to CSC subpopulation percentage.

From the second equation of System (7) neglecting the term with γ we obtained

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\omega_S N_{CSC} = (\omega_P+\delta_P)N_{PC_1}$$

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\frac{\omega_S}{\omega_P+\delta_P} > \frac{N_{PC_1}}{N_{CSC}} = M$$

and $M > 1$ because we know that CSC form is the smallest population inside the tumor. Therefore:

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\omega_S > M(\omega_P+\delta_P)$$

Knowing that $\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})} > 0.5$ from the first equation, we deduced that

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right) < 1$$

and so

$$\omega_S > 2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\omega_S > M(\omega_P+\delta_P)$$

This means that ω_S must be larger than the sum of ω_P and δ_P. Thus, we widely restricted the search parameter space for the data fitting problem speeding up the computation of the optimal vector a. Finally, the optimal vector derived through the weighted least squares method is $a = (p = 0.5101, \omega_S = 7.8952, \omega_P = 0.0094, \delta_S = 0.1312, \delta_P = 0.0009, \delta_{P_2} = 0.0107, \beta = 4.5339 \cdot 10^{-9}, \gamma = 1.0700 \cdot 10^{-5})$.

Figs. 6 and 7 represent the results of fitting data of tables III(left) and III(right) respectively.

To evaluate the goodness of fit of **Model 2** we analyzed the residuals. In Fig. 8 the normal probability plot of the residuals is an approximate straight line again. Moreover the Kolmogorov-Smirnov test was applied on the residuals obtaining a p-value equal to 0.60775. Therefore, we did not reject the null hypothesis and we stated that the residuals follow a normal distribution.

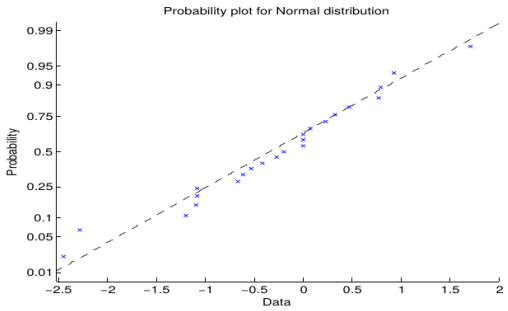

Fig. 8. **Model 2**: normal probability plot of the residuals.

Finally, Fig. 9 shows total population and subpopulations evolution over the time interval [0,1500] for **Model 2**. We noted, as for **Model 1**, that all the subpopulations oscillate, until reaching a stabilization around time $t = 1000$. The total population value is around $1 \cdot 10^6$, CSC final percentage is very small, while PC_1 and PC_2 final percentages are around the 40% and the 60%, respectively.

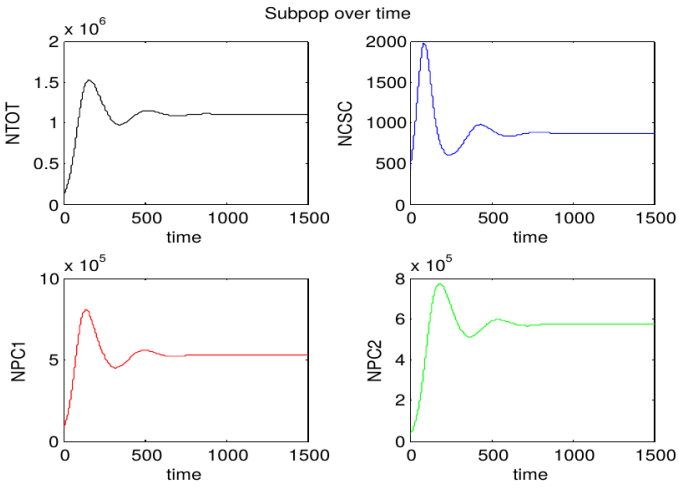

Fig. 9. **Model 2**: the evolution over the time interval T=[0,1500] of total population, CSC, PC_1 and PC_2 subpopulations,respectively.

Data fitting for Model 3.

For ODE System 8, the parameter vector a was defined by: $a = (p, \omega_S, \omega_P, \delta_S, \delta_P, \delta_{P_2}, \alpha, \beta, \gamma)$. As in the previous models we performed a qualitative study on the non trivial steady state of ODE system 8 whose aim was to reduce the search parameter space for the data fitting problem. We assume that all the parameters are non-negative and variables N_{CSC}, N_{PC_1} and N_{PC_2} are different from zero at this steady state.

From the first equation of System 8 neglecting the term with γ we obtained

$$\left(2\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}-1\right)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}=\delta_S$$

As the term at the right of the equal is positive, we deduced that

$$\left(2\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}-1\right)>0$$

$$\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}>0.5$$

From the second equation of System (9) neglecting the term with γ we obtained

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\frac{\omega_S}{1+\alpha(N_{PC_1}+N_{PC_2})}N_{CSC}=(\omega_P+\delta_P)N_{PC_1}$$

Therefore:

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\frac{\omega_S}{\omega_P+\delta_P}>$$

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\frac{\omega_S}{(1+\alpha(N_{PC_1}+N_{PC_2}))(\omega_P+\delta_P)}>\frac{N_{PC_1}}{N_{CSC}}=M$$

and $M > 1$ because we know that CSC form the smallest population inside the tumor. Therefore:

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\omega_S>M(\omega_P+\delta_P)$$

Knowing that $\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}>0.5$ we deduced that

$$2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)<1$$

and so

$$\omega_S>2\left(1-\frac{p}{1+\beta(N_{PC_1}+N_{PC_2})}\right)\omega_S>M(\omega_P+\delta_P)$$

This means that ω_S must be larger than to the sum of ω_P and δ_P.

As for **Model 1** and **Model 2** we substantially reduced the search parameter space for the data fitting problem speeding up the computation of the optimal vector a. Thus, the optimal vector derived through the weighted least squares method

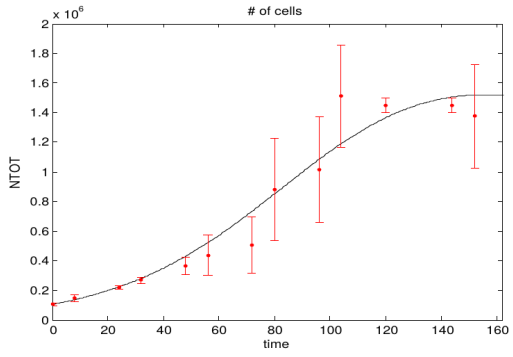

Fig. 10. **Model 3**: results of fitting data w.r.t. to cell count.

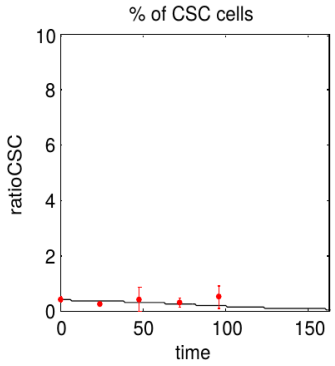

Fig. 11. **Model 3**: results of fitting data w.r.t. to CSC subpopulation percentage.

is $a = (p = 0.5080, \omega_S = 7.9711, \omega_P = 0.0123, \delta_S = 0.0970, \delta_P = 0.0007, \delta_{P_2} = 0.0227, \alpha = 1.0002 \cdot 10^{-9}, \beta = 4.3096 \cdot 10^{-9}, \gamma = 1.3921 \cdot 10^{-5})$.

Figs. 10 and 11 represent the results of fitting data of tables III(left) and III(right), respectively.

Afterwards, to evaluate the goodness of fit of **Model 3**, we analyzed the residuals. As for for **Model 1** and **Model 2**, Figs. 12, representing the normal probability plot of the residuals. states that residuals distribution is likely to be normal. Moreover Kolmogorov-Smirnov test was applied on the residual. We obtained a p-value equal to 0.51481 so that

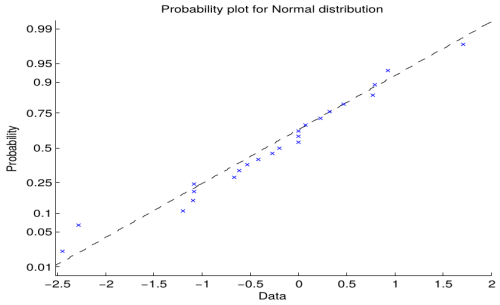

Fig. 12. **Model 2**: normal probability plot of the residuals.

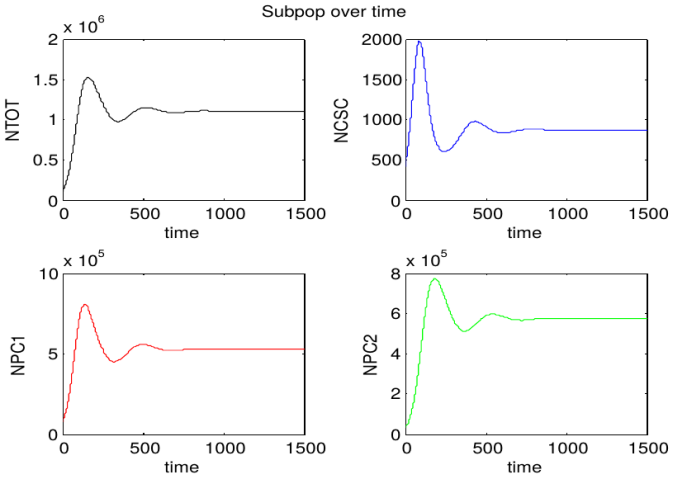

Fig. 13. **Model 3**: the evolution over the time interval T=[0,1500] of total population, CSC, PC_1 and PC_2 subpopulations,respectively.

	R^2	SSR	RMSE	AIC
Model 1	0.9970	24.48	1.2774	17.43
Model 2	0.9971	25.39	1.2655	16.99
Model 3	0.9971	23.72	1.3017	18.71

TABLE IV
R^2, SSR, RMSE AND AIC VALUES ARE REPORTED FOR THE THREE MODELS.

the null hypothesis cannot be rejected.

Finally, in Figure 13, we represented subpopulations evolution over the time interval $[0, 1500]$. We noted that, initially, subpopulations oscillate, until reaching a stabilization. However, this stabilization occurs at earlier times respect to System1 and System2. Final total population value is around $1.1 \cdot 10^6$, CSC final percentage is very small, while PC_1 and PC_2 final values are around $5 \cdot 10^5$ and $6 \cdot 10^5$, respectively.

V. COMPARING THE THREE NEW MODELS

In the previous section we showed that all the three systems can be used to explain experimental data of tables III(left) and III(right). Here, we investigate the R^2, SSR, RMSE and AIC values for the optimal set of parameters in three models.

Table IV shows R^2, SSR, RMSE and AIC values for the optimal set of parameters in the three models.

Among the three models, we observe that the best fit is the third one. Indeed, it has the minimum SSR value. Moreover, it has two feedbacks and, for this reason, it is the most versatile in the description of the microenvironment effects, needing only one more parameter compared to the other systems. RMSE and AIC values are a little higher for **Model 3** than for the other models, because these coefficients also consider the number of parameters involved. However, the difference is neglected, so **Model 3** is still preferable for its greater ability to describe complex microenvironment effects. Moreover, we noted in Fig. 10 that **Model 3** better describes total population stabilization in our experimental data, compared to **Model 1** and **Model 2** (see Figures 2 and 6).

VI. CONCLUSION AND FUTURE WORKS

Cancer is a complex system that can only be understood through the integration of theory and experiments. The goal of Mathematical Oncology is to describe processes in oncology using the tools of mathematics.

In this paper, we proposed three new models which extend the a mathematical model describing breast cancer cell growth in [1] considering micro-environment effects on tumor growth.

All these three proposed models resulted able to describe the experimental data quite well. Among them the **Model 3** seems to give the best fit of the data. More precisely: (i) it has the minimum SSR value, (ii) since it has two feedbacks, it is more flexible to model the complex microenvironment effects on CSC dynamics, having only 1 parameter more than the other 2 versions.

As future work, we are considering to develop a model that takes into account CSC variability among different cancer cell populations. In this way, we would not be obliged to do the preliminary manual work to align experimental data. This model modification consists in introducing a new term that modulates cell proliferation in accordance with CSC percentage. More precisely, cancer cells remain mainly quiescient when CSCs are above a certain threshold.

REFERENCES

[1] Fornari C, Beccuti M, Lanzardo S, Conti L, Balbo G, Cavallo F, Calogero R A, Cordero F., "A Mathematical-Biological Joint Effort to Investigate the Tumor-Initiating Ability of Cancer Stem Cells", PLOSE ONE , 9 (2014)
[2] Tang T., "Understanding cancer stem cell heterogeneity and plasticity", Cell Res 22, 457-472 (2012)
[3] R. Pardal, M.F. Clarke, S.J. Morrison. "Applying the principles of stem-cell biology to cancer", Nature Review Cancer, 3, 2003.
ARA Anderson V Quaranta. Integrative mathematical oncology. Nature Reviews Cancer, 8, 2008.
[4] Rao, C. R.; Toutenburg, H.; et al. . "Linear Models: Least Squares and Alternatives", Springer Series in Statistics (3rd ed.), Berlin:Springer (2008).
[5] A.R.A. Anderson, V. Quaranta. "Integrative mathematical oncology", Nature Reviews Cancer, 8, 2008
[6] Draper, N.R., Smith, H., "Applied Regression Analysis" (3rd ed.) John Wiley (1998)
[7] Casella G., Berger R.L., "Statistical Inference", Duxbury Pr., 2nd ed. (2001)
[8] Spiess A.N., Neumeyer N., "An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach", BMC Pharmacology, 10:6 (2010)
[9] Benzekry S., Lamont C., Beheshti A., Tracz A., Ebos J.M.L., Hlatky L., Hahnfeldt P., "Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth", PLoS Comput Biol 10(8) (2014)
[10] Frank N.Y., Schatton T., Frank M.H., "The therapeutic promise of the cancer stem cell concept", The Journal of Clinical Investigation 120, 41-50 (2010)
[11] Getto P., Marciniak-Czochra A.,"Mathematical modeling as a tool to understand cell self-renewal and differentiation", Methods in Molecular Biology 1293, 247-266 (2015)
[12] R. Bjerkvig, B.B. Tysnes, K.S. Aboody, J. Najbauer, A.J.A. Terzis. "The origin of the cancer stem cells: current controversis and new insights", Nature Review Cancer, 5, 2005.