POLITECNICO DI MILANO

FACOLTÀ DI INGEGNERIA DEI SISTEMI

Corso di Laurea Magistrale in Ingegneria Biomedica

PREDICTING THE EVOLUTION OF CERVICAL CANCER DURING THE TREATMENT USING ODE MODELS AND VALIDATION BY MEANS OF CBCT IMAGES

Relatore:
Prof. Pietro CERVERI

Correlatore:
Prof. Marco RIBOLDI

Tesi di Laurea di:
Antonella BELFATTO
Matr. 766125

Anno Accademico 2011/2012

Index

Figures index

Acknowledgments

I would like to express my gratitude to my Italian supervisors, prof. Pietro Cerveri and prof. Marco Riboldi, who helped me throughout the development of this project.

A very special thanks goes to prof. Marc Garbey (University of Houston). He welcomed me in his laboratory and assisted me in the most practical part of the work and in ordinary life, making me feel home.

I must also acknowledge Delia Ciardo and dr. Agnese Cecconi from the IEO (European Institute of Oncology) who provided us the data and kindly answered our (probably lay) questions

I would also like to thank my family that supported me during this five years, letting me chose my way even when it lead me far from them.

Finally I have to admit that I would not have been able to accomplish this work without the motivation and energy that my friends instilled in me in a very delicate time of my life. I felt the old friends so close even when I was far away, I also met wonderful people because of my travels and other fortunate coincidences. I would like to quote each of them, but I am sure they will feel my incommensurable gratitude all the same.

English abstract

In the fight against cancer, radiotherapy represents a highly effective localized treatment, in which image-guidance (Image Guided RadioTherapy, IGRT) plays a fundamental role. Images featuring diagnostic quality are acquired daily, in order to check the occurrence of anatomical changes, to adjust patient setup and to evaluate the need for treatment re-planning. In addition, functional imaging to visualize tumor characteristics (e.g., glucose consumption, hypoxia, repopulation) is applied, not only for the initial tumor staging, but also at given time intervals during the fractionated course of treatment. Despite advances in recognizing both the physics and radiation biological mechanisms underpinning radiotherapy treatment, early prediction of tumor response and toxicity effects remains difficult to achieve, particularly at the level of the individual patient. Over the last 10 years, attempts have been made to apply models to predict tumor response and toxicity occurrence on a patient specific level However the translation of computational approaches into effective clinical practice has been hindered by lack of convincing validation.

As far as tumor response and therapy outcome prediction are concerned, many efforts have been tackled to reinforce the clinical potentials of in silico models towards their translation into radiotherapy clinical practice, for the optimization and personalization of the treatment. Biological modeling of the tumor response however is still playing only a very limited role in commercial treatment planning system and is restricted to the use of classical tumor control probability models. These models make their predictions using population-based input parameters. Due to the dynamic character of tumor properties, this is not sufficient to predict the development of tumors in the individual case. Tumor heterogeneity caused by the different radio-sensitivities of cell sub-populations, as well as by varying environmental factors such as oxygenation, prevents researchers to get reliable predictions for the individual patient. This situation can only be improved if individual input parameters, as well as mechanistic models on the spatio-temporal development of the tumor, are systematically considered. To achieve this, specific in-silico models have been developed, which simulate the spatio-temporal growth of tumors, based on basic radiobiological principles. Single-cell-based models using the principles of cellular automata have been developed and applied to study a variety of radio-biological problems. Although these models may well be used to simulate the mechanisms underlying radiation response, they are limited with respect to clinical application as the morphological and functional information from medical images are difficult to include. This led to the development of voxel-based tumor-models, which preserve some of the properties of single-cell-based models, but which have the advantage that they can process information from medical images. Combining mechanisms on a cellular level and treating them with statistical sampling method within a voxel-based geometrical patient model provides the first step towards multi-scale models. Although some approaches in developing such models have already been made in the past, they have not been translated into clinical application yet.

In this thesis, we aimed at investigating and retrospectively validating computer based models for tumor response to radiation therapy applied to the cervical cancer, a malignant neoplasm located in the cervix. Imaging data of 9 patients were provided by IEO (European Institute of Oncology, Milan, Italy). For each patient, the dataset was constituted by one planning CT study (Computed Tomography) along with about 10 CBCTs (Cone Beam CT) studies, acquired in correspondence of the irradiation treatment fractions. The CBCTs served as control point for training and prediction evaluation. In the

patient dataset, two different groups, who differ for cancer type and therapies, were identified. First group included 5 patients affected by squamous cell carcinoma who were administered with radiation only. Second group included 4 patients affected by adenocarcinoma (AC) who were administered with radiation and chemotherapy.

Two different modeling approaches were studied: 1) cell-based discrete model, implemented through Cellular Automata; 2) tissue-level continuous model, implemented through ODE. Due to high computational burden, the cell-based discrete model underwent just preliminary simulated tests and was disregarded from extensive validation focusing the main work on the tissue-level continuous model. The tissue-level continuous model used Gompertzian or Logistic functions for the tumor growth while three different formulations were considered to represent the relation with the dose: a) classical linear/quadratic model; b) exponential; c) linear. The model training was implemented using an Evolutionary computation approach, it was repeated for different formulation of the tumor growth and therapy response terms, and performed across the two different patient groups. The tumor shape was simplified as a spheroid evolving throughout time. Volume and area of the tumor mass, measured in the CBCT datasets (expert manual segmentation and automatic 3D surface reconstruction) and predicted by the model, were chosen as training cost functions and validation quantities. Tumor volume regression rate has been suggested in literature as an important predictor of local control and long-term survival of the patient.

The residual errors of predicted volume and area for the first groups of patients were (absolute RMS) 5.3 cm^3 (relative error: 27%) and 10.4 cm^2 (relative error: 24%), respectively. Interestingly, the Gompertzian model combined with the exponential one provided best predictive results in both tumor volume and area. Smaller values were obtained in the second group, namely, 4.6 cm^3 (relative error: 21%) for the volume and 7.1 cm^2 (relative error: 16%) for the area. A sensitivity analysis was performed by studying the stability of the tumor mass volume-area prediction to the perturbation of specific model parameters. The natural tumor growth rate, represented by the first coefficient of each model, resulted generally the most critical, but in some cases after a while the weight of the therapy effects resulted to be the major cause of the prediction variation. In order to test extrapolation properties of the models, leave-one-out validation was considered. It consisted in removing one patient at a time from its group, and in training the model across the remaining patient datasets. As expected, the prediction error on the one-out patient was generally larger than on the rest of the group. Interestingly, for some patients, the prediction ability was comparable or even better than that of the remaining patient in the group. For example in the first group, considering the area results, the mean was about 25% while it reached 40% on the test, and the variance among the results was less than 0.05 on the first while above 0.25 on the second one. As expected, the quality of the prediction on the untrained patient was much poorer since the model fitted a small amount of data.

As a first consideration, the obtained prediction results were promising as they were in the range of the data uncertainty, qualitatively about 10% and 20% for area and volume, respectively. Noteworthy, such results were basically coherent with recent domain literature (Huang Z. et al. 2010) where relative prediction errors of about 25% were reported. As far as model training and validation is concerned, results highlighted that volume and area prediction criteria conflicted in between for the second group. In detail, while the combination of the Logistic curve and the linear/quadratic model turned out to be the most performing on the area, the volume was better fitted using the Gomperzian

model along with the linear model of the radiation response. The different behavior of the two indexes can be motivated by considering the effect of the chemotherapy on the overall treatment. The drugs are indeed delivered by mean of blood vessels that are more densely localized on the outer surface of the tumor and this peculiarity could cause an increase in the relevance of the events occurring to the most superficial layer of cells. While present results are not decisive, this finding reinforces the concept that the representation ability of the model is fundamental as well as the cost function to minimize for model parameter assessment (training). This could play a significant role in elaborating more complex predicting approaches.

A number of issues in this work must be highlighted. First, we represented tumor growth and therapy response, using tissue-level models implemented through continuous equations (ODE), thus limiting the overall process representation to just macro-scale mechanisms. Micro-scale processes at cellular level as for example cell cycle, cell-to-cell interaction, and tumor cell invasion weremodeled using cellular automata but the obtained models were not validated. Second, in the mathematical representation of tumor evolution we neglected the spatial dimension and only scalar validation criteria (area and volume) were adopted, approximating the tumor shape as a spheroid. This allowed to avoid registration across CBCT studies but loosing tumor progression in a common coordinate system. Third, the patient cohort was limited in size and this prevented us from getting any statistical evidence about the value of the prediction results and model performance. Fourth, the uncertainty on the data used for validation was just qualitatively evaluated. This aspect is however critical for any consideration about the model training feasibility and performance. Fifth, in the model training, the nominal isodose, administered to the patient, was used. However, the actual dose can be different due to tumor shape modification, soft tissue deformation, physical interaction with the surrounding organ and gravity effects. In order to overcome this drawback, actual dose computation should be carried out exploiting deformable registration between the planning CT volume, in which the nominal does is calculated, and the actual CBCT volume taken at treatment time.

Nonetheless, despite such shortcomings, the presented work showed how also not complex tissue scale models can be effective in predicting the interaction between tumor progression and the effects of radiation therapy. In conclusion, while being a preliminary study, the results are meaningful and promising for extending the methodology to include more sophisticated representation of the biological processes and therapy interaction, and making the approach suited for more extensive validation and future clinical translation.

Italian abstract

Nella lotta contro il cancro, la radioterapia rappresenta un trattamento localizzato altamente efficace, nel quale la guida per immagini (Image Guided RadioTherapy, IGRT) svolge un ruolo fondamentale. Le immagini con qualità diagnostica sono acquisite giornalmente, allo scopo di appurare la presenza di cambiamenti anatomici e valutare la necessità di una riorganizzazione del trattamento del paziente. Inoltre, l'imaging funzionale per la visualizzazione delle caratteristiche tumorali (ad esempio consumo di glucosio, ipossia, ripopolamento) è applicato, non solo per l'iniziale staging tumorale, ma anche a intervalli di tempo definiti durante il trattamento frazionato. Nonostante l'avanzamento nel riconoscimento dei meccanismi fisici e biologici legati alla radioterapia, la predizione tempestiva della risposta del tumore e degli effetti tossici rimane difficoltosa da raggiungere, in particolare a livello del singolo paziente. Negli ultimi 10 anni, si è tentato di applicare modelli alla predizione di questi eventi sui singoli casi di trattamento, tuttavia l'introduzione di approcci computazionali nella pratica clinica è stata impedita dalla mancanza di validazioni convincenti.

Per quanto riguarda la predizione dello sviluppo tumorale e degli effetti delle terapie, molti sforzi sono stati compiuti per rinforzare le potenzialità cliniche dei modelli in silico attraverso la loro traslazione nella pratica clinica per l'ottimizzazione e personalizzazione del trattamento. La modellizzazione biologica della risposta tumorale tuttavia ricopre un ruolo molto ridotto nei sistemi commerciali di pianificazione del trattamento, che si limita ai classici modelli di probabilità del controllo tumorale. Questi modelli compiono le loro predizioni attraverso parametri d'ingresso basati su popolazioni. A causa delle dinamiche delle caratteristiche del tumore, questo non è sufficiente a predire lo sviluppo dei tumori sui casi individuali. L'eterogeneità dei tumori, causata dalla differente radio-sensitività delle sottopopolazioni cellulari e dalla variazione di parametri ambientali come l'ossigenazione, impedisce predizioni attendibili paziente specifiche. Questa limitazione può essere superata unicamente considerando sistematicamente sia modelli meccanici dello sviluppo spaziotemporale del tumore che parametri tarati sul singolo individuo. Per raggiungere questo , sono stati sviluppati specifici modelli computazionale, in grado di simulare la crescita spazio temporale della massa tumorale, basati su principi radiobiologici. Modelli di interazioni tra le singole cellule sono stati sviluppati utilizzando i principi dei Cellular Automata e sono stati applicati allo studio di diversi problemi radio-biologici. Sebbene questi modelli possano essere usati per simulare i meccanismi che regolano la risposta alle radiazioni, sono limitati nell'utilizzo clinico dalla difficoltà di includere informazioni morfologiche e funzionali sufficientemente accurate tratte da immagini mediche. Questo porta allo sviluppo di modelli basati su voxel, che conservano alcune delle proprietà dei modelli a singola cellula, ma nei quali si ha il vantaggio di poter utilizzare le informazioni derivanti dall'imaging clinico. Combinando meccanismi a livello cellulare con metodi statistici all'interno di un voxel dalle dimensioni ridotte, si compie un passo verso l'approccio multi-scala. Nonostante alcuni tentativi in questa direzione siano stati approcciati in passato, non sono ancora stati tradotti in pratica clinica.

In questa tesi intendiamo investigare e validare retrospettivamente modelli computazionali sulla risposta tumorale alla radioterapia applicati al cancro della cervice uterina . Dallo IEO (European Institute of Oncology, Milano, Italia)sono stati forniti dati di imaging relativi a 9 pazienti . Per ciascun pazente , il dataset era costituito da uno studio CT (Computed Tomography) insieme a circa 10 studi CBCT (Cone Beam CT), utilizzati come punti di controllo per il training e la valutazione della predizione. Nel dataset dei pazienti sono stati identificati due distinti gruppi. Il primo gruppo includeva

5 pazienti affetti da carcinoma squamocellulare (SC) i quali sono stati trattati unicamente per mezzo di radiazioni. Del secondo gruppo facevano parte 4 pazienti colpiti da adenocarcinoma(AC) che hanno subito sia trattamento radioterapico che chemioterapico.

Sono stati analizzati due differenti approcci modellistici: 1) modello discreto a livello cellulare implementato attraverso la tecnica ei Cellular Automata; 2) modello continuo a livello tissutale che prevede l'utilizzo di ODE. A causa dell'alto costo computazionale, il modello basato sulle interazioni tra cellule è stato spinto solo fino a test e simulazioni preliminari, mentre la validazione estesa è stata focalizzata unicamente sul secondo tipo di approccio. Il modello su scala tissutale utilizzava funzioni Gomperziane o Logistiche per la crescita tumorale mentre sono state considerate tre possibili formulazioni per rappresentare la relazione con la dose: a) classico modello lineare/quadratico; b)esponenziale; c)lineare. L'addestramento del modello è stato implementato utilizzando tecniche evolutive ed è stato ripetuto per le diverse combinazioni delle formulazioni del termine di crescita tumorale e di risposta alla terapia per entrambi i gruppi. La forma del tumore è stata semplificata considerandolo uno sferoide in evolvuzione nel tempo. Il volume e l'area della massa tumorale, ricavata dal dataset delle immagini CBCT (attraverso segmentazione manuale di un esperto e ricostruzione automatizzata della superficie 3D) e predetta dal modello, sono state utilizzate nella funzione di costo per l'ottimizzazione del modello e per la sua validazione. Il tasso di regressione del volume tumorale è stato presentato in letteratura come un importante indicatore di riuscita del controllo locale e predittore della sopravvivenza a lungo termine del paziente.

Gli errori residui relativi a volume ed area predetti per il primo gruppo sono (in valori assoluti RMS) 5.3 cm^3 (errore relativo: 27%) and 10.4 cm^2 (errore relativo: 24%), rispettivamente. È interessante notare come il modello Gomperziano combinato con quello esponenziale abbia fornito i risultati migliori sia per il volume che per l'area. Valori più bassi sono stati ottenuti per il secondo gruppo, ossia, 4.6 cm^3 (errore relativo: 21%) per il volume e 7.1 cm^2 (errore relativo: 16%) per l'area. Un'analisi di sensitività è stata svolta studiando la stabilità della predizione di area e volume in risposta alla perturbazione di specifici parametri. Il tasso di crescita naturale del tumore, rappresentato dal primo coefficiente di ciascun modello, è risultato generalmente il più critico, ma in alcuni casi durante il corso del trattamento il peso degli effetti della terapia è risultato essere la maggiore causa di variazione nella predizione. Allo scopo di testare la capacità di estrapolazione del modello è stata utilizzata la tecnica di validazione leave-one-out. Questa consiste nel rimuovere on paziente alla volta dal suo gruppo ed addestrare il modello sui rimanenti. Come ci si aspettava, l'errore di predizione commesso sul paziente lasciato fuori è risultato generalmente maggiore di quello ottenuto mediamente sul resto del gruppo. Interessante notare come per alcuni pazienti rimossi la capacità predittiva sia risultata comparabile, o addirittura migliore, a quella del resto del gruppo. Per esempio nel primo gruppo, considerando i risultati ottenuti per l'area, la media dell'errore è risultata del 25% mentre raggiungeva il 40% per i pazienti di test, e la varianza, molto contenuta tra i primi (0.05) era molto più elevata nei secondi (0.25) Come ci si aspettava, la qualità della predizione sui pazienti non utilizzati per l'addestramento è risultata ridotta dal momento che il modello aveva a disposizione solo pochi esempi, insufficienti per una corretta generalizzazione.

Come prima considerazione, i risultati di predizione ottenuti sono stati promettenti poiché non lontani dal range di incertezza dei dati , qualitativamente intorno al 10% e 20% rispettivamente per area e volume. Per quanto riguarda i modelli, i risultati hanno evidenziato una discrepanza tra il criterio di

predizione di area e volume nel secondo gruppo. In dettaglio, mentre la combinazione del modello lineare/quadratico e della funzione Logistica si è rivelata la più performante per quanto riguarda l'area, nel caso del volume il fitting migliore è stato ottenuto con la curva Gomperziana combinata con la risposta al trattamento lineare con la dose. Il diverso comportamento può essere spiegato considerando l'effetto ella chemioterapia sul trattamento complessivo. I farmaci sono trasportati per mezzo dei vasi sanguigni maggiormente presenti sulla superficie esterna della neoplasia, questa caratteristica potrebbe portare ad una particolare rilevanza dei fenomeni che si presentano nello strato cellulare più esterno. Sebbene questi risultati siano preliminari, è possibile supporre che oltre alla tipologia di modello anche il metodo di ottimizzazione e validazione giochino un ruolo importante nella realizzazione di approcci predittivi più complessi.

Vanno evidenziate alcune criticità di questo lavoro. Per prima cosa, abbiamo rappresentato la crescita tumorale e la risposta alla terapia usando modelli continui a livello tissutale, implementati per mezzo di equazioni differenziali (ODE), limitando la rappresentazione ai meccanismi presenti su scala macroscopica. I processi su micro sala, come il ciclo cellulare e le interazioni tra cellule sono stati studiati attraverso l'approccio dei Cellular Automata ma il modello ottenuto non è stato validato. In secondo luogo, nella rappresentazione matematica dell'evoluzione tumorale abbiamo trascurato la spazialità, ma abbiamo fatto riferimento unicamente a grandezze scalari (area e volume) approssimando la massa tumorale ad uno sferoide. Questo ci ha permesso di evitare la registrazione dei vari studi CBCT tra loro, ma ci ha fatto perdere la progressione morfologica del tumore in un sistema di riferimento cartesiano univoco. Terzo, la dimensione limitata del dataset ci ha impedito di ottenere evidenze statistiche significative riguardo il valore dei risultati di predizione e performance del modello. Quarto, l'incertezza sui dati utilizzati per la validazione è stata valutata solo qualitativamente. Questo aspetto è tuttavia fondamentale per qualunque considerazione riguardo i risultati della predizione, una valutazione oggettiva del rumore sui dati sarebbe indispensabile alla valutazione della qualità dei risultati ottenuti. Quinto, nell'addestramento del modello è stata utilizzata la dose nominale somministrata al paziente, tuttavia la dose reale può differire sensibilmente dalla precedente a causa di modificazioni morfologiche del tumore, deformazioni dei tessuti molli e fisiologici, spostamenti degli organi circostanti e altri fenomeni. Per superare questo problema, bisognerebbe sfruttare una registrazione con deformazione per calcolare la dose effettivamente rilasciata nella regione di interesse e comprendere quale sia il volume realmente trattato.

Tuttavia, nonostante questi inconvenienti, il presente lavoro mostra come, anche modelli non complessi e su scala tissutale, possano essere efficaci nel predire l'interazione tra la progressione tumorale e gli effetti della radioterapia. I principali risultati della predizione sono stati essenzialmente coerenti con quanto rintracciabile nella recente letteratura (Huang Z. et al., 2010) dove sono stati riportati errori relativi di circa 25%. In conclusione, pur essendo uno studio preliminare, i risultati ottenuti sono significativi e promettenti per estendere la metodologia, includere rappresentazioni più sofisticate dei processi biologici e delle interazioni dovute al trattamento, e rendere il modello adatto per una validazione più estesa con la conseguente traduzione in pratica clinica.

Summary

Problem

Radiation Oncology is the one of the most common therapy, used in the treatment of 40% of patients cured of cancer. Nowadays the radiotherapy planning and delivery rely on medical imaging, actually several technological advances, such as Image Guided Radiotherapy (IGRT), have become part of the clinical practice. The huge amount of data, stored in the databases, is a priceless source of information. Cancer development and response to therapy can generally be thought as a system evolving in time and space, therefore many efforts have been spent to apply all the mathematical and statistical approaches to this problem in order to describe it in the most complete and realistic way. Being able to predict the development of the disease or even the outcome of a therapy could help the physician to make a decision about the best treatment for every patient. Despite their obvious advantages this mathematical models play a small role in the medical environment. What limits their use in the clinical practice is the lack of an objective and quantitative validation of the quality of their performance. The aim of this work is to propose two different modeling approaches: 1)the continuous one, developed using Ordinary Differential Equations (ODE), considering the tissue level only; 2) the discrete one, based on the Cellular Automata technique, able to describe the single cell interaction. We decided to apply our research to some real clinical cases provided by the IEO (European Institute of Oncology, Milano, Italy) of patient affected by cervical cancer.

Cervical cancer is a malignant neoplasm located in the cervix (Latin for neck) uteri . The cervix, cylindrical anatomic element, is the lower portion of the uterus and consists basically of two structures: the ectocervix that lays in the outer part and it is paved with epithelial cells and the endocervix, that consist of columnar cells. Cervical Adenocarcinoma affects the inner part of the cervix (endocervix) while Squamous cell Carcinoma, which is the most frequent kind of cervical cancer, in fact it represent the 80% of all cases, develops in the ectocervix (American Pathologists, 2011). It usually reacts well to therapies but there is more than one possible approach, therefore it might be useful to predict the possible outcome of each kind of treatment.

Mathematical models background

The models can be subdivided in two main classes, the Ordinary Differential Equation (ODE) models and the Cellular Automata (CA) models (Boondirek A. et al., 2006; Hwang M. et al., 2009). The ODE models are able to catch the global behavior of the system looking at the evolution of a macroscopic measure. They start from the knowledge of general rules and carefully try to put them in formulas in order to get a global representation of the problem. Anyway they consider only the evolution trough time of scalar quantities and have no link to the spatial domain were the tumor is developing; one of the most well-known ODE model for cancer application is the reaction-diffusion model. A more complex analysis can be carried on

by means of Cellular Automata models that use a bottom up approach, the single discrete units (cells) interact leading to complex pattern formation unpredictable from the basilar rules driving the single cells. There are also hybrid experiments that put together the main characteristics of the two classes for example using the continuous models to represent the role of nutrient and oxygen together with a discrete one describing the tumor development itself. Moreover there are complex models able to study the problem at several levels of detail, from the molecular to the macroscopic point of view (Stamatakos G. S. et al., 2009). The mayor issue in mathematical modeling is the lack of validation and error estimation which prevents these tools from being used in clinical practice. Nowadays the evolution history of many tumor has been successfully documented by imaging techniques. It is often a major requirement to visualize the region of interest in order to plan the therapy and check the evolution of the disease, the real data acquired for these purposes can be an useful source of information.

Project aim

A set of CT acquisition, belonging to 9 different patient divided into two groups according to the kind of tumor and treatment, were provided by the IEO (European Institute of Oncology). The images are unregularly acquired on a period of time up to 80 days while the therapy was delivered. Since the CT resolution did not allow a precise reconstruction of all the structural elements, as for example the net of ducts, we decided to implement both the discrete and continuous models but to focus on the ODE representation of the problem as far as it concerns the validation part of the project. Our goal is to build a model able to fit the real data on which we will train him and also some new one belonging to the same class of patients. The ODE model will consider the tumor as a sphere evolving trough time varying its radius, therefore volume and area should keep a constant ratio. In order to fit and validate the model we will analyze the dynamic of both volume and area of the tumoral mass. The analysis of the evolution considering both the values can prove or invalidate the radial symmetry assumption. It is reasonable to expect some differences among the classes of patients in terms of prediction's errors and model's coefficient. It will be interesting to see if some type of models are more suitable to represent a certain group of patients than others, how the parameters change in presence or absence of chemotherapy or accordingly to the type of tumor, how the error varies inter and intra classes, if the model is able to reveal outliers (if there are any) and so on. In the end the best models for each class will be presented and an error estimation of the prediction will be provided comparing it to an approximated estimation of the data-noise.

Data and protocol

The subjects have been divided into two groups:

- 5 patients affected by squamous cell carcinoma (**SC**) treated by means of radiation therapy (**RT**) only;

- 4 patients affected by adenocarcinoma (**AC**) administered with radiation therapy (**RT**) and chemotherapy (**CT**).

Every patient underwent a CT (computed tomography) scan, usually a couple of week before the first treatment session. A radiation therapy plan was developed on this first acquisition and several CBCT (cone beam CT) were taken, one at any scheduled session of therapy. All the acquired images are Dicom files and consist of the volume itself plus an header in which the information about the patient, the date of treatment and other meaningful data are stored. The main structure is a 3D matrix containing the grey level of the pixel corresponding to a specific point of the volume. There are also some extra fields containing the contours of the region of interest. The time schedule of the radio-therapeutic treatment corresponds to the dates at which the CBCT images were acquired. The planned dose is specified in another 3D matrix where the values of each pixel correspond to the dose delivered to the corresponding point in the CT volume. Usually 1.8 Gy or 2 Gy were delivered during each session, and the treatment lasted up to 80days from the date the planning CT image was acquired (from 25 to 28 fractions). Only some of the patients were administered with both radio and chemotherapy, moreover they did not get the same drugs, (generally Cispaltino and Paclitaxel), in the same amount and with the same schedule (once a week, 5 or 6 cycles). However the chemotherapy treatment in our representation was thought to be only an adjuvant therapy, that made the tumor more sensitive to the irradiation. Since the drugs are delivered by the bloodstream, they may affect more the superficial layer of the mass than the inner part because of its resistance to perfusion. The images had to be elaborated in order to extract the volume and area information, anyway a bigger effort would have been required in order to make them compatible with the C.A. model requirements.

The elaborations can be subdivided into the following steps:

a) The GTV (Gross Tumor Volume) was countered by the physicians;
b) A binary volume was derived from the CT and CBCTs, each voxel was considered either healthy or not depending on its location (outside or inside the GTV, respectively);
c) A triangulation of the point belonging to the contours was accomplished;
d) The global volume and surface were computed from the binary image and the triangulation respectively.

The binary image was obtained from the acquired one setting all the voxels corresponding to the Gross Tumor Volume to one and the others to zero. This new image was saved in mha format and it contained the tumor volume in white on a black background. In this way we were able to calculate the overall number of cancerous cell summing up the values of each voxel of the binary image. The voxels were parallelepipeds whose sides length corresponded to the image spacing (Δx, Δy and Δz respectively). The volume (ΔV) of one volumetric unit was therefore the product of the three dimensions, while the overall volume (V) could be calculated, as showed in Equation 1, as the product of ΔV and the number of unhealthy cells (N).

$$V = \sum \Delta x * \Delta y * \Delta z = N * \Delta x * \Delta y * \Delta z = N * \Delta V \quad (1)$$

In order to understand if the mass grows in the same way in every direction or elongates and changes its shape we decided to calculate the overall area of the tumor and to compare it to the volume.

From the points of the GTV contour, a Delaunay triangulation was obtained, and the area was calculated summing up all the areas of the single triangles, than it was possible to calculate the volume over area ratio as in Equation 2.

$$\text{ratio} = \frac{V^{\frac{1}{3}}}{A^{\frac{1}{2}}} \quad (2)$$

If the mass had a spherical shape its volume could have been calculated as showed in Equation 3 and the area as in Equation 4, where r is the radius.

$$V = \frac{4}{3} * \pi * r^3 \quad (3)$$

$$A = 4 * \pi * r^2 \quad (4)$$

In this case the ratio would have been a constant value through time (0.4547) a different constant value would imply an elongation but would also suggest that the tumor keeps its shape during the treatment, while a variation in the ratio through time could invalidate our initial assumption of a spheroidal mass.

Data-error estimation

The contours can be affected by different kind of errors:

1. The accuracy of the acquired data and the dimensions of the voxels
2. The uncertainty in the segmentation of the images due to the physician precision in the choice of the key points of the contour
3. The approximation of the tumor to a convex structure made in order to calculate the area of the tumor from the triangulation.

Since there was no possibility to implement a protocol able to objectively assess the average error, especially on the physician manual work, we decided to calculate an approximation of

the data-noise starting from the assumption of radial symmetry. We calculated the radius of a sphere having a volume equal to the collected one, afterwards an average error of ±1mm on the radius was assumed and the consequent error on the volume and area was computed. The image spacing varies between 0.87mm and 3.0mm with our approach the maximal uncertainty on the radius is supposed to be 2 mm, which is within this range.

Model definition and data fitting

First of all it was necessary to choose how to face the problem. We proposed two different kind of approaches, one discrete, able to describe the cellular level, implemented using the Cellular Automata technique, and one continuous, that provides only a macroscopic description of the problem using the Ordinary Differential Equation (ODE).

The Cellular Automata model was implemented using MatLab®. The basic structure was a 3D grid of voxels containing a certain number of cells. The tumor growth dynamic was obtained subdividing the problem into two parts: the single voxel evolution, that includes the cell cycle and the damaging effects of the therapies, and the tissue evolution, that mimics the cell interactions (e.g. tumoral cell invading healthy cells). Anyway it was not possible to evaluate its performances because of the lack of accuracy of the clinical data and the huge computational burden of such a detailed simulator.

The ODE model consisted of two terms, as showed in Equation 5. Where $f(t,V)$ represents the independent growth term while $g(t,V)$ is the decay term due to therapy effects, P can be either the volume or the surface.

$$\frac{dP}{dt} = f(t,P) + g(t,P) \tag{5}$$

The two formulations that $f(t,V)$ could acquire are the one described by the Gompertzian (Equation 6) and Logistic (Equation 7) curves, where a is the growth factor and b the asymptotic value for the index P (volume or area).

$$f(t,P) = a * \log\left(\frac{b}{P}\right) * P \tag{6}$$

$$f(t,P) = a * P * \left(1 - \frac{P}{b}\right) \tag{7}$$

While the $g(t,V)$ function was supposed to be as simple as a linear function of the dose with only one parameter, or an exponential one with two unknown coefficients up to a more

complex function developed from the linear quadratic model (Swanson K. et al., 2009) using three different coefficients.

A genetic algorithm implemented in MatLab ™ was used in order to compute the best parameters of the model Each gene consisted of the sequence of parameters that had to be set, they were expressed in binary code to allow the most common mutation techniques. During the evolution the genes who got the best results of the fitness function were selected and randomly mutated in order to explore new combinations even far from the starting point or they were mixed using the crossover techniques that does not add new information but allows to tune the parameters. In this case the fitness function (z) was linked to the sum of the errors (ε) due to difference between the real volume size ($\bar{V}_i$) and its prediction (V_i) obtained using the model (Equation 8). Since we wanted the fitness function to be maximal when the error was as small as possible, we defined $z=1/\varepsilon$

$$\varepsilon = \frac{1}{N}\left(\sum(V_i - \bar{V}_i)^2\right)^{1/2} \quad (8)$$

The computational effort was quite considerable because for each gene, that is each combination of parameters, a numerical ODE solver had to be called in order to calculate the model outcome, but it was sensibly reduced comparing it to the one required by the discrete model. This approach should lead to the values of the parameters that best fit the data but it could get stuck in a local maximum in case the rate of mutations is too small, or it could escape from the right direction of maximization if each generation is too much different from the previous one. In order to check if the solution is reliable we tried the code several times varying the settings of the genetic algorithm (e.g. population size, number of generation, mutation rate etc.).

Results

The best model is the one that minimize the error ε (Equation 8).

The residual errors on the volume and area prediction for the first groups of patients were (absolute RMS) 5.3 cm^3 (relative error: 27%) and 10.4 cm^2 (relative error: 24%), respectively. These results were achieved using the combination of the Gompertzian curve and the exponential therapy module both for Area and Volume .

Smaller values were obtained in the second group, namely, 4.6 cm^3 (relative error: 21%) for the volume and 7.1 cm^2 (relative error: 16%) for the area. These patients are affected by Adenocarcinoma cervical tumor and their treatment required both radiotherapy and chemotherapy . The combination of the Gomperzian curve and the third therapy model resulted the most performing as far as it concerns the volume, on the other hand , the area evolution pattern was better represented by the logistic model and the first therapy-module. This is the only case where the best outcomes were achieved using the logistic curve, and the

difference between the volume and area results could be related to the localized effect of chemotherapy, which is delivered by blood and may affect more the outer layer of cells than the rest of the tumor.

Stability analysis and model validation

The Linear Sensitivity analysis was performed, it consists of varying one coefficient at a time in order to analyze the effect of a wrong calibration of a specific parameter on the overall error. Each coefficient was perturbed of ± 2% or ±5% and the variation of the Volume or Area prediction over the variation of the coefficient was computed. Noteworthy the same results have been achieved using the two percentages. The obtained curve can be thought as a derivative, it was normalized by the value of the prediction obtained using the unperturbed coefficients. It is interesting to notice that the error of a wrong parameter setting increase trough time and that the critical coefficients are the one representing the growth rate of the uncontrolled tumor, and the one that weights the effect of the therapy. This two parameters have to be carefully set in order to get reliable results.

The validation leave-one-out is based instead on removing one patient at a time from its class, calculating the best parameters as already done before, and testing the solution on the one left out. This approach is needed order to understand if the solution found is general enough to predict the evolution of the tumor in a new patient with the same pathology and administered with the same therapy, without using their data to train the model. It turned out that the error on the patient left out was generally larger than the one that affected the rest of the class, but if we analyze every single case we can see how some models were able to fit the new data better than the one they were trained on, this is the reason why the error variation was generally bigger among the test patients than on the training class.

Conclusions

The preliminary results are very promising, actually the values of the errors are close to the range of the data uncertainty, qualitatively about 10% and 20% for area and volume, respectively. Moreover, such results were basically coherent with recent domain literature (Huang Z et al., 2010). Here we summarize the most critical issues of this work:

1. the tumor growth and therapy response were represented using continuous equations (ODE), limiting the description to the tissue level and macro-scale mechanisms. Micro-scale processes at cellular level, as for example cell cycle, cell-to-cell interaction, and tumor cell invasion were studied using the cellular automata model which was not validated though;
2. the spatial dimension of the mass and its location were neglected and only scalar variables (area and volume) were adopted, approximating the tumor shape as a spheroid. This allowed to avoid registration across CBCT studies but not to collocate the tumor progression in a common coordinate system;

3. the patient cohort was limited in size and this prevented us from getting any statistical evidence about the value of the prediction results and model performance;
4. the uncertainty on the data used for validation was not evaluated in a rigorous way which is fundamental in order to assess the performance of the model ;
5. the nominal value of the dose delivered to the patient was used despite the possible occurrence of errors due to tumor shape modification, soft tissue deformation and physical interaction with the surrounding organs.

In order to overcome this limits more patients and more accurate data would be needed, including the dose map, functional images and histological information. In this way it could be possible to validate the discrete model implemented as well. It would also be interesting to implement a patient-specific model using the ODE , in order to provide some meaningful information to the physician at run time so that he can adjust the schedule of the therapy while it is delivered.

1 Introduction

This project aims to give a contribute in the deployment of computer-based models for multi-scale tumor response in radiation oncology for clinical practice working on real data belonging to patients affected by celvical cancer. Cancer Modeling is quite a recent research field with a wide application possibility since the the 'in silico' reproduction of the growth dynamic of this disease, and its response to different kinds of therapies, might sensibly improve the efficacy of the treatments. In the last decades several modeling approaches have been selected in order to mimic the behavior of the mutated cells and to customize the therapy accordingly to the needs of a specific patient. They can be roughly subdivided into continuous, macroscopic models and discrete, microscopic ones, but the separation line is not always so neat, there are also some hybrid solutions, involving complex multi-scale techniques. In any case their purpose is to select and model the relevant factors in the tumor evolution in order to early predict the effects of a therapy; this will give the physician the chance to select the treatment that will lead to the best possible outcome in terms of health and life quality for the patient. A more extended dissertation on this topic will be provided in chapter 2. Currently, there is a mismatch between the large amount of multimodal image based (and non-image) based data available, and the human ability to realize accurate and trustworthy models, valuable in clinical decisions on treatment. This is the reason why this project starts from real clinical data belonging to cervical cancer patients, and propose some different modeling approaches in order to investigate their ability to catch the dynamic underneath the evolution of the disease.

1.1 Cervical tumors

Cervical cancer is the second most common malignancy in women throughout the world, with over 250,000 deaths per year worldwide; it is located in the cervix (Latin for neck) uteri.

The cervix, cylindrical anatomic element, is the lower portion of the uterus and consists basically of two structures: the ectocervix and the endocervix. The ectocervix is the outer portion of the cervix, it is reachable through the vagina, while the endocervix is the internal canal leading to the uterus. The aspect of the cervix varies among women in accord to age, hormonal state and other factors. The cervix (Figure 2) contains two types of cells, columnar cells, located in the endocervix, and squamous epithelial cells, that can be found in the more distal part, near the vagina. The columnar epithelial cells are organized in a single layer and secrete cervical mucus, a substance composed for the most part by water, that among other things helps spermatozoa's movements.

Cervical cancer is still a weighty burden in developing nations, while in the wealthier countries there has been a 75% decrease in its incidence and mortality over the past 50 years. This is due to the introduction of several screening programs. In 2008 Harald zur Hausen has been awarded with The Nobel Prize for his discovery about the relationship between human papilloma viruses (HPV) and cervical cancer. HPV is an incredibly common sexually transmitted disease, usually, the infections are temporary and fade away in 1 year (70%) or 2 years (90%). The problem arises in the young females, when the infection persists, because

there is high risk of developing cervical cancer. The HPV infections may cause genital warts but is not unlikely that an affected woman shows no symptoms at all, until the development of pre-cancerous lesion of the cervix. Prevention and early detection are very important, because the pre-cancerous lesion are treatable even if this might necessitate preventive surgeries with possible loss of fertility.

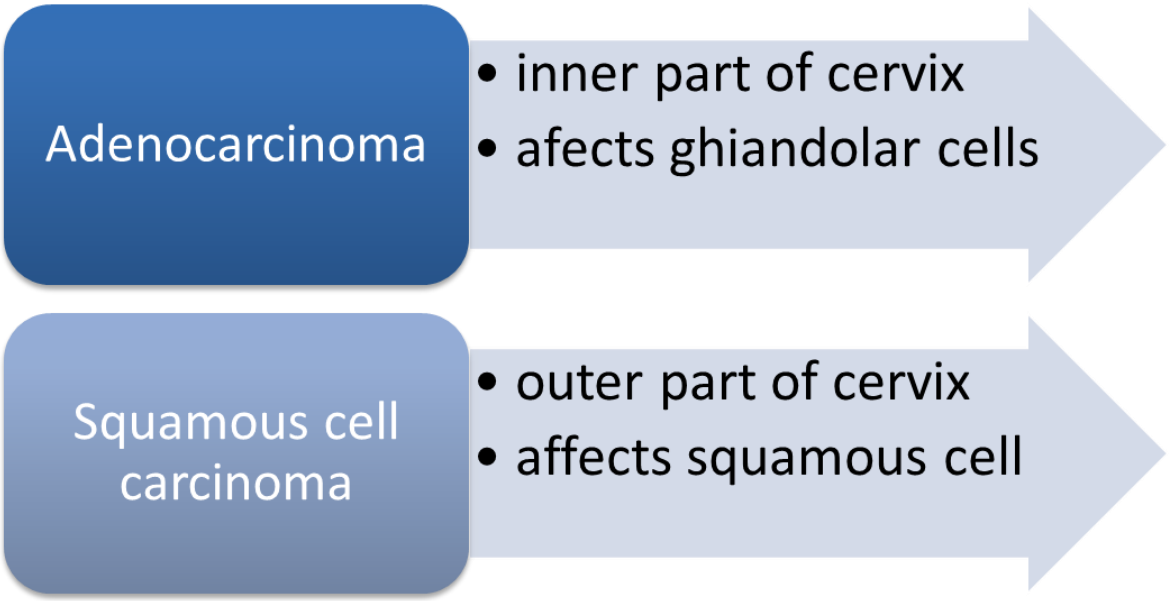

Figure 1
Cervical cancer classification

The most frequent kind of cervical cancer (represents about 80% of all cases) is named squamous cell carcinoma since it affects the squamous cells. Lately, since 1970s, another pathology has increased its incidence: the adenocarcinoma, a tumor that originates from cells that attack some glands in the cervix. Their characteristic are summarized in Figure 1. The quoted ones are definitely the most common types, anyway it exist a small percentage of cervical cancers which combine the two (adenosquamous carcinomas) and others even more rare ones that we will not mention (Dolinsky C. et al., 2011).

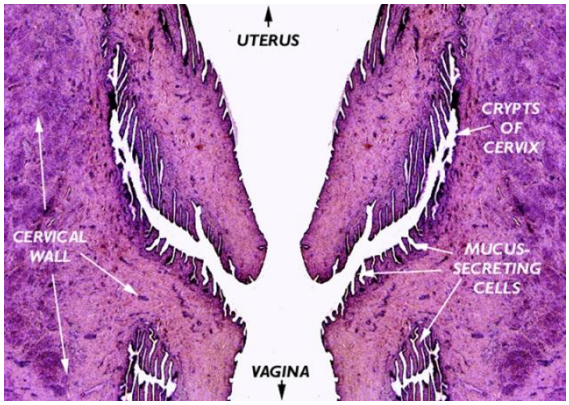

Figure 2
Cervical anatomy

1.2 Project motivation

The mathematical models are useful tools commonly used in many fields, and since a tumor can be seen as a mass evolving in space and time (Figure 3) it can be modeled as a system describing its properties and predicting its outcome. The enormous potential of this tool is the possibility to add as many terms or modules as needed to better represent the problem, but obviously the more we complicate the formulas the greater will be the computational effort.

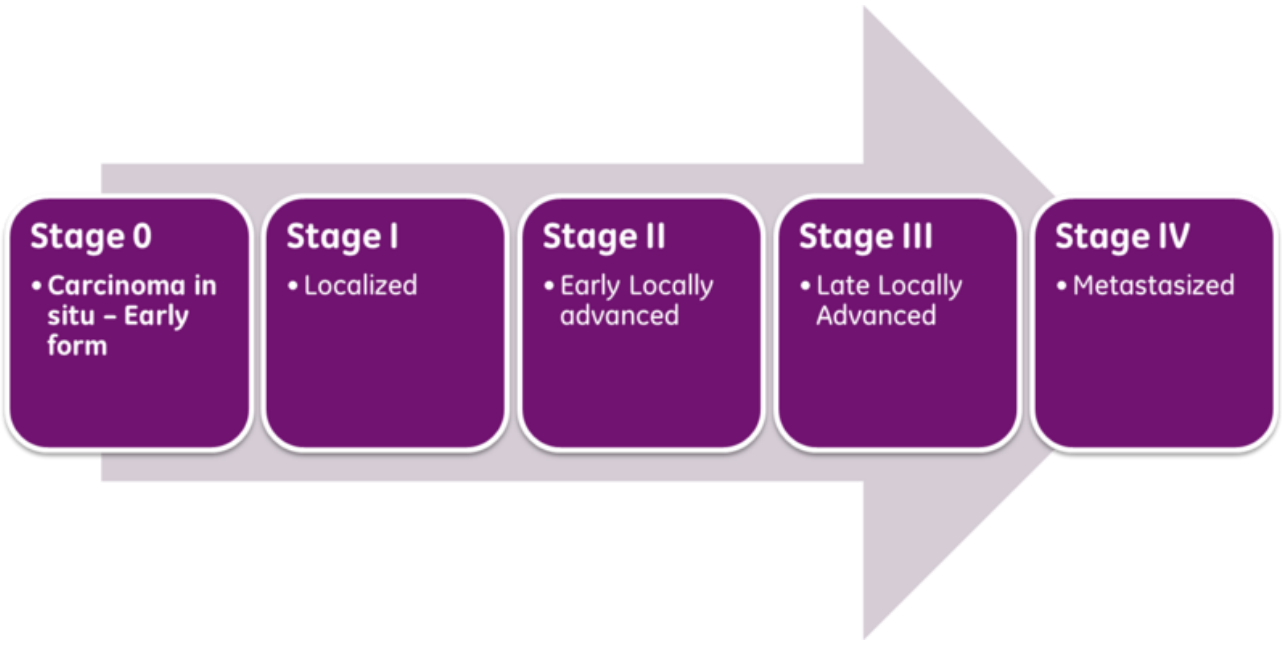

Figure 3
Cancer staging

There are many different possible cancer treatment plans, therefore the therapeutic decision taken by the oncologist may be a matter of life or death for the patient because different approaches may lead to very different outcomes. Looking at this context it is easy to understand the advantages deriving from a model able to predict the influences of surgery, radiotherapy , chemotherapy on the dynamic of the tumor, considering the doses or other meaningful parameters: they could effectively help the physician choosing the most suited therapy for the specific patient. The main reason that prevents these models from being applied, despite their potentials, is the lack of a convincing validation on real data able to prove their efficacy and the quality of their performances. The trials set using in vitro experiments are not sufficient to assess the quality of their results and to turn them into clinical practice. Lately the development of new medical devices like 'IMRT' , (Intensity Modulated Radio-Therapy) that is a kind of image guided radiation therapy, and many other imaging techniques caused the availability of daily acquired images stored in huge database. The idea of our work is to use this large amount of clinical data to build and validate a tumor growth model.

1.3 Project proposal and expectations

In this thesis two different approaches will be proposed, in order to compare their ability and issues. The first type of model that will be presented is based on Ordinary Differential Equations (ODE). This method allows the continuous fitting of some scalar variables evolving through time, as for example the volume, providing the physician an idea of the global system's evolution. The second one is a discrete approach that exploit the Cellular Automata technique in order to get an accurate spatial reconstruction of the problem at a cellular level. The project will consist of two main parts: 1) the formalization of the models 2)the data fitting (optimization) and validation. The second part will be applied only to the ODE models because of the lack of some histological information needed to inspect the problem to the microscopic/cellular level and the high computational burden. The parameters that will be monitored during the treatment period will be volume and area of the tumoral mass, the main assumption of our ODE model is a radial symmetry of the cancer. The availability of both volume and area will allow us to see the evolution of two different phenomena since the value of the volume is linked to the dimension of the necrotic core, while the area represents the layer of outer cells that proliferate and invade the surrounding tissues, moreover this will allow us either to support or to confute the assumption of spherical geometry. We also want to provide the level of uncertainty on the collected data and to compare it to the prediction error; a further step will be to test the stability of the model and the dependence of the result on the variation of the parameters in order to show the robustness of our software. Due to the subdivision of the patients into two classes we will be able to compare the models and results obtained for each of them . We might even be able to draw some conclusions about the possibility to infer some of the physical characteristics of the diseases from the structure of the model itself. We expect some models to fit best certain class of patients than others, the interesting part will be to make some hypothesis on the

reasons of this behavior and to link it to some physiological event or meaning. We want to define the best model for each analyzed class and to show the error made in the prediction comparing it to the level of data-noise. We will also try to assess the validity of the hypothesis underlying the ODE representation. In conclusion we will try to implement two different solutions able to predict the cancer evolution in presence of therapies, and we will validate one of them using real data in order to provide the confidence of the result achieved. This will be hopefully a starting point for developing more complex models and it will set a path for further analysis.

The IEO (Istituto Europeo di Oncologia, European Institute of Oncology, Milano, Italy) partnership will allow us to work on some real anonymous data belonging to two different groups of patient affected by cervical cancer. Moreover, thanks to the Atlantis project we will take advantage of the already existing cooperation with professor Mark Garbey (University of Houston) and his team which already developed an high level of expertise on the field of tumor modeling, this is why a short stay in the USA will be required to work under their supervision.

2 Background

In this paragraph will be introduced both the theoretical basis of the project and an overview about the previous attempts of tumor growth modeling, some possible classifications will be also suggested.

2.1 Microscopic, macroscopic and multi-scale models

All the works developed in the field of tumor growth modeling can be coarsely classified into two different groups: microscopic and macroscopic models (Konukoglu E. et al., 2008). The microscopic models are the most studied and concentrate on cellular level, this means that they take into account the cell cycle's phase, the presence of nutrients and oxygen and all the dynamics observable at a scale of a few microns. On the other hand macroscopic models focus their attention on the tissue level and the overall dynamics of the pathology. The way of juxtaposing the microscopic and the macroscopic description is one of the main open problem in this field. In order to build a microscopic model information on what happens at subcellular and cellular scale are required. It is not easy to obtain such data because of the invasiveness of the exams and the cost of the acquisition's tool. A biopsy, even more than one, would be needed which is not easily feasible. Anyway many in vitro experiments have been set to study the real behavior of the single cancerous cell and to understand how the nutrients are absorbed and so on. In fact this kind of models concentrate on the single cell evolution, regulated by some rules, and take into account the possible causes of mutation, the effects of external factors and many other parameters that can be meaningful in order to achieve a realistic representation. The models that focus on the cellular level are usually discrete since in such a small scale the stochastic effect and the discrete texture of the tissue cannot be neglected (Deroulers C. et al., 2009). The macroscopic models start from the big picture, catching the global dynamic of the system, and are usually subdivided in two main categories: Diffusive and Mechanical models. The diffusive models describe the diffusion flux of migrating cells and nutrients through concentrations, densities and their derivatives, while the mechanical models have to describe the mechanical properties of the involved structures in order to predict deformations and other effects of the tumor's presence. These tools have the obvious advantage of considering average values, therefore they do not need any special effort to get the data that can be easily extracted by imaging techniques as computed tomography scans (CT) , magnetic resonance imaging (MRI),single photon emission CT(SPECT), positron emission tomography(PET), cone beam CT (CBCT) etc. (Konukoglu E. et al., 2008). The limit of this approach is the level of detail it can reach, since it is linked to the spatial resolution (usually about $1mm^3$): it does not take into account the underpinning mechanisms regulating the macroscopic evolution of the disease. A tissue-level model, for example, can evolve ruled by differential equation considering the average cell density or the survival fraction pursuant to irradiation therapy or drugs administration. Therefore this models usually assume tumor cells to be continuously distributed and the formulations consist often of ordinary or partial differential equations able to describe the global growth process. During the last decades some discrete macroscopic models have been introduced from the first studies on lattice (Duchting W. et al., 1985) to early cellular automata models (Boondirek A et al., 2006). Dealing with a system constituted of a large number of cells, it could be reasonable to consider the statistical properties of the interactions which allow a model simplification as it happens in the kinetic cellular theory (Bellomo N. et al., 1999).

One of the most compelling problems at present is that, while the microscopic approach does not consider the overall dynamic, the macroscopic one on his own is not always sufficient to get an accurate description of the system. The necessity to realize a tool capable of dealing with all the different aspects rises the problem of putting together different timescales as well as different views (the microscopic and macroscopic), in other words the solution requires the implementation of a multi-scale model.

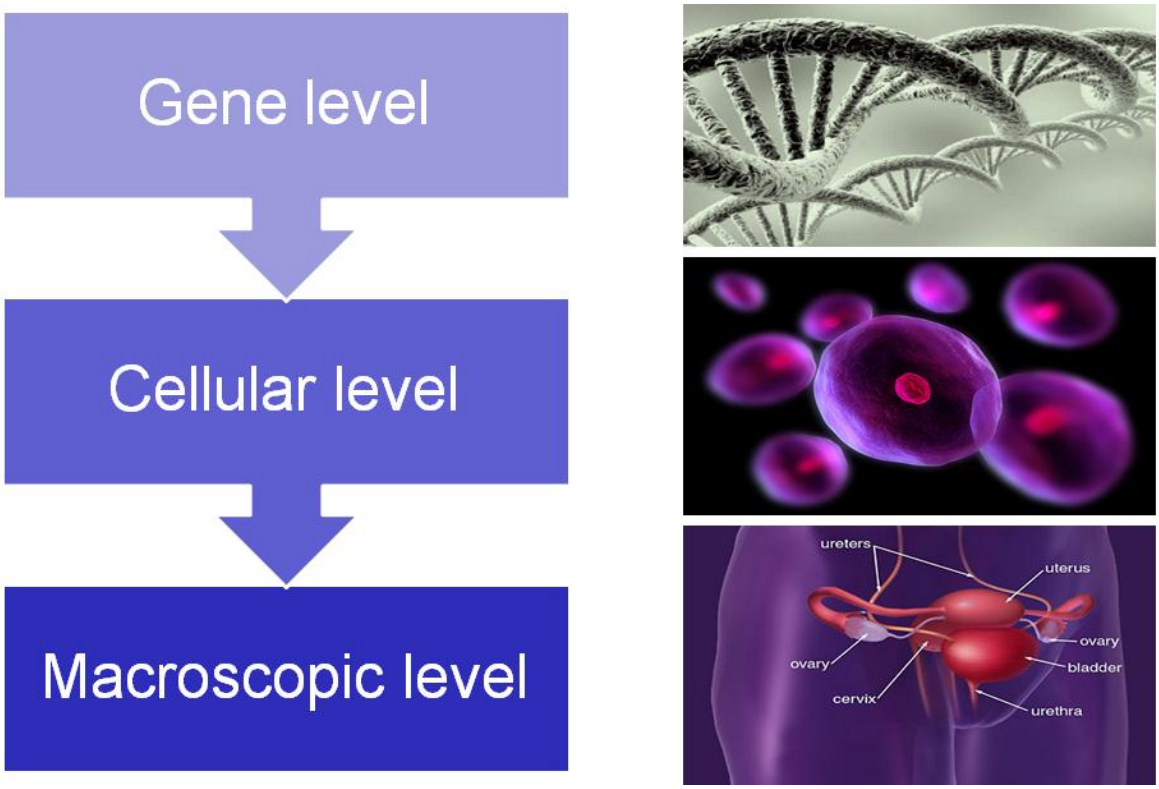

Figure 4
Level of detail taken into account during the modeling

As showed in Figure 4 there are three main levels:

- The gene level, that considers the distortions to the cell cycle as well as the presence of nutrients or even the response to stimuli and therapies (DNA modifications and damages).
- The cellular level, where the single-cell dynamic is analyzed, each cell can proliferate following the physiological cycle or ignore the normal control signals, mutate and compete with the immune system.
- The tissue level, that correspond to the macroscopic level, refers to events modeled as continuum phenomena: cell migration, diffusion, aggregation, formation of metastases and any other interaction with the surrounding.

The usage of a discrete model may allow the parallel representation of the different scales, an exemplum is the work of G. Stamatakos et al. (Figure 5).They described an hyper-matrix notation which permits the use of a composite operator able to work on different levels at the same time (Stamatakos G. et al., 2009).

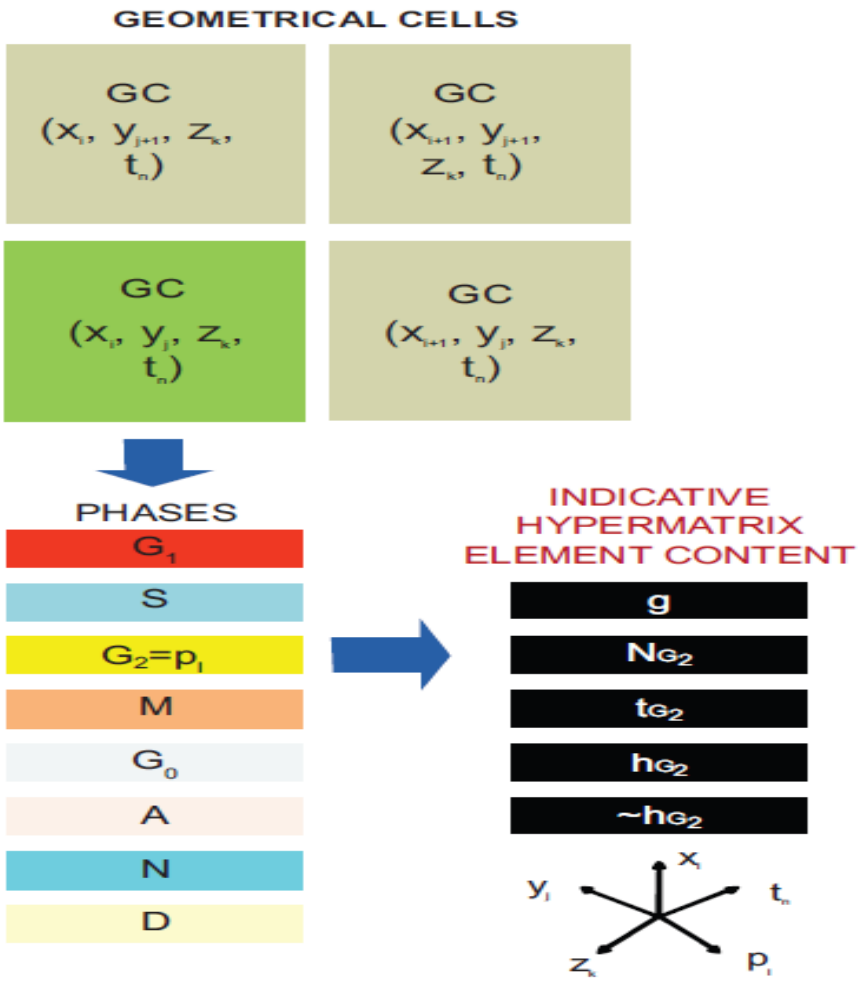

Figure 5
Hypermatrix notation **(Stamatakos G. et al., 2009)**

Another interesting work has been carried on by Ribba who approached the problem in a different way (Ribba B. et al., 2005). He and his team considered the antigrowth factors that regulate (healthy) cell density accordingly to supplies and oxygen availability, developed three different modules to concentrate on three different levels plus one to include radiotherapy effects (Figure 6). The discrete evolution of the model is showed in Figure 7 where cell cycle is described highlighting the damages, overpopulation and hypoxia signal check-points.

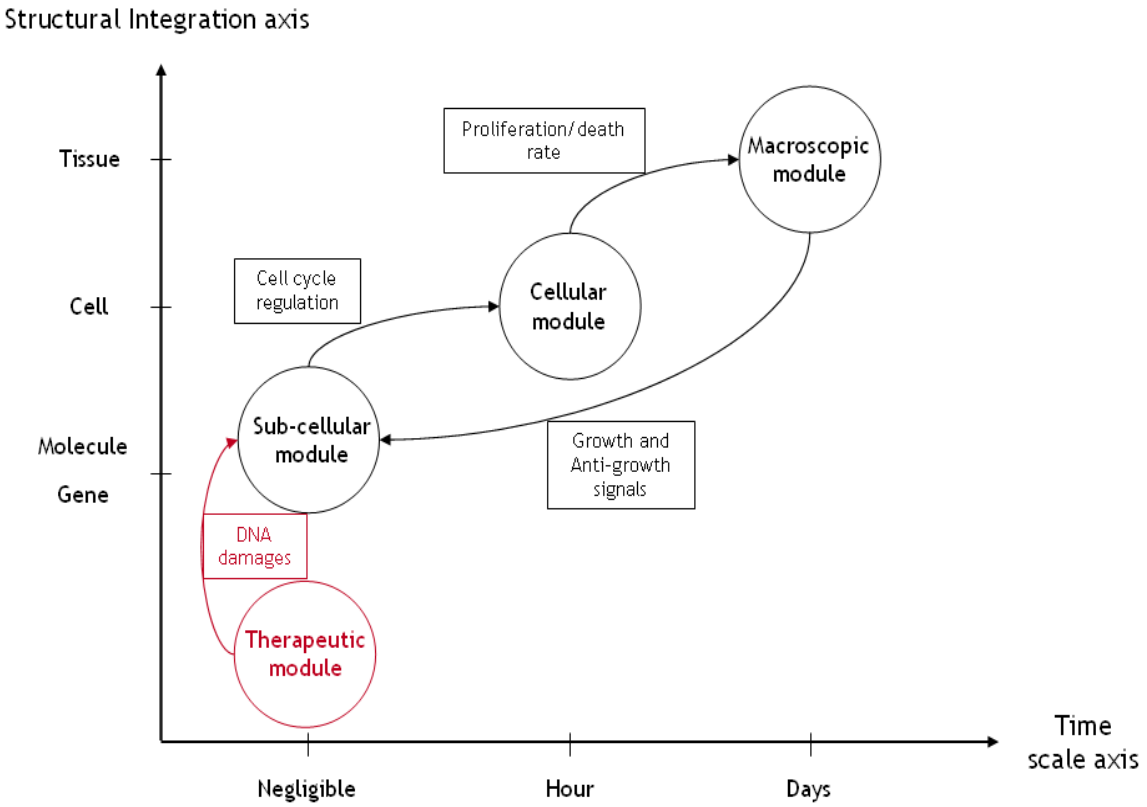

Figure 6
Overall model view **(Ribba B. et al., 2005)**

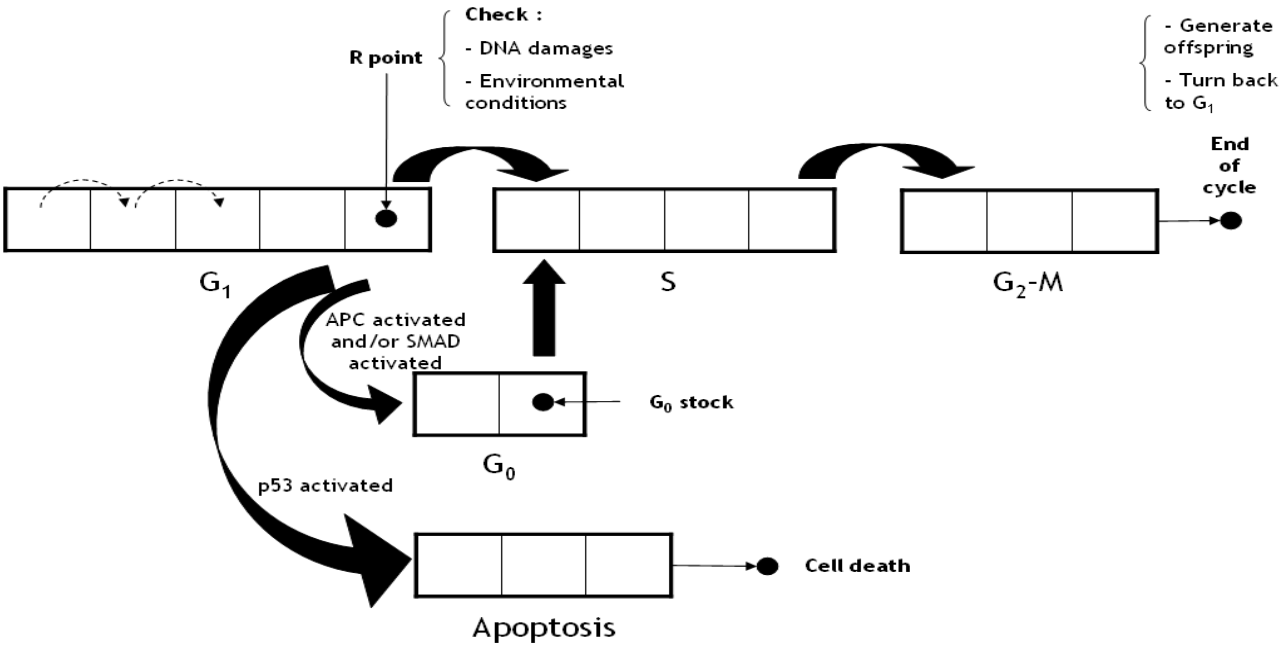

Figure 7
Cell cycle evolution **(Ribba B. et al, 2005)**

2.2 Continuous models and Discrete models

The model may evolve in a discrete environment or continuously in space (and time). Basically it depends on the specific needs of the project and the kind of analysis that has to be performed. In the following lines advantages and disadvantages of the two different approaches will be examined. The early attempts of tumor modeling have been mostly based on differential or partial differential equation already quoted in the previous paragraphs (Mayneord W. V. et al., 1932) (Hill A. et al, 1928). One of the most used model is the so called reaction-diffusion one. It is able to represent one or more evolving populations, and takes into account the simultaneous occurrence of two processes, the local chemical reactions that turn one into the other, and diffusion which causes the substances to spread out. The one-component formulation is ruled by the following equations (9, 10).

$$\frac{\partial u}{\partial t} = \nabla(D\nabla u) + R(u,t) \tag{9}$$

$$(\eta\nabla)u = 0 \tag{10}$$

Equation 9 describes two characteristics of the tumor growth: diffusion and proliferation. In fact, the term on the right hand side represent the invasion of tumor cells by means of a brownian motion, where **D** is the diffusion tensor, **u** the tumor cell density and the gradient operator implies a differentiation with respect to time. ***R(u, t)*** is referred instead to the proliferation of tumor cells. Equation 10 mimics the fact that tumor cells cannot diffuse towards some kind of structures, where η is the normal direction with respect to the boundaries (Konukoglu E. et al., 2008). The solution of the system usually relies on well-known curves whose formulation was widely studied such as the Gompertzian one or the Logistic equation.

Gatenby and Cruywagen have been pioneers in the application of competition theory to the propagation of tumor cells, anyway both of them customized the reaction diffusion model considering the contemporary evolution of two different populations (Gatenby R. A. et al., 1995), (Cruywagen G. C. et al., 1995). This new system catch the interactions between healthy and pathological cells. Afterwards many improvements have been introduced modifying the formulation, for example the characterization of the diffusion tensor that evolved from isotropic-homogeneous (Cruywagen G. C. et al., 1995). to isotropic-non homogeneous (spatially dependent) (Swanson K. A. et al, 2002), and finally anisotropic-non homogeneous (Clatz O.et al., 2005). The continuum modeling approach just illustrated, is able to capture the tumoral structure at the tissue level, but it cannot describe the process at the cellular level and beyond. This may be a critical limit because the introduction of the cellular and subcellular mechanisms helps to explain and anticipate the overall outcomes at tissue level: without the inclusion of some of the stochastic, discrete rules regulating the cell's development and mutation the model description might fail in effectively predicting the real evolution of the pathology.

The idea of basing the modeling of tumor growth ,which is nothing but a multi-cellular biological system (MCBS), on single cells interactions can be easily justified since it is well known that cancer arises from cell mutations. Currently, a standard cell-based model does not exist, anyway a rough

classification could be to distinguish the models in which the elements can move freely through continuous space from the ones structured in lattices. The latter have been more extensively used because of their intuitive way of working and implementation and are frequently called Cellular Automata (CA). Probabilistic Cellular Automaton models (CA) as well as Agent Based Model (ABM), are artificial ecologies approaches used to model population dynamics that utilize a bottom up approach and often exploit stochastic rules to evolve (Boondirek A. et al., 2010). Conway's Game of Life has been one of the first and most famous examples of C.A., but in order to show the variety of fields in which this approach can be used we may quote also the work realized by Frisch on computing the Navier–Stokes equations (Frisch U. et al., 1986). As already said cancer can be thought as a multicellular system but the problem with biological systems is that each element is alive and this generally implies a fairly complex system of rules that may depend not only on local environment but also on global factors (Hwang M. et al., 2009). Cellular automata are defined by Moreira J. et al. as a class of spatially and temporally discrete, dynamical systems based on local interactions. In their review the researchers present the basic structure and the main characteristics of the C.A. approach showed in Table 1

Discrete space	the evolution space is a two- or three- dimensional (regular) lattice
Discrete states	each site takes one of a finite number of states
Discrete dynamics	the evolution time steps are discrete and equal for each cell
Local rules	the transition rules depend only on the site's spatial neighborhood configuration and can be deterministic or probabilistic
Homogeneity	all sites are equivalent (since the lattice is regular) and the transition rules are the same for each of them

Table 1
The principles of cellular automaton **(Moreira J. et al., 2002)**

This boundaries can be extended, obtaining several variants:

- **Coupled-map lattices** (CA where cells do not only assume a finite number of discrete states)
- **Asynchronous CA** (there is no more restriction of simultaneous update of all the states).
- **Non-homogeneous CA** (transition rules may vary from site to site or in time)

(Moreira J. et al., 2002)

The lattice is a grid of spatially defined geometrical cells that evolves trough time in accord with neighbors' characteristics. Both cell's shape and neighborhood definition may vary (Figure 8), the parameters as well as the evolution rules obviously, may strongly influence the resulting prediction.

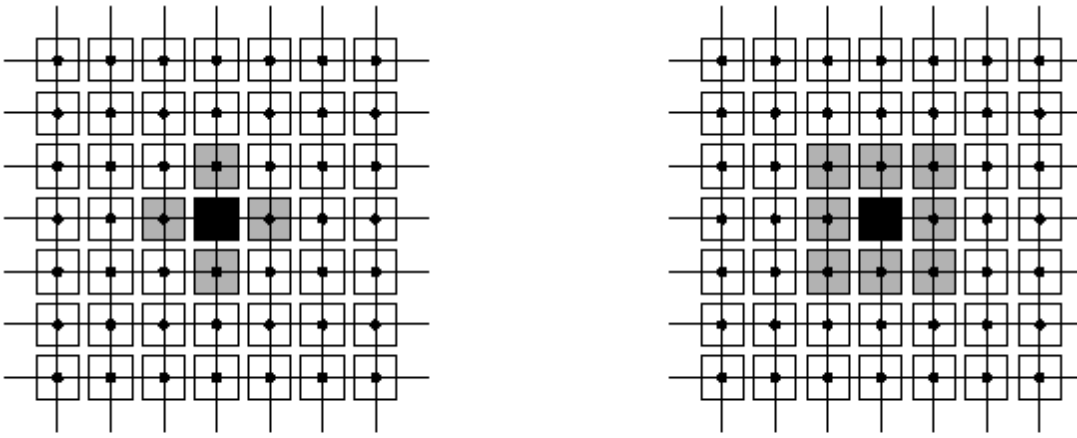

Figure 8
Neighborood definition **(Moreira J. et al., 2002)**

Another way to deal with discrete models may be to use an hypermatrix notation as described in the work of Stamatakos (Stamatakos G. et al., 2009).

In conclusion every type of model has some positive aspects and drawbacks. Continuum models, as previously remarked, take advantage of the properties of their formulation usually characterized by the usage of ordinary or partial differential equations. This models are able to catch the overall relationship among the various elements but they cannot represent the discrete nature of individual cells, and it may fail in handling many variables. On the contrary, the discrete models such as cellular automata (CA) or agent-based models (ABM) can easily and effectively describe individual cell behavior and interactions among them but may not fit the overall dynamic properly and require a consistent computational effort. Therefore, continuum models and discrete models may be combined to realize a more effective tool. The idea is to model some components such as chemical distribution in MCBS with the continuum approach while the overall system keeps evolving under discrete rules resulting in a more accurate prediction (Hwang M. et al., 2009).

2.3 Therapies modeling

In order to achieve a realistic prediction and realize an effective tool able to assist the oncologist in choosing the most suited therapy for his patient it is necessary to include the therapies outcomes in the model evolution. In this section we will deal with the most common treatments which are radiotherapy and chemotherapy. The surgery is not considered because it will not be included in our model in this work, anyway, in a discrete lattice tissue removal can be easily modeled clearing the content of the corresponding geometrical cells.

2.3.1 Radiotherapy

The purpose of radiotherapy is to maximize local tumor control while minimizing the probability of damaging the healthy tissue. In order to achieve this goal many different radio treatments were introduced, the beam thickness was reduced, the energy band of the radiation was limited, and some dynamic tumor tracking techniques were developed in order to irradiate the very critical region only, saving the adjacent tissues. The major hypothesis at the base of this technique is that the sensitivity to

radiation is greater in mutated cells, this idea is supported by the fact that the tumoral cells generally get more oxygen, because they stimulate angiogenesis, than healthy cells and it is the interaction with this element that causes most of the damages. Modeling the cell's response to radiation is usually based on the assumption that the critical damages are the one delivered to DNA, which can be subdivided into three main categories:

- **Bases damages** that are assessed to be 2000 to 4000 per Gray (Gy, dose measure unit) and cell, the most common.
- **Single strand breaks** of the sugar desoxyribose (1000 per cell per Gy)
- **Double strand break** of the sugar deoxyribose is the fatal damage, not repairable, but also the less frequent (40 per cell ang Gy).

The cell survival after irradiation can be modeled by the Linear Quadratic Model (LQM) as showed in Equation 11 where **S(D)** is the survival percentage relative to the administered dose **D**, alpha and beta are the coefficients related to the radio-sensitivity of the specific cell type.

$$S(D) = \exp[-(\alpha D + \beta D^2)] \qquad (11)$$

This model is named LQ because the relation between the survival fraction and the dose is both linear and quadratic in a logarithmic scale. The linear component correspond to lethal events that are caused by a single particle (double strand breaks) while the quadratic component represent the occurrence of two sub-lethal events (single strand break) in a narrow period of space and time, this does not allow the cell to repair the first damage and causes its death (Figure 9).

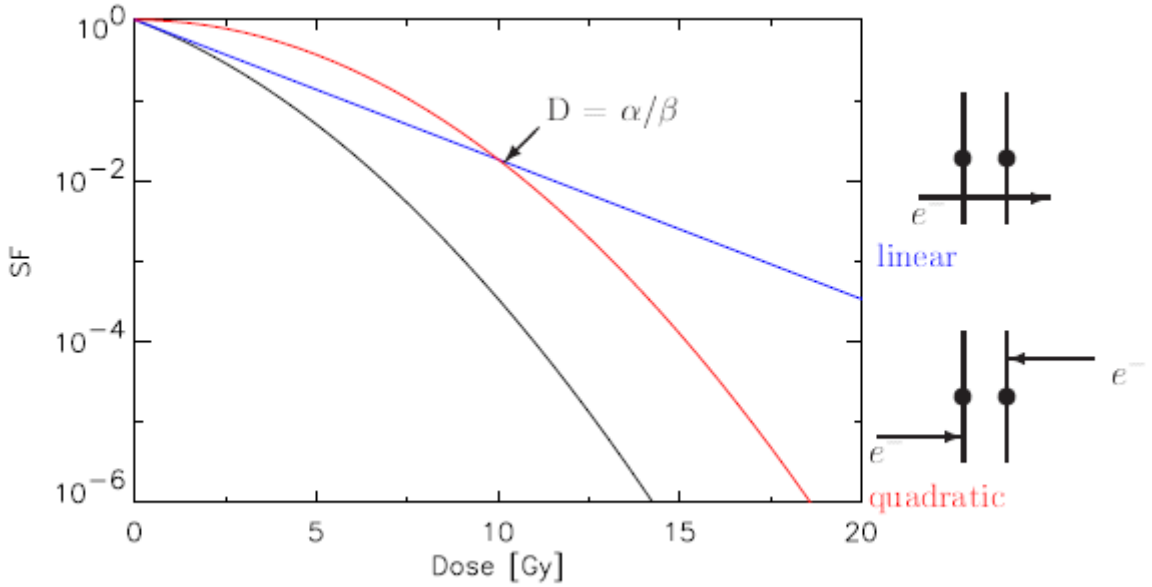

Figure 9
Survival fraction **(Borkestein K. et al., 2001)**

2.3.2 Chemotherapy

When metastasis occur, generally, a systemic treatment is required as, for example, chemotherapy, which is delivered through bloodstream to all the cancerous sites. The aim of chemotherapy is to suppress the mutated cells using toxic drugs that are effective both on cancerous and healthy cells so the main problem is to keep the toxicity acceptable on the normal tissue while maximizing the mutated cells death. Mathematical models for cancer chemotherapy have been extensively researched in the last decades. It is quite difficult to represent all the aspect of the underlying biological processes also because our understanding of the dynamics is still incomplete. The dosage, the permeability of the tissues and the specificity of the chemicals have to be taken into account, but there are many other meaningful parameters, especially if we are dealing with multi-drug treatments. Another critical issue in the design of the mathematical framework is the assessment of the negative side effects of the treatment. In clinical trials the new drugs are delivered progressively increasing the dosage to find out the level at which side effects show up, but the actual biological mechanisms is not completely understood yet, therefore the drawbacks of the treatment are often included only implicitly in mathematical models (e.g. minimizing the drug dosage in the objective). A recent type of drug model is called 'cell-cycle-specific' because the mutated cell are only poisoned during a specific stage, usually the mitotic phase. The reason for this behavior is quite intuitive, the cell wall is thinner while it duplicates and the diffusion coefficient is maximal. Because of this characteristic of the drug we can divide the cells in the mitotic phase (or the one sensitive to the drug effects) from the others. In this way it is possible to formalize the transition from one group to the other introducing some coefficients : in case of chemotherapy the number of cell exiting from the first group will be more than the one entering the other because some will be poisoned by the drug (Ledzewicz U. et al, 2004). The chemotherapy model usually consist of two parts, **pharmacokinetics** (drug penetration) and **pharmacodynamics** (drug cytotoxicity) as showed in Figure 10.

$$c' = -fc + h \tag{12}$$

The **pharmacokinetics** concentrates on the obstacles and barriers that the drug may encounter on its route to the target, this barriers reduce the overall effect of the treatment, this phenomena is known as multidrug resistance (MDR). A possible approach (Murray J. at al., 1994) to pharmacokinetic is to use a first-order linear system as the one described by Equation 12 with the condition $c(0) = 0$ where c represent the drug concentration in the plasma and f and h are positive constants.

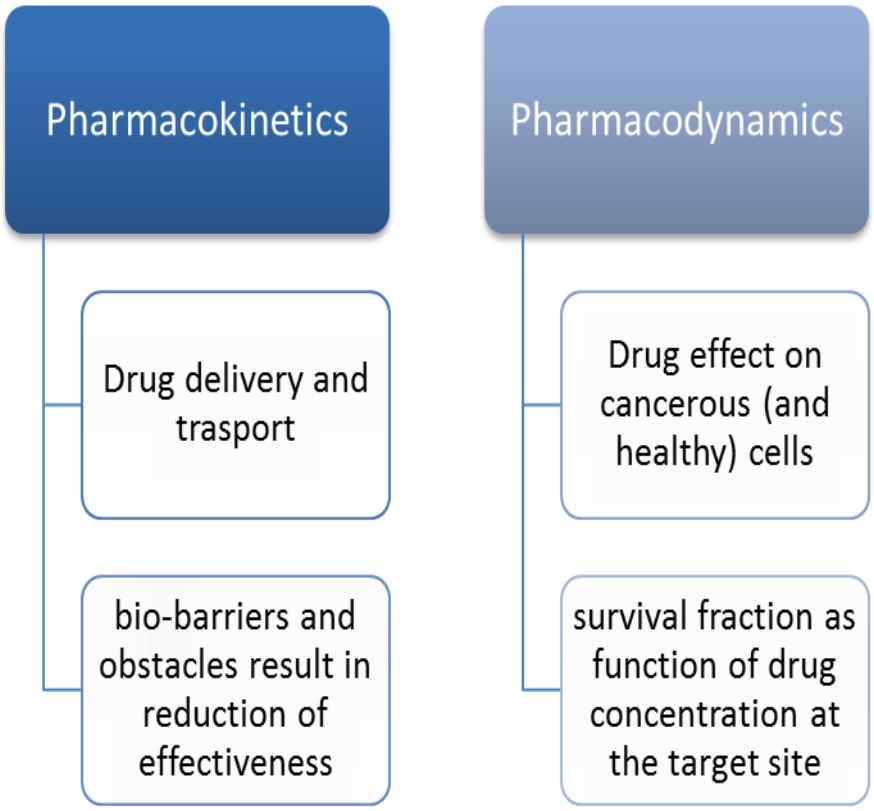

Figure 10
Differences between Pharmacokinetics and Pharmacodynamics The sum of this two behaviors explains the overall effects of the chemotherapeutic treatment.Anyway the knowledge about the mechanisms that regulate them is still partial.

It is reasonable to think of the organism as a net of pools (compartments) which can be defined by their properties (e.g. permeability) and the presence of fluxes. Conventional pharmacokinetics uses this approach to model drug interactions at cellular level. In their work (Figure 11), Marger et al, developed a system consisting of three compartments to analyze the effects of the biological barriers and obstacles on the efficacy of the drug and found out that the decreased intracellular sequestration was the main cause of the worsened effectiveness of the treatment (Dordal MS et al., 1995).

On the other hand **pharmacodynamics** models the effect of the drug on cancerous and, eventually, on healthy cells using often a statistical approach to the problem . In some works it is assumed that the effectiveness e of the drug is proportional to the number of ineffective cell division and the process is simulated by the impairment between the number of cells exiting one group and entering the other as previously described (Ledzewicz U. et al.,2004).

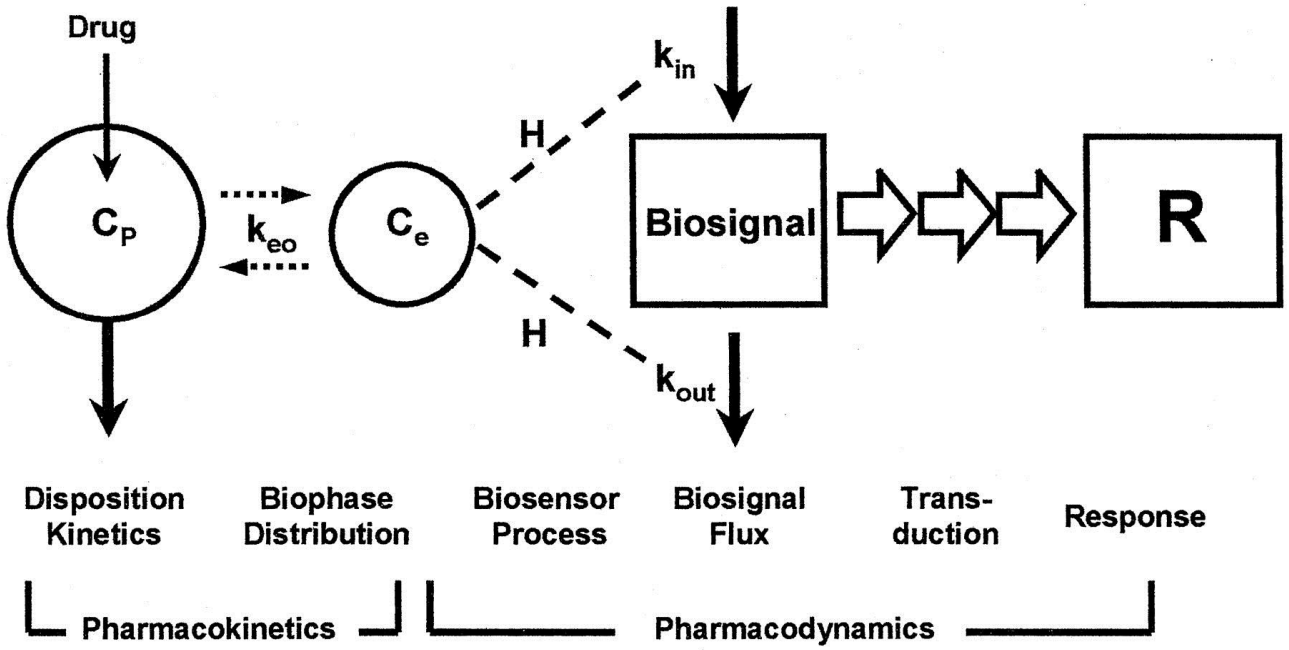

Figure 11
General scheme of drug delivery and effects **(MagermmD. E. et al., 2003)**

The mechanisms that regulates drug efficacy are not completely understood, the early models were based on linear regression, it was recognized that the intensity of many pharmacological effects is linearly related to the logarithm of dose. This behavior has been formalized in Equation 13 (Levy G. et al., 1966) where **E** are the effects, **A** is the dose, and **m** and **e** are the slope and intercept terms respectively. After that the Hill model was proposed in order to describe the in vivo concentration-response relationship (Wagner J.G. et al., 1968). Further information and a complete review about this models can be found in the works of Mager. (Mager D. E. et al., 2003) who proposes a general scheme presented in Figure 11, and many others (Ledzewicz U. et al.,2004) (Fister R. K. et al., 2000).

$$E = m * \log A + e \tag{13}$$

In some cases the chemotherapy is used together with the irradiation and its effect can be modeled as an increased effectiveness of the radiotherapy instead to thinking of it as a completely different process. In their work 'A mathematical model of brain tumor response to radiotherapy and chemotherapy considering radiobiological aspects' Barazzuol and colleagues compared the results of the two different approaches (Barazzuol L. et al., 2009). They tested two different models, one considering chemotherapy and radiotherapy separately the other considering only the enhancement of radio sensitivity due to drugs, the best prevision in terms of life expectancy was achieved with the second one (Figure 12).

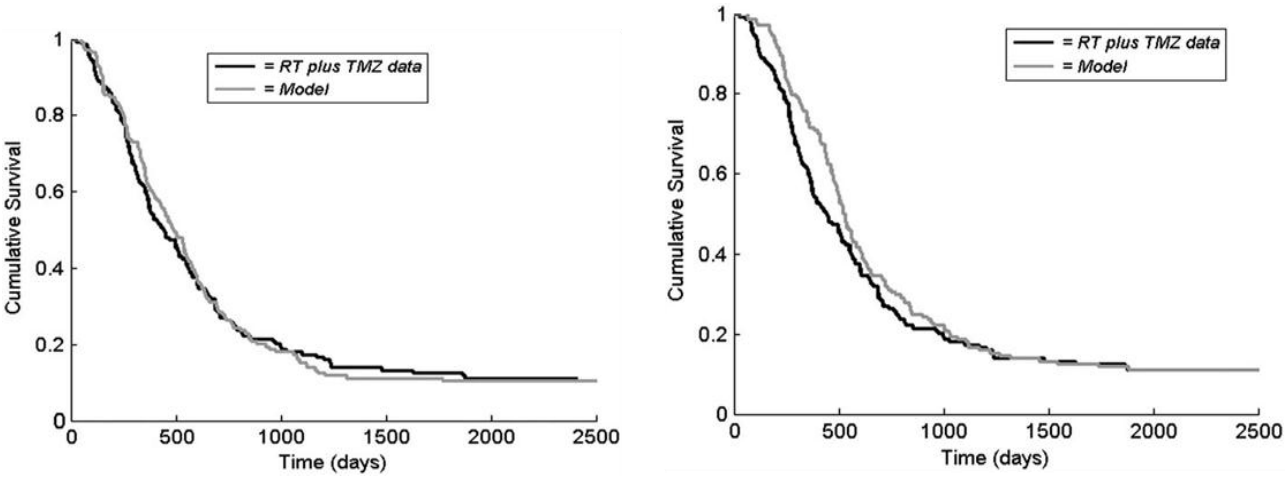

Figure 12

Model vs trials data (Barazzuol L. et al., 2009).On the left: Radiosensitizing effect of chemotherapy. On the right: Indipendence of radiotherapy and chemotherapy effects

3 Materials and methods

One of the main issues was to decide which kind of approach to choose. As already described in the Background chapter, there are several options to model the dynamic of a tumor, each of them has some advantages and drawbacks, therefore we decided to present two different possible solutions and to analyze which one is more likely to meet the needs of this specific dataset. The bottom-up approach peculiar of the Cellular Automata was selected among the discrete models because of its flexibility. Its modularity allows this model to be as complex as the problem requires and to predict unexpected behaviors starting from simple rules of interactions among the smallest units: the cells. The classic technique that involves the use of the Ordinary Differential Equation (ODE models) requires the computation of a numerical solution of the equations. Despite the mathematical complexity it is actually a very simple tool, able to give some hints about the overall dynamic without considering the spatial evolution of the tumor. In this chapter the two models will be presented along with the data at our disposal and the elaboration needed in order to make them compatible with our purposes.

3.1 Patients, image acquisition and treatment protocol

The women selected for this study did not underwent any surgical operation. The patients were initially subdivided into three groups but only the data belonging to two of them (9 subjects) were already available to be analyzed.

PATIENT	IMAGES	TUMOR	THERAPY	GROUP
Patient 2	7 images	SC	RT	1
Patient 3	8 images	SC	RT	1
Patient 4	9 images	SC	RT	1
Patient 5	10 images	SC	RT	1
Patient 6	9 images	SC	RT	1
Patient 8	9 images	AC	RT+CT	2
Patient 9	7 images	AC	RT+CT	2
Patient 10	8 images	AC	RT+CT	2
Patient 11	10 images	AC	RT+CT	2

Table 2

Data classification according to the kind of tumor and therapy.(SC= squamous cell carcinoma; AC= adenocarcinoma; RT= radiotherapy; CT= chemotherapy.) Patients 1 had to be discarded because of its reaction to drugs that forced the physician to interrupt early the chemotherapy, while patients 7 has not been considered because of data inconsistency

		ACQUIRED IMAGES	CONTOURED
Patient2	CT	31 January 2011	31 January 2011
	CBCT	17 18 22 23 24 25 28 February 1 2 3 4 7 8 9 10 11 14 15 16 17 18 21 22 23 24 March	17 18 22 February 4 11 18 March
Patient3	CT	18 February 2011	18 February 2011
	CBCT	1 2 3 4 7 8 9 10 11 14 15 16 17 18 21 22 23 24 25 28 29 30 31 March 1 4 April	1 2 3 8 15 22 29 March
Patient4	CT	11 March 2011	11 March 2011
	CBCT	28 29 30 31 March 1 4 5 6 7 8 11 12 13 14 15 18 19 20 21 26 27 28 29 April 2 3 4 5 6 May	28 29 30 March 4 12 18 27 April 6 May
Patient5	CT	01 July 2011	01 July 2011
	CBCT	11 12 13 14 15 18 19 20 21 22 25 26 28 29 July 2 3 4 5 8 9 10 11 12 16 17 18 19 22 August	11 12 13 18 25 July 2 9 16 22 August
Patient6	CT	13 September 2011	13 September 2011
	CBCT	27 29 30 September 3 4 5 6 7 10 11 17 18 19 20 21 24 25 26 27 31 October 2 3 4 7 8 November	27 29 30 September 6 10 17 24 October 3 November
Patient8	CT	05 May 210	05 May 210
	CBCT	11 12 13 14 17 18 19 20 21 26 27 28 31 May 1 3 4 7 8 9 10 11 14 15 16 17 June	11 12 13 20 27 May 3 10 17 June
Patient9	CT	13 June 2011	13 June 2011
	CBCT	23 24 27 28 29 30 June 1 4 5 6 7 8 11 12 13 14 15 18 19 20 21 22 25 26 28 July	23 24 27 June 7 15 21 July
Patient10	CT	29 June 2011	29 June 2011
	CBCT	12 13 14 15 18 19 20 21 22 25 26 28 29 July 2 3 4 5 8 9 10 11 12 16 17 18 August	12 13 14 21 28 July 4 9 August
Patient11	CT	16 November 2011	16 November 2011
	CBCT	30 November 1 2 5 6 9 10 12 13 14 15 27 28 29 30 December 2 3 4 5 9 10 11 12 13 16 19 20 23 January 2012	30 November 1 2 12 27 December 3 10 16 23 January 2012

Table 3
Image acquisition dates and actual (contoured) data

The subdivision is due to the location of the tumor inside the cervix and the kind of therapy delivered:

1. Patients affected by squamous cell carcinoma (SC) threated by means of radiation therapy (RT) only
2. Patients affected by adenocarcinoma (AC) administered with both radiation therapy (RT) and chemotherapy (CT) .

The complete scheme of the available data is summarized in Table 2. Every patient underwent a CT (computed tomography) scan, usually a couple of week before the first treatment session. A radiation therapy plan was developed on this first acquisition and several CBCT (cone beam CT) have been taken, one at any scheduled session of irradiation, in order to check the tumor evolution during the treatment.All the acquired images are dicom files and include the volume itself plus an header in which the information about the patient, the date of treatment and other meaningful notes are stored. The spatial information are structured in 3D matrices containing the grey level of the pixel in the corresponding position. Moreover there are some fields containing the coordinates of the contours of some important body structures, especially the Gross Tumor Volume. The schedule of the radio-therapeutic treatment correspond to the dates of the CBCT acquired images (Table 3), while the planned dose is specified in a 3D matrix where the value in each pixel correspond to the amount of gray (Gy) delivered to the corresponding point in the CT volume. In Table 4 the nominal dose delivered to the patient during each session, the number of fraction in which the therapy has been subdivided and the overall irradiated dose can be found.

	Dose per fraction	Fractions	Total dose
Patient2	2 Gy/fr	25	50 Gy
Patient3	1.8 Gy/fr	25	45 Gy
Patient4	1.8 Gy/fr	28	50.4 Gy
Patient5	1.8 Gy/fr	28	50.4 Gy
Patient6	1.8 Gy/fr	25	45 Gy
Patient8	2 Gy/fr	25	50 Gy
Patient9	2 Gy/fr	25	50 Gy
Patient10	2 Gy/fr	25	50 Gy
Patient11	1.8 Gy/fr	28	50.4 Gy

Table 4
Radiotherapeutic data

The second group of patients underwent both radio and chemotherapy but among them there were sensible differences in the kind of drugs (Cisplation or Paclitaxel), the dosage, and the time schedule (usually once a week, 5 or 6 cycles) more information about the treatment are available in Table 5. However we suppose chemotherapy to be only an adjuvant factor, that makes the tumor more sensitive to the irradiation. Since the drugs are delivered by the bloodstream, this therapy can be thought to affect more the surface of the tumor than the inner part because of its resistance to the perfusion

	Chemo	Num. of cycles	Drugs
Patient8	yes	6	CDDP (60 mg) and PTX (45 mg)
Patient9	yes	5	CDDP (60 mg)
Patient10	yes	5	CDDP (70 mg) and PTX (50 mg)
Patient11	yes	6	CDDP (65 mg)

Table 5
Chemotherapy data (CDDP= Cisplatino, PTX= Paclitaxel)

3.2 Pre-processing of the data

3.2.1 Image analysis

The elaboration of the data in order to meet the model requirements can be summarized in a few steps:

a) The GTV (Gross Tumor Volume) was countered by the physicians
b) A binary volume was derived from the CT and CBCT images, each voxel was either healthy or not depending on whether its position was inside or outside the GTV
c) The 3D surface of the mass was reconstructed (e.g. Figure 17).
d) The global Volume and Surface values were computed at each time step

Obviously all the data had been previously anonymized. The contouring of the images demanded a good level of expertise, therefore this step was accomplished by a physician, anyway this was a critical aspect in the data uncertainty estimation. The contours were nothing but some points chosen on the perimeter of the ROI (Region Of Interest). In the first image of each patient, the only one collected by a CT scan, several body structures were highlighted (Figure 14Figure 14), here we quote some of them:

- Ovaries
- Uterus
- Bladder
- Intestine
- kidney
- Femur
- Gross Tumor Volume (GTV)
- Clinical Target Volume (CTV)
- Planning Target Volume (PTV)

The last three are all referred to the tumoral mass but while the GTV (red in Figure 13) is the part of the lesion that is actually visible in the image, the CTV is a larger structure which includes also the small and unnoticeable extent of the pathology, and the PTV takes into account the uncertainty due to the patient movements and tissue deformation, PTV is the overall irradiated area.

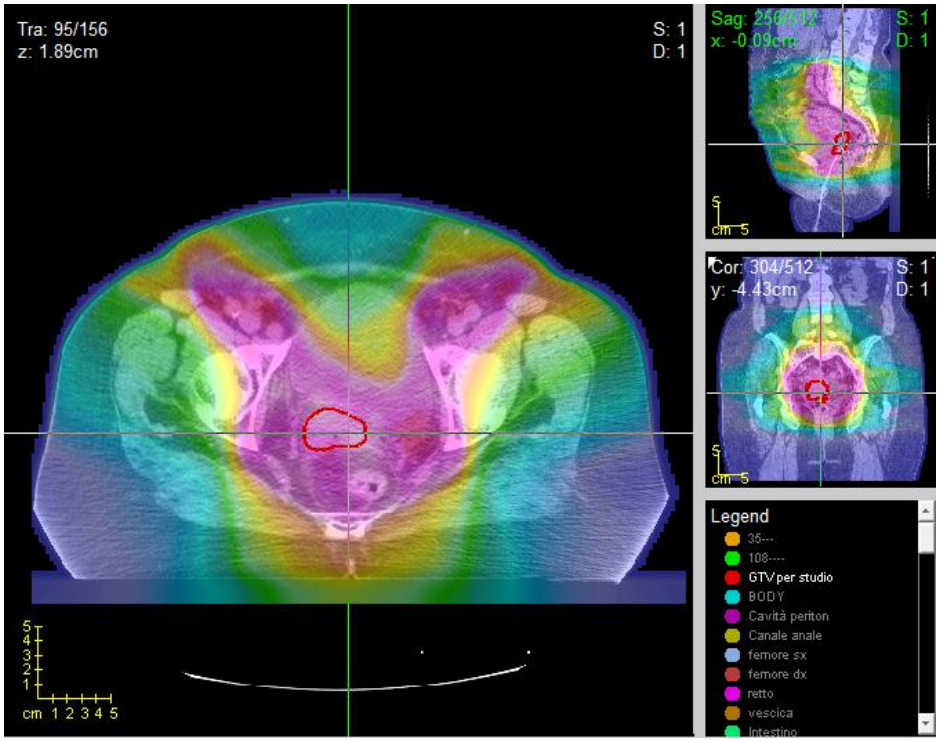

Figure 13

Example of image contouring: the red circle highlights the Gross Tumor Volume on a specific slice. The colored areas show the way the dose has been delivered to the patient, the pink one is the zone that underwent the maximal irradiation

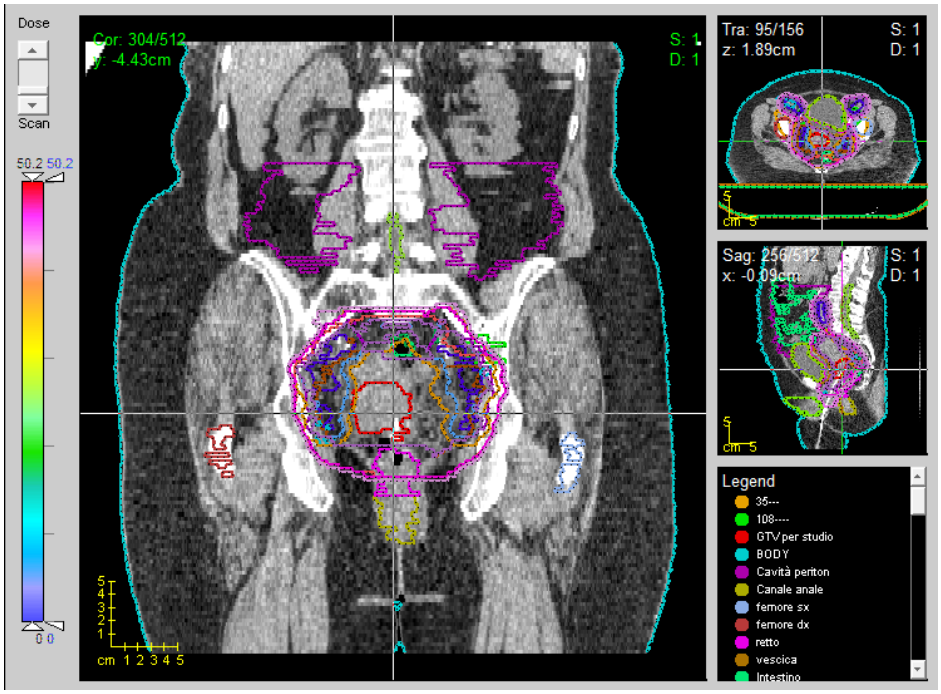

Figure 14

Example of CT image where several structures have been contoured

3.2.2 Binarization & global indexes calculation

The binarization of the images consist in discriminating the pixels that lay inside the contours of the gross tumor volume, setting them to 1 (white), from the background that assumes the value 0 (black). This operation will be helpful to proceed with the volume calculation, and it could be used in the Cellular Automata settings. The obtained black and white image is then saved as an mha file. Since we know the image spacing, that corresponds to the dimension of the voxels of the 3D matrix, it is easy to calculate the global volume value (V), which is needed to fit and validate the ODE models, as the product of the volume of a single voxel (ΔV) and the number of cancerous units (Equation 14).

$$V = \sum \Delta x * \Delta y * \Delta z = N * \Delta x * \Delta y * \Delta z = N * \Delta V \quad (14)$$

Where Δx, Δy and Δz are the image resolutions in each direction. Since all the cells have the same shape and fixed dimensions, we can easily calculate the volume of one of them (ΔV) and multiply it for the overall number of unhealthy cells. In order to understand if the mass grows in the same way in every direction or elongates and changes its shape we decided to calculate also the area of the tumor and compare it to the volume. From the points of the GTV contouring a Delaunay triangulation is obtained, and the overall area can be calculated summing up all the areas of the triangles. Moreover it is possible to calculate a peculiar ratio between volume and area, which does not depend on the radius, in order to understand if there is some elongation (Equation 15)

$$\text{ratio} = \frac{V^{\frac{1}{3}}}{A^{\frac{1}{2}}} \quad (15)$$

If the mass had a spherical shape the volume could have been calculated as $V = \frac{4}{3} * \pi * r^3$ while the area as $A = 4 * \pi * r^2$ where r is the radius, therefore the ratio would have been a constant value through time (0.4547) a different but quite constant value would suggest that the tumor is elongated but keeps its shape during the treatment. Some example for the second group of patients is shown in the following chart (Figure 15).

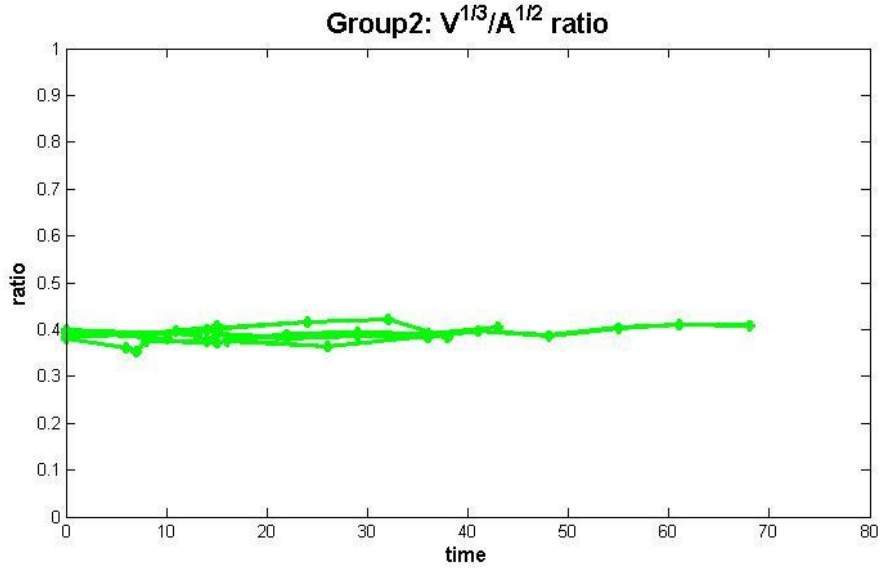

Figure 15

In this graph the $\frac{V^{\frac{1}{3}}}{A^{\frac{1}{2}}}$ ratio for each Adenocaricinoma patient is represented. If the tumor had been perfectly spherical during all the therapy it would have been constant and equal to ~ 0.4547. Time is expressed in days.

3.2.3 Data uncertainty

Many errors could possibly occur during the data acquisition and elaboration, depending on the well-known CBCT resolution and the contouring accuracy which is mainly due to the expertise of the physician which is not so easy to be esteemed. Using a spherical approximation the equivalent radius has been calculated from the inverse formulas of the area (Equation 16) and volume (Equation 17).

$$r_A = \sqrt{\frac{A}{4*\pi}} \tag{16}$$

$$r_V = \sqrt[3]{V * \frac{3}{4*\pi}} \tag{17}$$

Considering the image resolution, all the indices have been recalculated introducing an error of ±1 mm on the radius (Figure 16). In this way an approximation of the minimal and maximal values for both volume and area were found. The difference between them can be considered as a rough estimation of the data uncertainty

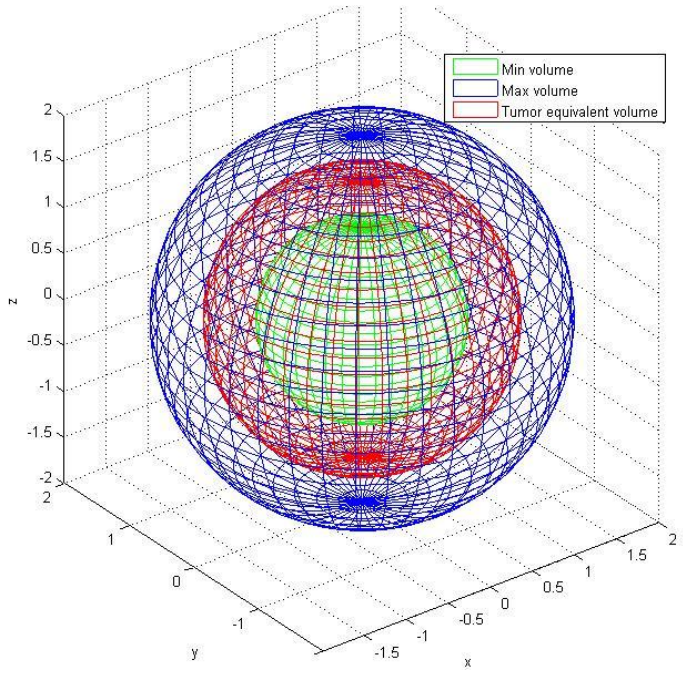

Figure 16

Data uncertainty estimation using the spherical approximation. in red, sphere with the equivalent volume as the acquired tumor; in blue, radius increased by 1 mm; in green, radius decreased by1 mm. The values used in the pictures are not related to the real data at our disposal.

.

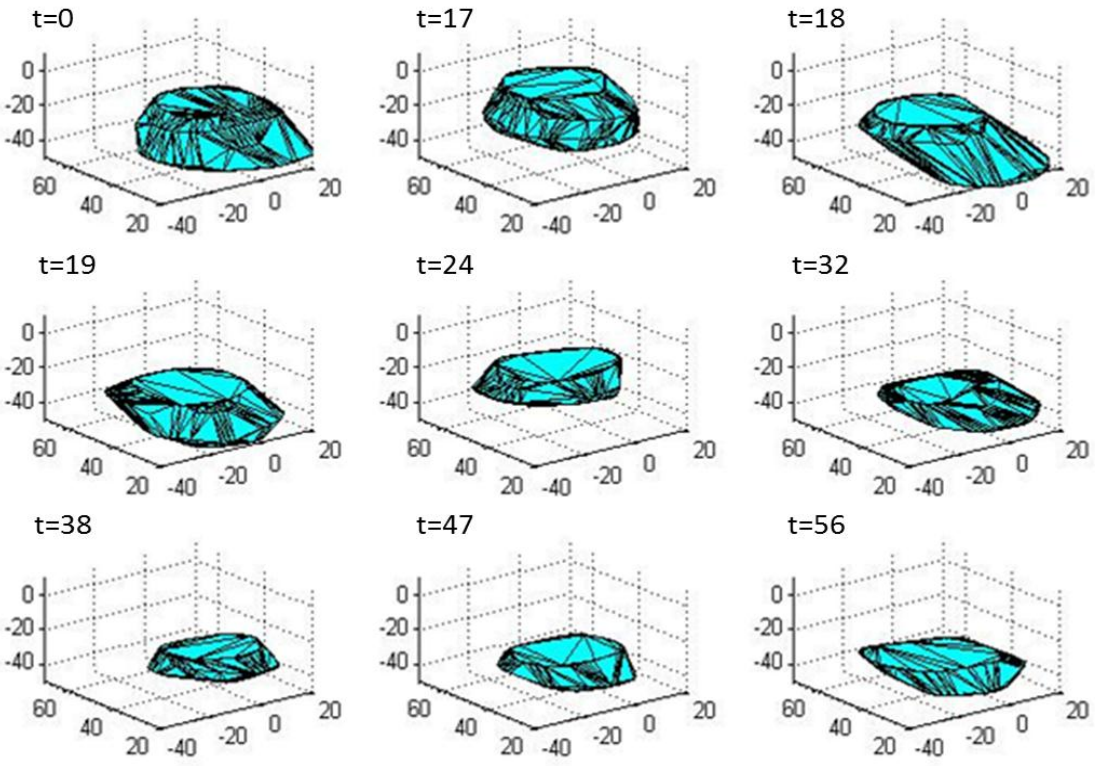

Figure 17

Evolution through time of the GTV of a patient. Time (t) is expressed in days while the dimensions on the axis in mm

3.3 Cellular automata

The model we realized consist of two main modules: the single cell evolution and the tissue evolution, the effect of the radiotherapy is simulated using the linear quadratic model while the chemotherapy is considered only as an enhancer of the effect of the irradiation as suggested by the outcomes of the work of Barazzuol (Barazzuol L. et al., 2009) The code has been implemented using the MatLab® environment.

3.3.1 Data structure

The basic structure is a 3D matrix where each geometrical cell represents a voxel and contains a certain number of physiological cells (depending on the site density), considering that we are dealing with a pixel resolution of a few mm^3 of the CT/CBCT images. The main fields of the structure are showed in the following picture(Figure 18) and their purpose will be discussed further in the next paragraphs. Each geometrical cell is labeled as 'C'(cancerous), 'H'(healthy) or 'N'(necrotic), this information goes under the field '**type**' and determines the characteristics of all the elements of that unit. During the preprocessing of the data, the initial type will be assigned to the cell accordingly to the segmentation operated by the expert who contoured the neoplasm.

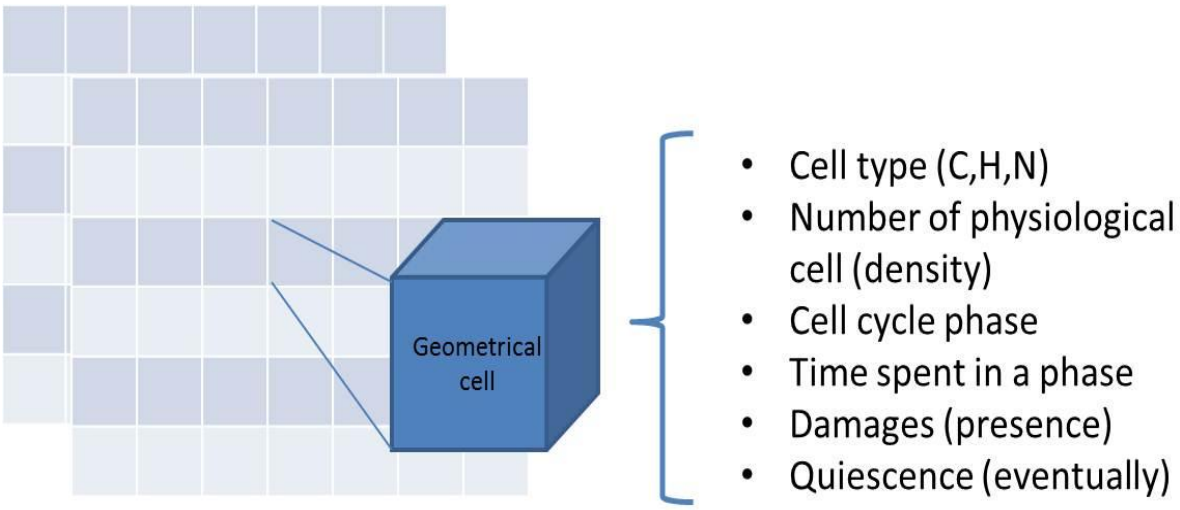

Figure 18
Data structure and stored information

As previously said, each geometrical cell will contain a certain number of biological cells depending on its dimensions and the cell-density of the site where it is located. This value goes under the name of '**numel**' (number of elements) in our structure and it could be set using the data available in literature or could be obtained through an histological analysis, it could be even possible to tune the parameter in accord to the grey level present in the corresponding site of the CBCT image. Using a statistical simplification, we considered every element belonging to the same voxel as if they were all in the same phase of the cell cycle (label '**phase**'), the phase is assigned randomly but with a probability proportional to its time duration. We established the same time unit (1 hour) for both healthy and unhealthy cells in order to let them evolve in parallel, but we have set an higher rate of duplication and

a shorter duration of some phases of the cell cycle in the unhealthy cells to model the uncontrolled cancer growth Another field related to this topic is the one named '**timeinphase**' that stores the information about the time spent in the current phase, in this way the moment of transition to the following phase of the cell-cycle can be deterministically fixed accordingly to the average phase duration values present in literature . Because of the radiotherapy, the tissues might result damaged. This fact is recorded in the field '**damage**' which is set to 1 when some particular condition occur, the mechanism will be explained in detail in the next paragraph. Another possibility, accordingly to the biology of the problem, is that the nutrients or the oxygen concentration became too low to support the uncontrolled growth of the neoplasm. In order to simulate this occurrence we suppose that when the cell-density overcomes an upper limit and the unit is not able to invade the surrounding cells, it goes quiescent for a certain time period. After loading the initial structure customized for the specific patient the program starts asking to select the file (.txt) containing the **dose profile.** The profile of the given dose is described by a 3D matrix, with non-zero values in correspondence to the coordinates of irradiation and equal to the dose (Gy) delivered in that specific site. The knowledge of the scheduled time at which radiotherapy is delivered is necessary to mimic the effect of the treatment, a *.txt file has been saved ('**timefractions**.txt') with a row of increasing numbers representing the time expressed in hours, e.g. if the radiotherapy's session are scheduled every 24 hours in 6 days the file will contain the values [24 48 72 96 120].

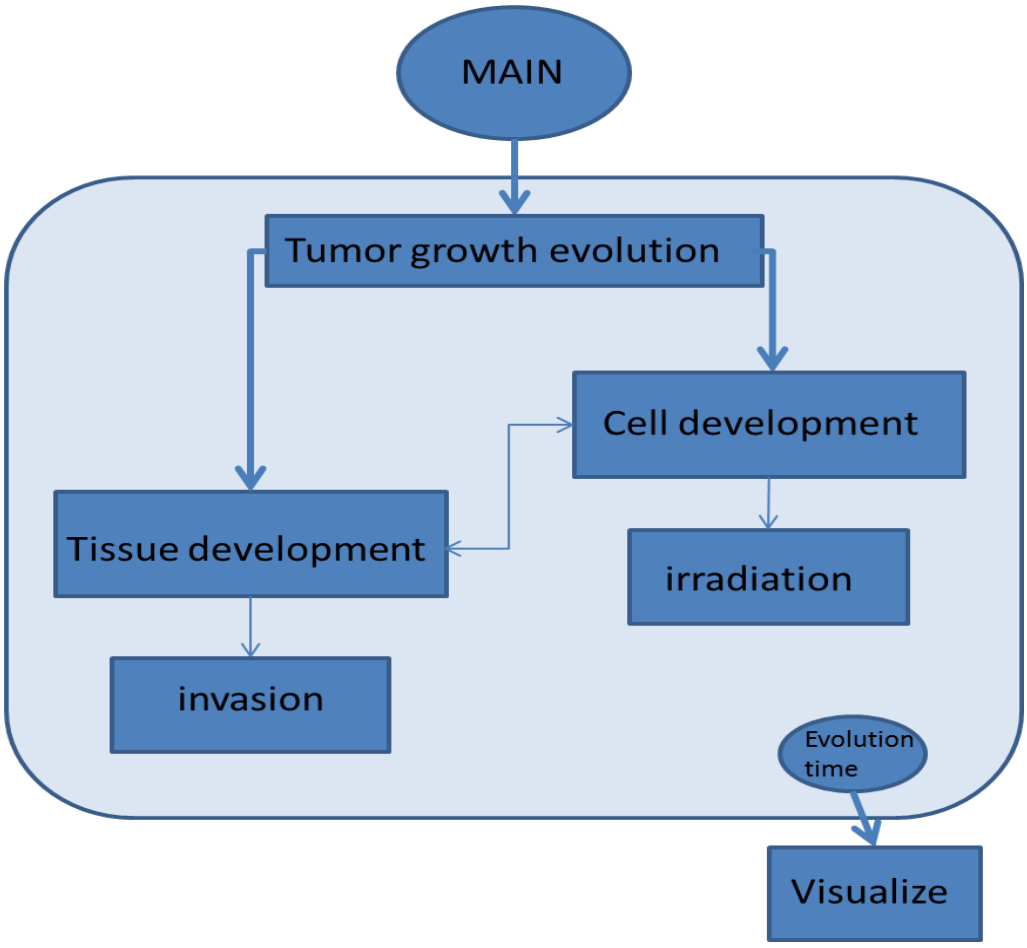

Figure 19
simplified general scheme of the cellular automata algorithm

3.3.2 Evolution

This model simulates the development of the tissue and the effects of radiotherapy, as showed in the simplified scheme (Figure 19). In the main code all the files and data are loaded and it is asked to the user to define the **time period** he wants to simulate ('evolutiontime'), the first module activated is the so called '**tumor growth evolution**' which is nothing but a for-type loop that ends only when the number of iteration required (evolution time) are reached. -The presence of radiotherapy is checked: a flag called *radio* is set to 1 only if the evolution-time is equal to the time at which radiotherapy is delivered and the corresponding *dose profile* is saved. In the loop the tumor is analyzed considering two different scales, the *cellular level* is simulated by the function '**celldevelopment**.m' and the *tissue level* is implemented by means of '**tissuevevelopment**.m'. The *cellular level* (Figure 21) evolves with a time step of one hour, each geometrical cell is checked, if it is labeled as quiescent or necrotic it is skipped, in the first case a counter that takes into account the time spent in quiescence is incremented . Both the healthy and the mutated cells that have been damaged during a radiotherapy session, keep evolving accordingly to their cell-cycle, the only difference is the value of some coefficients (LQ model): the same dose will cause the death of a greater number of cells. At each time step (hour) the time-counter of the damaged cells is increased, when it reaches a certain value, they recover, unless another session of radiation therapy is delivered in the meantime. When the cell exit the mitotic phase the '*numel*' is increased accordingly to the Gompertzian equation (Equation 18) (Gompertz, 1825):

$$X(t) = K \exp\left(log\left(\frac{(X(0)}{K}\right) exp(-\alpha t)\right) \quad (18)$$

If the considered cell is *healthy* or *cancerous* and not *quiescent* at each iteration the *time-in-phase* is incremented or the element enters a new phase of the cell-cycle. In case of radiotherapy (*radio*=1) the function '**irradiation**.m' is called, it takes as an input the data structure of the 3D geometrical cells (we will refer to this structure as *dstruct* from now on) and the dose profile and calculates the new values of the field *numel* in accord to the equation of the linear quadratic model we already quoted; the surviving fraction of cells is calculated as showed in Equation 19 where D is the dose, S the percentage of survived cells, alpha and beta two coefficients related to the specific radiosensitivity. If most of the cells are damaged (S(D)<0.5) the whole cell is considered as damaged (damage=1).

$$\mathrm{S(D)} = \mathrm{e}^{-\mathrm{alpha}*\mathrm{D}-\mathrm{beta}*\mathrm{D2}} \quad (19)$$

The function '**tissuedevelopment**.m' modifies only the cells that are entering the phase 'g2'. It checks if the geometrical cell type is 'H' otherwise the cell is skipped because:

- the 'N' type has a negligible cellular density, the whole voxel is considered as necrotic/fibrotic tissue, so it is not evolving anymore (or at least not in the timescale we are dealing with, the healing process will occur, hopefully, only much later);
- the 'C' type does not stand the influence of such a regulation.

First thing, '**tissuedevelopment**.m' calls another function: '**invasion**.m'. This function checks the 'C'-type cells, if their *numel* is greater than a threshold (*thnumelC*) their neighborhood may be at risk of invasion. In '**invasion**.m' a mechanism of quasi-random selection of a neighboring cell is implemented as showed in Figure 20. If it fails the invader became quiescent (quiescence=1) for some hours (*thtimequiescence*).

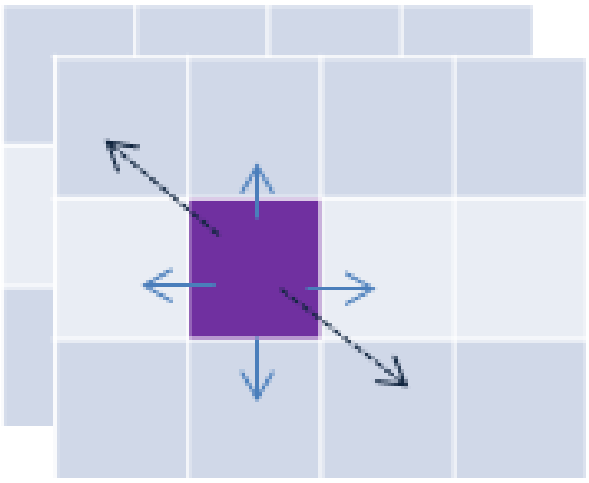

Figure 20

The cell(I,j,k) highlighted in violet is potentially invasive(type 'C' & numel>thnumelC), it picks randomly a cell increasing or decreasing the indexes (one or more) of one unit obtaining new values (I',j',k') . The neighborhood consist of 26 cells (9 on the upper slice, 9 on the lower one, and 8 on the same level as the invader). If the picked element is a 'H' cell the corresponding element (I',j',k') of a matrix called 'invasione' is set to 1, otherwise the process is repeated. If the invader fails the invasion 20 times, which means that is probably already surrounded by 'C'-cells) it becomes quiescent for a certain time period. When 'invasion.m' is called all the whole structure is scanned to complete the invasion matrix.

In the end there is the visualization function that shows in red the cancerous cells and in green the necrotic tissue due to the effect of the irradiation. The detailed complete scheme is showed in Figure 22.

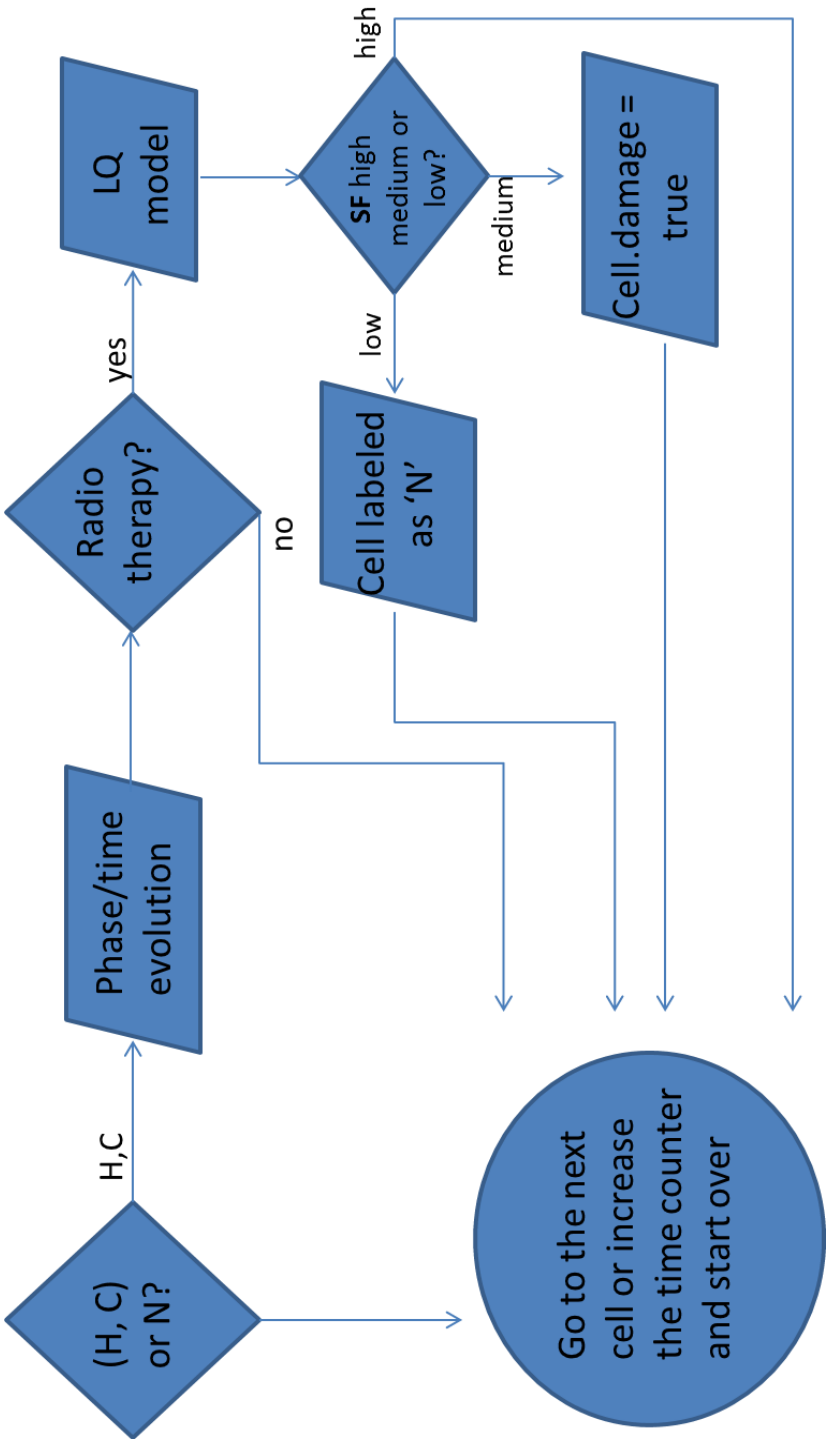

Figure 21

Single cell evolution and radiation therapy module scheme: the necrotic cells (N) do not evolve and it is not further damaged by radiotherapy, while the cancer and healthy cell have two different life cycles. When a certain dose of radiation is delivered the effect vary according to the kind of cell that it reaches,both healthy (H) or cancer (C) cell get decrease their 'numel' according to the Linear Quadratic model formula when struck by the dose, but the parameters are different for the two kind of population: Cancer cells are supposed to be more sensitive otherwise there would be no gain in using radiation therapy. Moreover if most of the cell population die due to the incoming dose the whole cell is marked as necrotic, if only a percentage survives it will stay 'damaged' for a certain amount of time, therefore it will be more sensitive to a new treatment (the damaged cells have higher LQM parameters).

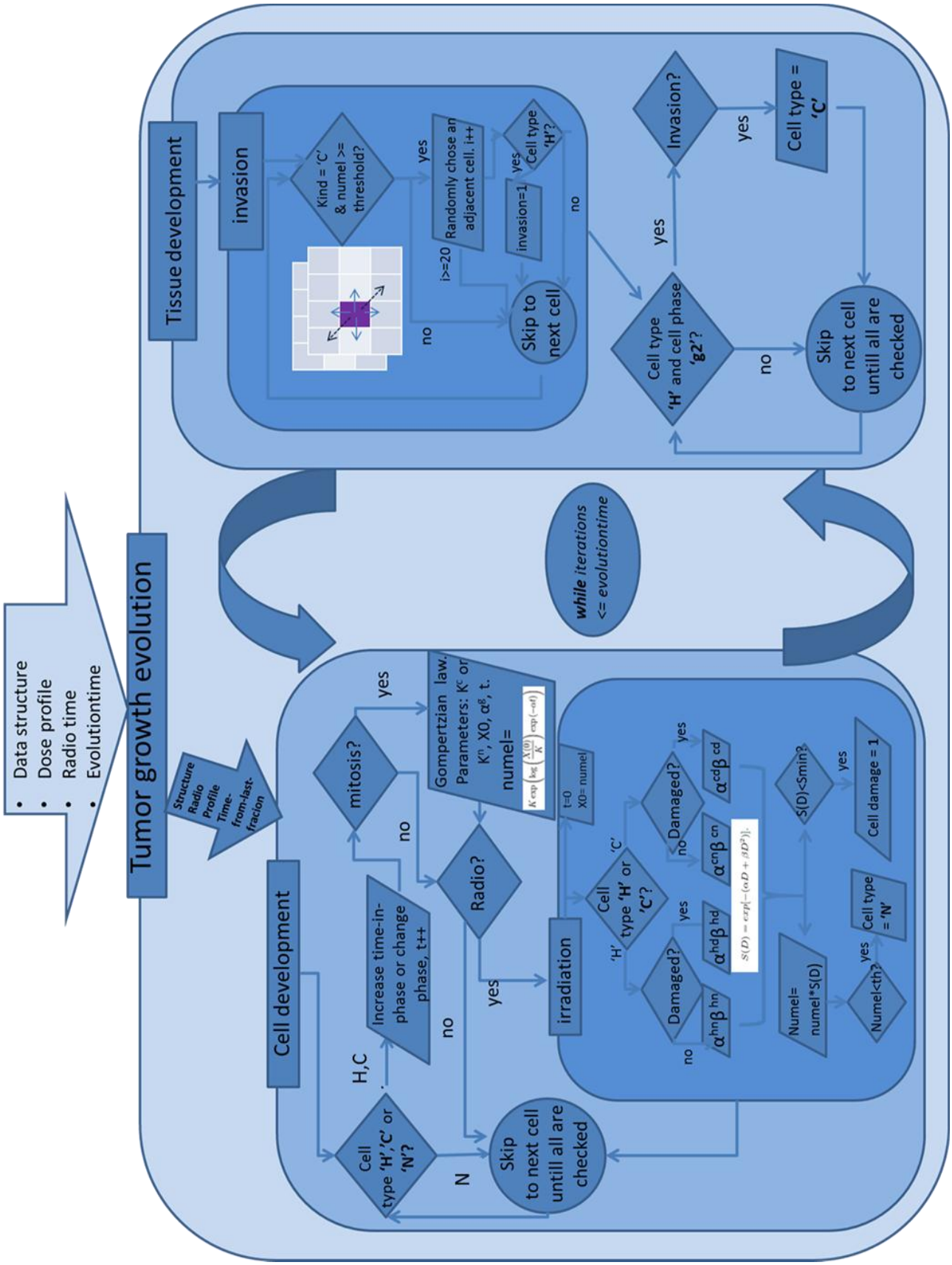

Figure 22
Complete cellular automata scheme

3.3.3 Parameters

A complex issue that arose together with the computational effort implementing the Cellular Automata model is the parameters setting, the most critical ones have been collected in Table 6.

Symbol	Parameter	Description
cd	Cell cycle duration	It takes into account the time spent in each phase of the cell cycle.
k	Gompertzian parameter	Defines the value that the curves tends to achieve (max cell density)
X(0)	Gompertzian parameter	Represents the starting point (initial density value)
α	Gompertzian parameter	Represent the slope, and influences the time the curve takes to get to its maximum
α_h, β_h	Linear-quadratic model parameters	Model the radio-sensitivity of the healthy cells
α_{hd}, β_{hd}	Linear-quadratic model parameters	Model the radio-sensitivity of the healthy cells that have been damaged
α_c, β_c	Linear-quadratic model parameters	Model the radio-sensitivity of the cancer cells
α_{cd}, c_d	Linear-quadratic model parameters	Model the radio-sensitivity of the cancer cells who have been damaged
r_t	Recovery time	Time that a cell takes to recover when damaged by the therapy
q_t	Quiescence time	Time that a cell stays in a quiescent status before trying again to invade other tissues
d_{th}	Density threshold	If the cell density of tumor site overcomes this threshold it could invade the neighbourood
n_{tri}	Number of trials	Maximum number of trials allowed to invade surrounding sites before becoming quiescent

Table 6
Parameters of the Radiotherapy Model

Considering, for example, the Gompertzian curve (Equation 19) we can observe:

- X(t) is the number of cells at time t in a certain geometrical cell;
- X(0) is the starting number (or the number of cells remaining after the last irradiation dose delivered);

- K is the maximum value of cell density it can reach;
- αg influence the speed to reach K.

All of them have to be set either accordingly to values of cellular densities already studied in literature (Lyng H. et al., 2000) or using an optimization method and fitting real data. Anyway the role of the parameters is critical and the literature values could be misleading, therefore a huge amount of data would be needed in order to tune them all and obtain some meaningful results. Unfortunately this is not the case of the rather small dataset already described, this is why we decided to focus our attention on a simpler approach.

$f(t, P)$		
Curve name	Parameters	Formula
Gompertz curve	2 (a,b)	$a * \log\left(\frac{b}{P}\right) * P$
Logistic curve	2 (a,b)	$a * \left(1 - \frac{P}{b}\right) * P$
$g(t, P)$		
Model	Parameters	Formula
Model 1	3 (c,f,g)	$-c\ P\ R(1 - e^{\left(-fD(t-t_i)-gD^2(t-t_i)\right)})$
Model 2	2 (c,f)	$-c\ D\ P\ e^{\left(-f(t-t_i)\right)}$
Model 3	1 (c)	$-c\ D\ P\ (t - t_i)$

Table 7
Possible combinations of growth curves and therapy formulation

3.4 ODE models

3.4.1 Model definition

The most common formulation for the tumor growth curves (f(t,P)) are the so-called Gompertzian (Equation 21) and Logistic (Equation 22) equations, both of them represent a value P that increases non-linearly trough time going towards an asymptotic value (b) with a certain growth rate (a). This behavior mimics the occurrence of necrosis and other mechanism due to the lack of nutrients and oxygen which prevents the tumor from growing when the when its size increases too much.

Anyway, since the cervical cancer is usually treated by means of irradiation and chemo therapy a term (g (t,P)) that considers the effects of the treatment is needed in order to catch the global dynamic of the

system. Therefore we proposed a combination of two factors as showed in the Equation 20, where f(t,P) epresent the uncontrolled growth and g(t,P) the shrinkage due to the dose delivered to the mass. To give an example we quote the first therapy-term suggested, linked to the LQ model formulation (Equation 23). The esponential part $E=e^{\left(-fD(t-t_i)-gD^2(t-t_i)\right)}$ represents the surviving fraction calculated using the LQ model (D is the dose, f and g the alpha and beta coefficients respectively), since we want to now the decrease we needed to compute (1-E) which had also to be multiplied by the population value (P) and a rescaling factor (c) in order to get the volume or area.

$$\frac{dP}{dt} = f(t,P) + g(t,P) \tag{20}$$

$$\frac{dP}{dt} = a * \log\left(\frac{b}{P}\right) * P \tag{21}$$

$$\frac{dP}{dt} = a * P * \left(1 - \frac{P}{b}\right) \tag{22}$$

$$g(t,P) = -c\ P\ R\left(1 - e^{\left(-fD(t-t_i)-gD^2(t-t_i)\right)}\right) \tag{23}$$

All the possible combinations of the two terms are summarized in Table 7.

3.4.2 Data fitting

The optimization has been achieved separately for the two groups of patients, we were looking for a class-model, not a patient specific one, in this way we could achieve a general knowledge about the problem. This approach should allow us to predict the volume (and area) evolution of a new patients with the same pathology and scheduled treatment even if its images were not present in the training dataset. The procedure can be summarized as follows:

1. Chose the formulation of the two terms
2. Set the value of the parameters
3. Calculate the values of the function at the time of each acquisition (i) with implicit or explicit numerical methods (Eulero, Runge-Kutta, Crank-Nicolson)
4. Evaluate the mean square error (ε, Equation 24), which means to calculate the root mean square sum of the squared difference between the real volume size ($\bar{V}_i$) and its prediction achieved using the model (V_i)
5. These operations were repeated several times changing the values of the parameters using a genetic algorithm, in order to select the combination that minimizes the error for each kind of model.

$$\varepsilon = \frac{1}{N}\sum(V_i - \bar{V}_i)^{2^{\frac{1}{2}}} \quad (24)$$

A genetic algorithm was implemented in MatLab ™ in order to compute the best parameters of the model (Figure 23).

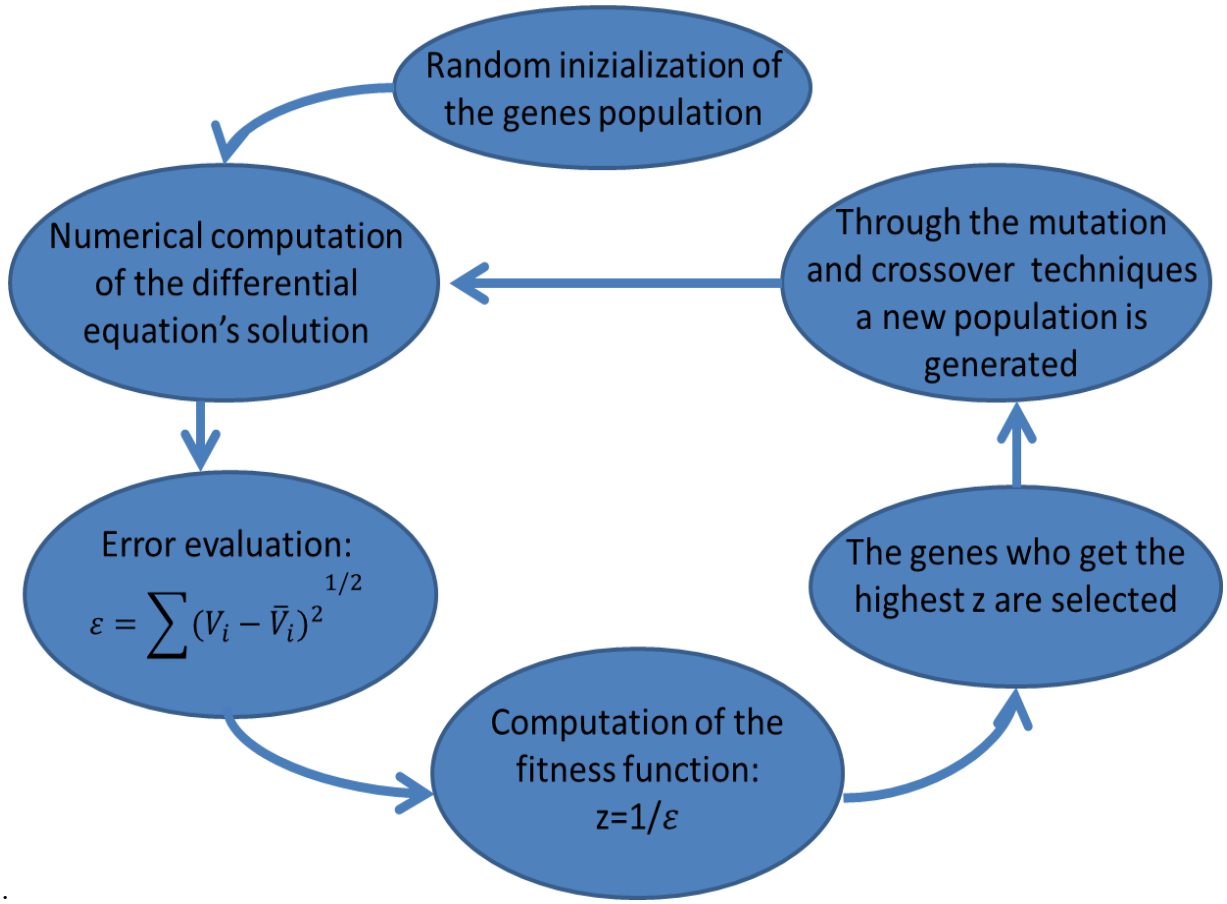

Figure 23
Genetic Algorithm
Each gene is a binary codification of the parameter's values, $\boldsymbol{V_i}$ is the volume of the tumor predicted by the model at time t_i, $\bar{\boldsymbol{V}}_{\boldsymbol{i}}$ is the collected volume value at time t_i The same cycle is implemented for the area models

Each gene consisted of the sequence of parameters that needed to be set, they are expressed in binary code to allow the most common mutation techniques. During the evolution the genes who get the best results of the fitness function were selected and randomly mutated in order to explore new combinations even far from the starting point or mixed using the crossover techniques which does not add new information but allows to tune the parameters. In this case the fitness function (z) was inversely proportional to the sum of the errors made by the model predicting the volume (or area) at every acquisition time, since we want the fitness function to be max when the error is as small as possible, we define $z=1/\varepsilon$ and ε as already formulated in Equation 23. The calculations took some time because for each gene, which means each combination of parameters, and each time step, a numerical ODE solver had to be called in order to calculate the model outcome. This approach should lead to the values of the parameters that best fit the data but it could get stuck in a local maximum in case of a too small mutation rate, or it could escape from the right direction of maximization of the fitness function

if each generation differs too much from the previous one (high mutation rate). To check if the solution is reliable we tried the code more than once choosing the best solution among the ones found.

3.5 Stability analysis and validation of the model

A number of issues prevented us from applying the Cellular Automata models to the real data, the fitting part and the analysis carried on for the ODE models were not achievable in case of the discrete model. First of all, a lot of information concerning cellular level as for example the presence of blood vessels (source of nutrients and oxygen) or the histological structure of the region of interest, where missing,. Second, the limited cohort was a sever limit for such a complex model, the data available where not enough to provide a meaningful settings of the CA parameters. Third, the computational effort needed to simulate the evolution of the tumor using this approach was quite massive due to the level of detail of the representation, so it was not worth trying because of the just highlighted limits. In conclusion the methods showed in the following lines are only referred to the continuous approach presented in paragraph 3.4.

In order to understand the role of each parameter in the overall prediction the ***Linear Sensitivity Analysis*** was performed using the models that resulted in the least prediction-error for each class of patients. This technique consist in perturbing one parameter at a time leaving the others unvaried. The change in the predicted index (volume or area) over the variation of the single parameter can be thought to be an approximation of the derivative. This information can be used to understand the necessity of an accurate setting and the possible effect of an error in the pursuit of the best coefficients.

The ***leave-one-out*** technique is useful to analyze the behavior of the model when used to predict a set of data it was not trained on. In order to achieve this goal the coefficient were set using all but one patients for each class cyclically. The data belonging to the remaining woman were used to test the combination of parameters that turned out to be optimal for the rest of the class. The average prediction-error was expected to be higher on the test patients.

The outcomes of both these analyses will be discussed in the next two chapters.

4 Experimental results

The results achieved, as expected, are different for each class of patients, and of course depend also on the type of ODE model used to fit the data. We remind the general formula used to mimic the evolution of K (area or volume):

$$\frac{dP}{dt} = f(t,P) + g(t,P)$$

In the following paragraphs it will be illustrated which is the combination of terms that best represented each group, and some possible explanations will be presented in the discussion part. Anyway, in order to understand if the assumption of radial symmetry is plausible we first analyzed the V/A ratio . Considering that the area of a sphere (As) of a given radius (r) is:

$$As = 4 * \pi * r^2$$

While its volume (Vs) can be calculated by the formula :

$$Vs = \frac{4}{3} * \pi * r^3$$

The ratio $\frac{V^{1/3}}{A^{1/2}}$ should be a constant value but it changes if the solid elongates in one direction. Using this indicator, we verified that the shape does not change much for patients 6, 10 (variance <7e-5) while it has a limited variation in the other cases(var<3e-4)

4.1 Errors and parameters comparison

4.1.1 Group 1: Squamous cell carcinoma (RT only)

In this paragraph the main results for the first class of patient are shown, it is very interesting to see how different are the numbers achieved by alternatives ODE formulations. One patient was removed because of its bad reaction to the chemotherapy treatment which had to be interrupted early; those data could have been a source of error in the search for the best parameters for the class.

	model1		model2		model3	
Group1	eps	epsrel	eps	epsrel	eps	epsrel
Volume	5.54E+03	2.83E-01	5.28E+03	2.71E-01	6015.877	3.97E+03
Area	1.11E+03	2.58E-01	1.04E+03	2.35E-01	1.08E+03	2.38E-01

Table 8

Results achieved for the first class of patients both for area and volume prediction using the Gompertz equation as evolution term but different formulas to model the therapy.(eps= absolute error in mm^2 or mm^3; eprel=relative error)

	model1		model2		model3	
Group1	eps	epsrel	eps	epsrel	eps	epsrel
Volume	8.32E+03	4.85E-01	6.20E+03	3.51E-01	6.08E+03	3.40E-01
Area	1624.005	0.373141	1562.327	0.353612	1537.18	0.376114

Table 9

Results achieved for the first class of patients both for area and volume prediction using the Logistic equation as evolution term but different formulas to model the therapy.(eps= absolute error in mm^2 or mm^3; eprel=relative error)

Table 8 summarize the result achieved using as first term the Gompertz model while Table 9 does the same for what concerns the Logistic one. It is worth notice that the smallest errors were achieved using the combination of the Gompertzian curve and the therapy module number 2 both for Area and Volume. Anyway there are a few models whose errors do not differ much. The percentage of the errors made by the best models are respectively 24% for the area and 27% for the volume, but this values have to be compared to the high noise level of the data. Some example of the quality of the fitting achieved with the combination of the Gomperzian curve and the exponential model can be seen in the following pictures (Figure 24, Figure 25). It is plain to see that the prediction error increases during the evolution time, this might imply the occurrence of delayed mechanism we did not consider in our two-terms formulation of the tumoral dynamic

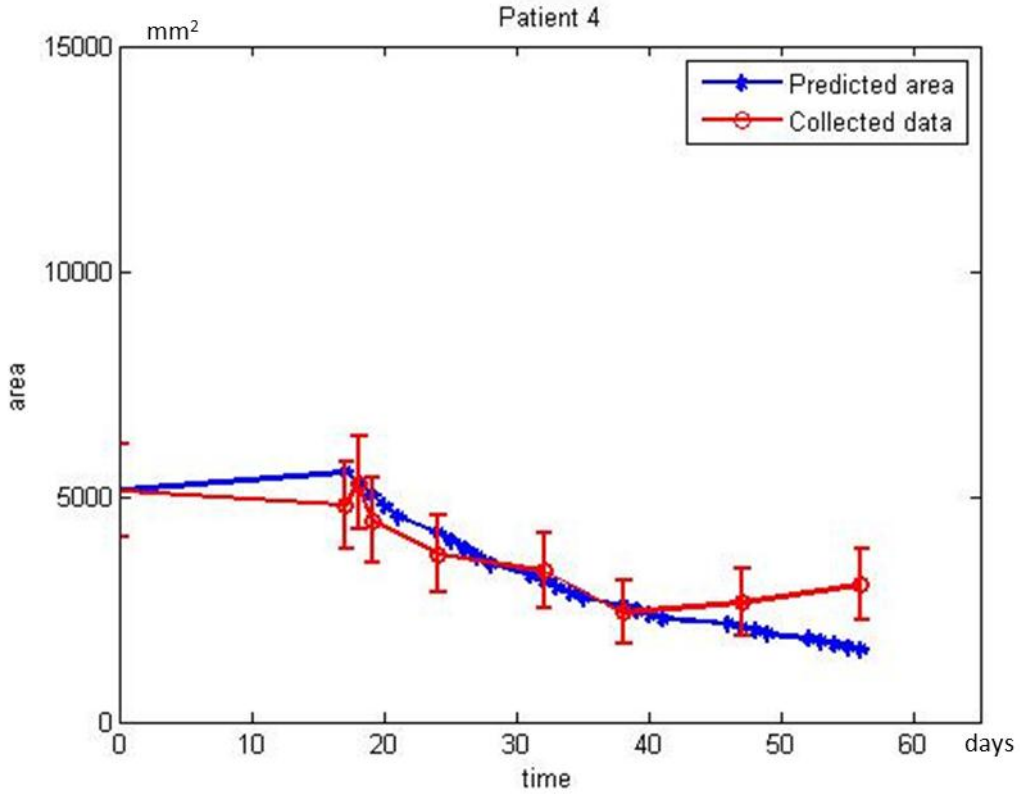

Figure 24

Area fitting example: Patient 4 (Gompertz curve and exponential response to the delivered dose)

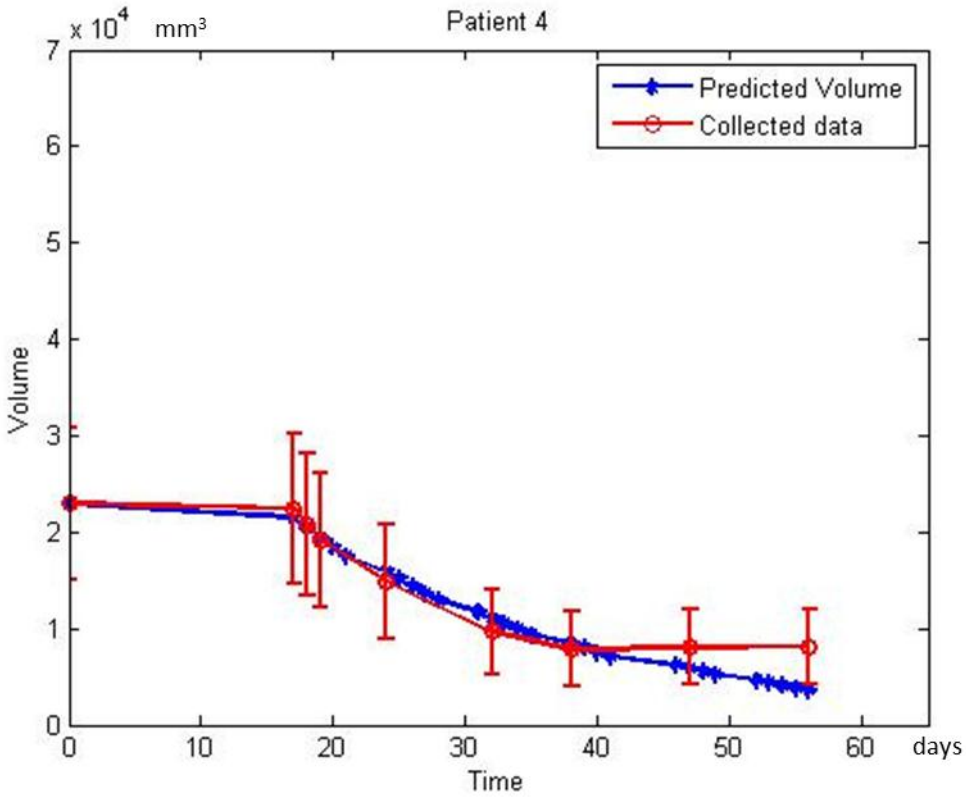

Figure 25
Volume fitting example: Patient 4 (Gompertz curve and exponential response to the delivered dose)

4.1.2 Group 2: Adenocarcinoma (RT+CT)

Group 2 is the one that involves patients affected by Adenocarcinoma cervical tumor who undergo both radiotherapy and chemotherapy . The following tables (Table 10, Table 11) show how restrained are the errors achieved with both the Gompertzian model and the Logistic one. The combination of the Gomperzian curve and the third therapy model, linearly linked to the dose, (Table 10, red value) resulted in the least relative error as far as it concerns the volumetric data. On the other hand, the area evolution pattern is better represented by the logistic model plus the first therapy module (LQ model). The values of the errors achieved with this group are much smaller, 16% and 21% for area and volume fitting respectively, this might be due to the limited number of patients belonging to this class. This is the only case where the best results are achieved using the logistic curve. The graphs (Figure 26, Figure 278)show some examples of the area and volume fitting achieved.

	model1		model2		model3	
Group2	eps	epsrel	eps	epsrel	eps	epsrel
Volume	5.18E+03	2.53E-01	4.57E+03	2.09E-01	3.97E+03	2.07E-01
Area	9.55E+02	2.05E-01	1.18E+03	2.44E-01	9.02E+02	1.97E-01

Table 10
Results achieved for the first class of patients both for area and volume prediction using the Gompertz equation as evolution term but different formulas to model the therapy.(eps= absolute error in mm^2 or mm^3; eprel=relative error)

	model1		model2		model3	
Group2	eps	epsrel	eps	epsrel	eps	epsrel
Volume	4.64E+03	2.44E-01	6.17E+03	3.17E-01	1.03E+04	5.15E-01
Area	710.9431	0.164341	983.6704	0.201244	2284.318	0.48245

Table 11

Results achieved for the first class of patients both for area and volume prediction using the Logistic equation as evolution term but different formulas to model the therapy.(eps= absolute error in mm^2 or mm^3; eprel=relative error)

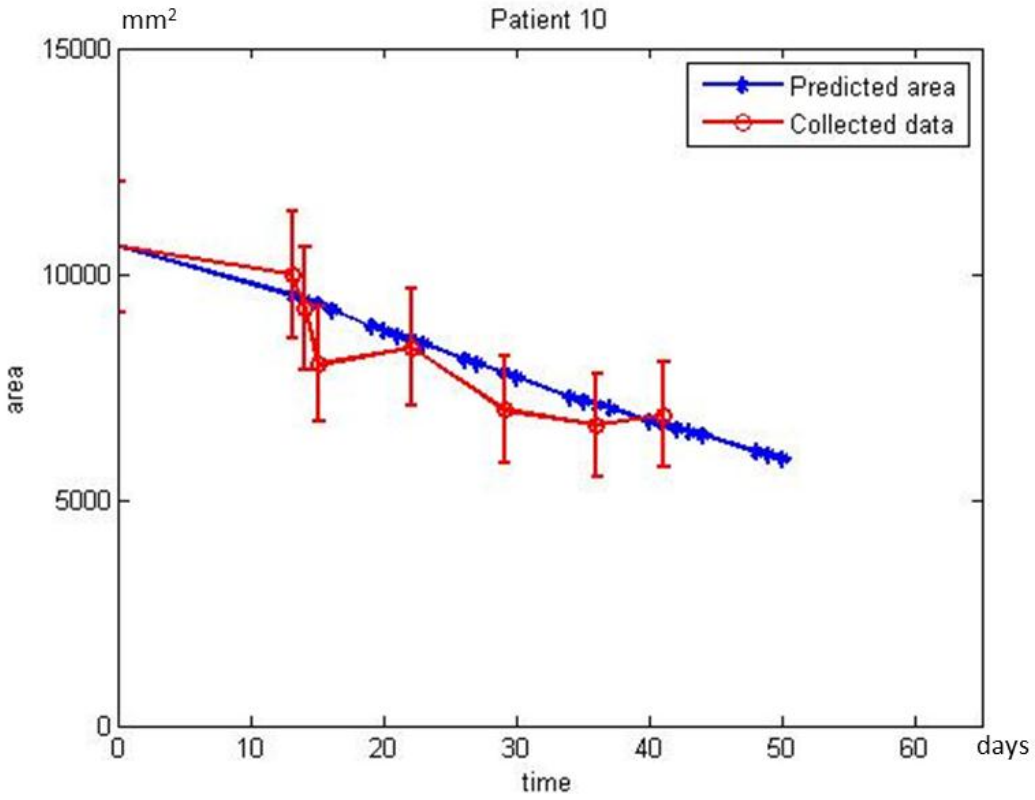

Figure 26

Area fitting example: Patient 10 (Logistic curve and LQ model)

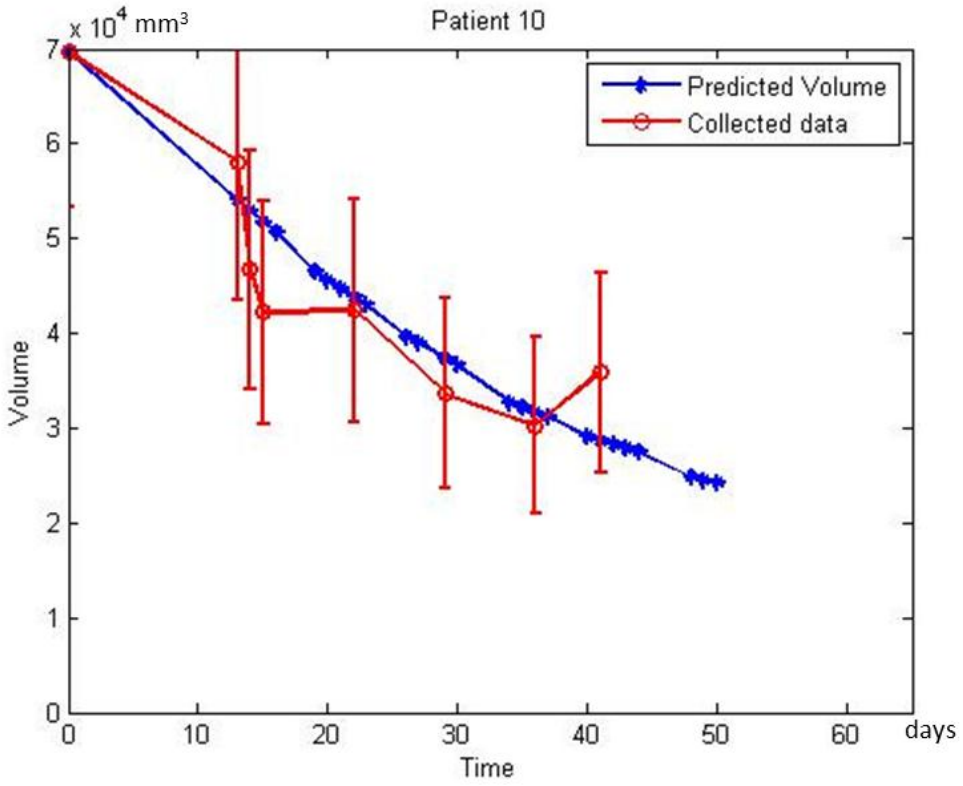

Figure 27
Volume fitting example: Patient 10 (Gompertz curve and linear response to the delivered dose)

4.1.3 Coefficient comparison

The coefficient's values obtained for the most successful combinations of the models (Table 12) were analyzed in order to understand and explain the differences among groups. The 'a' and 'b' coefficients both in the Gomperzian and in the Logistic model corresponded respectively to the growth rate and to the asymptotic value that should be reached by the tumoral mass in case of no therapies. The models presented sometimes an unexpected slightly negative growth rate (Table 13), this may be due to a very slow dynamic of the tumor invasion compared to the effects of the radiation therapy, meanwhile the 'b' coefficient varies significantly in accord to the selected model and the class of patients analyzed. The coefficient named 'c' generally represents the incidence of the treatment and assumes almost identical value both for the area (0.039) and volume (0.040) for the group of patients affected by the Squamous cell Carcinoma, while in case of the second group affected by the Adenocarcinoma its value for the area development is sensibly higher(0.0078) than for the volume (0.0039), this peculiarity reinforce the fact that volume and area follow similar dynamics when only radiation therapy is delivered while the two patterns differ in case of the introduction of a non-uniform treatment such as chemotherapy. In Model1 'f' and 'g' can be thought as the alpha and beta parameters of the linear quadratic model therefore they represent respectively the cell death due to a single event or to a summation of chain-breaks close in space and time.

	Volume	Area
Group 1	Gompert, RT2	Gompertz, RT2
Group 2	Gompertz, RT3	Logistic, RT1

Table 12
Growth and Therapy's modules combination that resulted in the smallest prediction error for each class of patients both for volume and area fitting

	AREA		VOLUME	
parameters	**Group1**	**Group2**	**Group1**	**Group2**
a	0.011811	-0.009804	-0.006387	0.011811
b	7843	61176	40000	11765
c	0.039216	0.007843	0.040000	0.003922
f	0.556863	-0.031373	0.926196	-
g	-	-0.243137	-	-

Table 13
Coefficient founf througn the genetic algorithm that led to the best results for each class of patients both for area and volume

4.2 Sensitivity to parameters variation

The role of each parameter was analyzed separately in order to understand the effects of a suboptimal setting of the model. The four models that resulted in the smallest error for each group both for area and volume (Table 12) were tested with different parameters combination:

1. One parameter was perturbed of a percentage 'h' (± 2% or ±5%) while the others have been left unvaried
2. The new predictions are calculated and plotted
3. The variation of the output (Volume or Area)over the variation of the input (coefficient) can be considered as the partial derivative of the output on that specific parameter
4. Using the infinity norm the most critical parameter can be identified for each patient trough time.

Noteworthy the same results have been achieved using the two percentages of variation (2% and 5%). We plotted for each patient the highest value of the derivative at each time step. It is interesting to notice that the curves obtained increase trough time and that the critical coefficients always are the first one, which represent the growth rate of the uncontrolled tumor, and the third one, that weights the effect of the therapy (Figures 28, 29, 30, 31). This two parameters have to be carefully set in order to get reliable results.

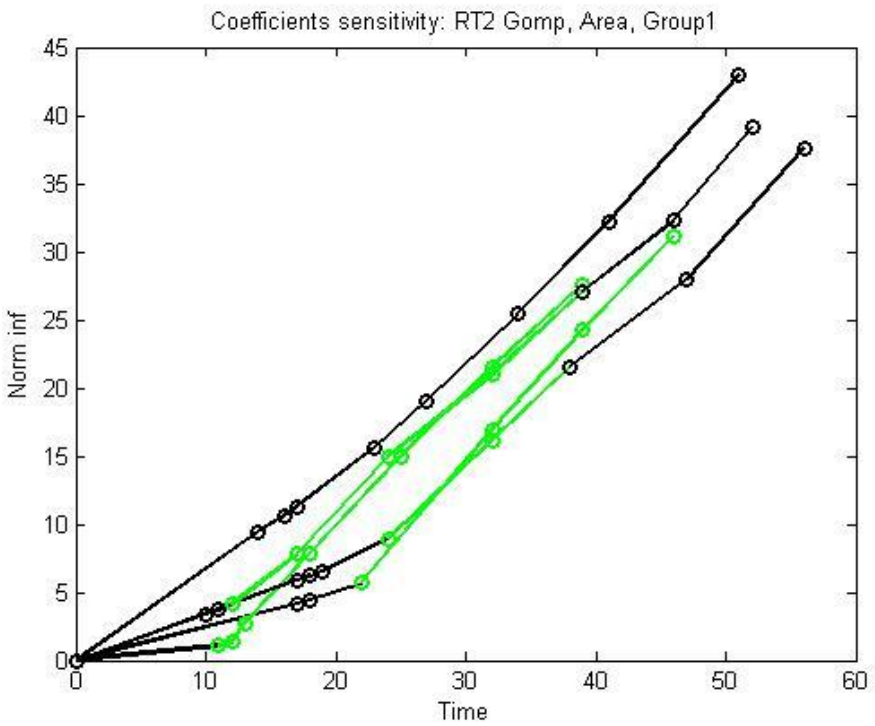

Figure 28

In this figure the infinity norm of the derivative's vector has been plotted trough time for each patient of Group1. The green dots correspond to the 'c' coefficient while the black ones to the 'a' parameter. Time is expressed in days while the norm is an absolute value

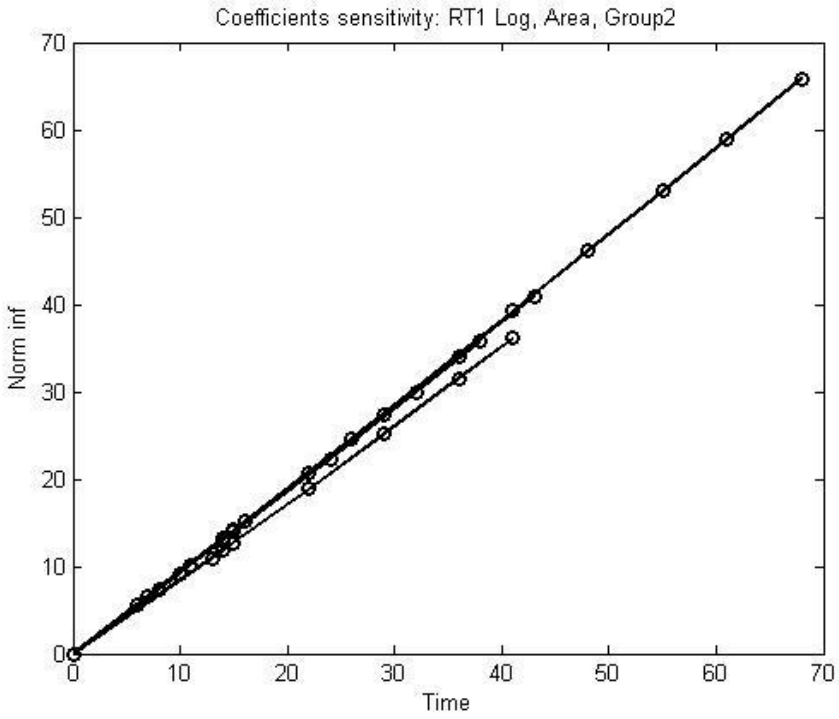

Figure 29

In this figure the infinity norm of the derivative's vector has been plotted trough time for each patient of Group2. The green dots correspond to the 'c' coefficient while the black ones to the 'a' parameter. Time is expressed in days while the norm is an absolute value

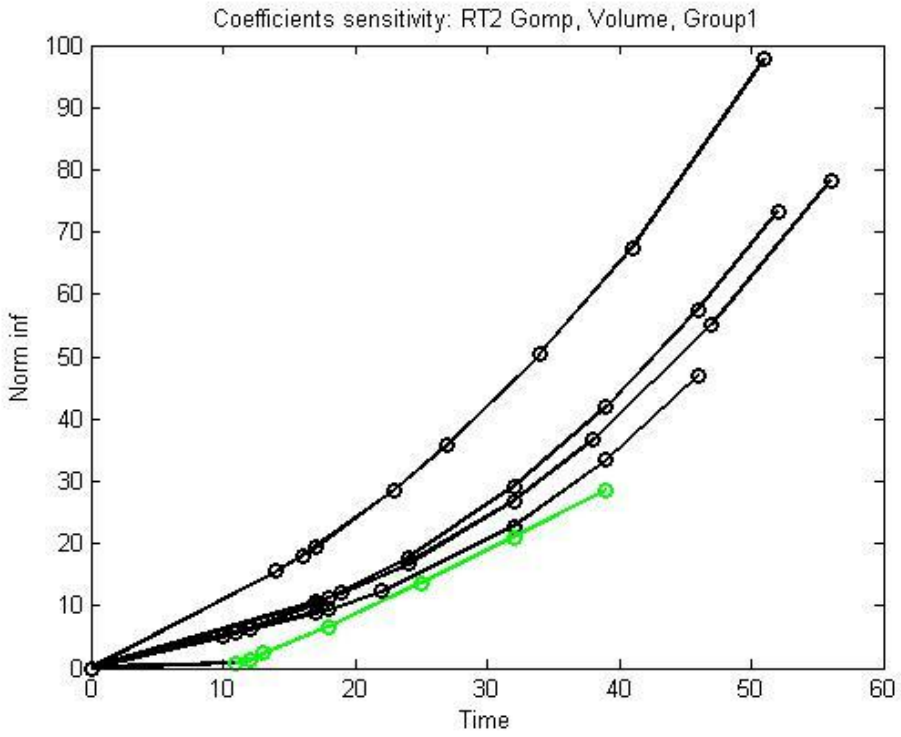

Figure 30

In this figure the infinity norm of the derivative's vector has been plotted trough time for each patient of Group1. The green dots correspond to the 'c' coefficient while the black ones to the 'a' parameter. Time is expressed in days while the norm is an absolute value

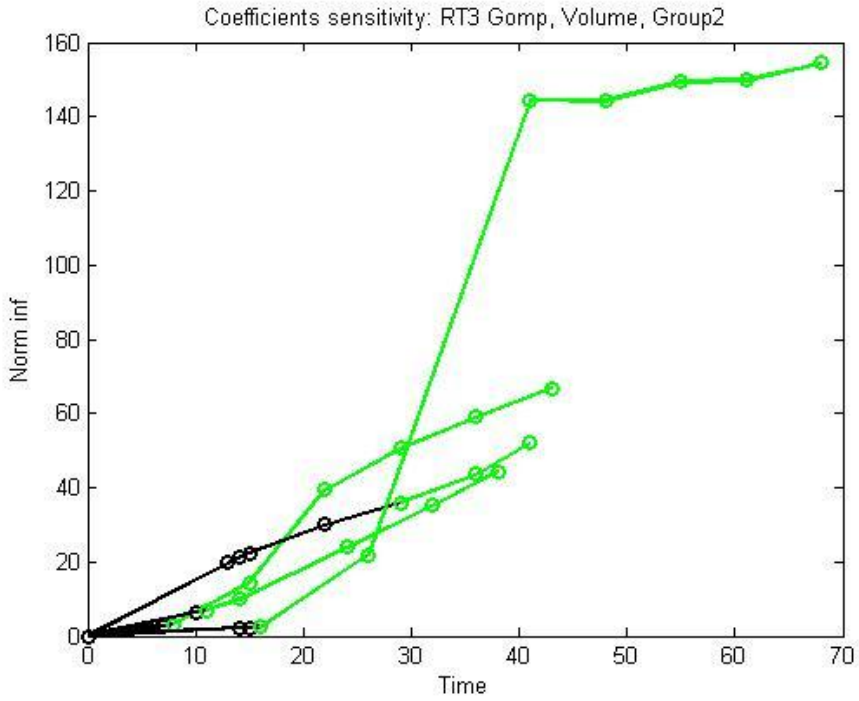

Figure 31

In this figure the infinity norm of the derivative's vector has been plotted trough time for each patient of the Group1. The green dots correspond to the 'c' coefficient while the black ones to the 'a' parameter**.** Time is expressed in days while the norm is an absolute value.

4.3 Validation leave-one-out

The technique of leaving out of the class one patient at a time can both show if the solution is able to catch the general evolution pattern and if there are outliers among the patients. Actually, since the numbers we are dealing with are too small to allow a meaningful statistic, this approach was likely to lead to much bigger errors on the test patients than on the others. Each class of patients were split into two clusters: one consisting of only one patient which was used to validate the model, while the other, containing all the remaining population, was used to train the coefficient adopting the same genetic algorithm as the one already presented in this chapter. Each patient was excluded cyclically and the search for the optimal parameters was repeated. Observing the results (Figures 32, 33, 34, 35) it can be seen that most of the values obtained for the test-patients are comparable and similar to the one achieved considering the whole group. It turned out that the error on the patient left out was generally larger than the one that affects the rest of the class, but if we analyze the single cases there are sometimes models that are able to fit some of the new data even better than the one they were trained on; this is the reason why the error variation is much bigger among the test patients than on the training class.

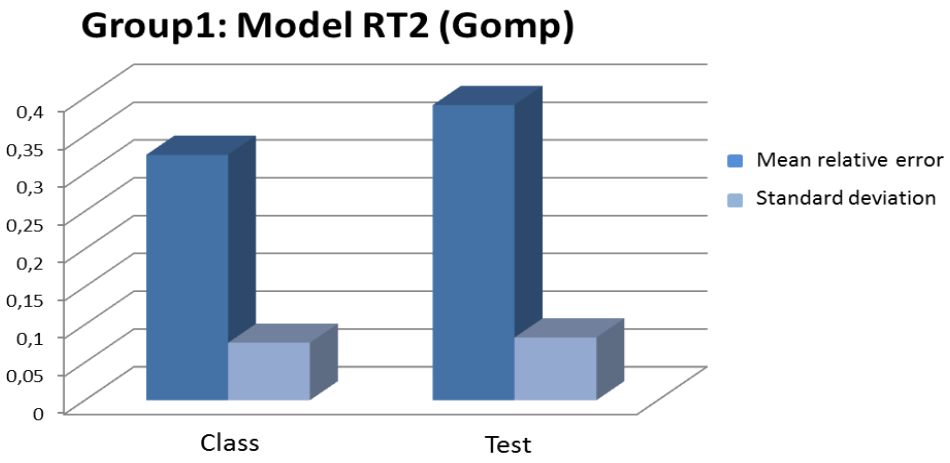

Figure 32
Leave one out validation: Group1 (Volume) error comparison between class and test

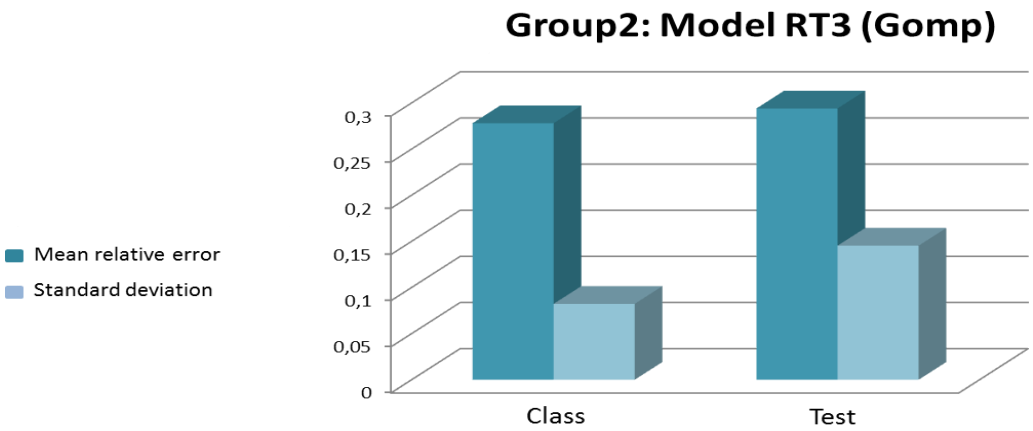

Figure 33
Leave one out validation: Group2 (Volume) error comparison between class and test

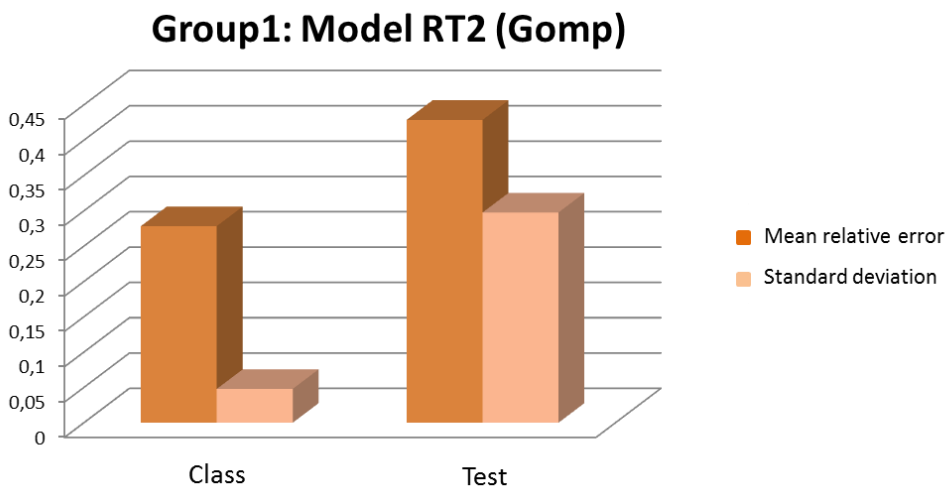

Figure 34
Leave one out validation: Group1 (Area) error comparison between class and test

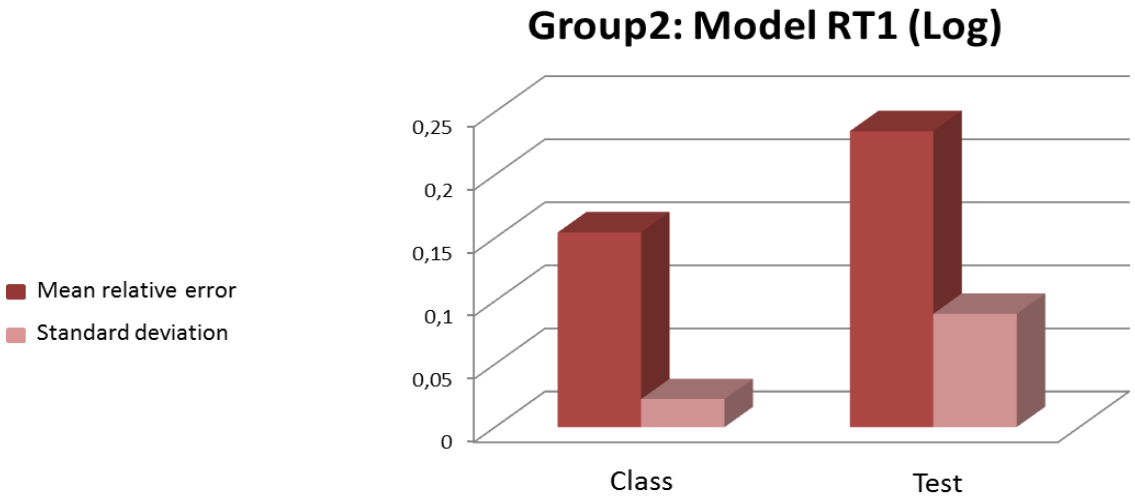

Figure 35
Leave one out validation: Group2 (Area) error comparison between class and test

5 Discussion and conclusions

In this work many important issues have been discussed and some preliminary results have been achieved. The information gathered through the elaboration of a real set of data might be a precious starting point for further analysis on tumor growth and reaction to therapies in case of cervical cancer.

5.1 Discussion

The errors percentage obtained both on the area (24%group1, 16% gruop2) and on the volume (27%group1, 21% gruop2) fitting are quite promising since close to the range of the data uncertainty, qualitatively about 10% and 20% for area and volume, respectively. Moreover such results were basically coherent with recent domain literature (Huang Z. et al., 2010) where relative prediction errors of about 25% were reported. It is worth notice that volume and area prediction criteria conflicted in between for the second group: while the combination of the Logistic curve and the linear/quadratic model turned out to be the most performing on the area, the volume was better fitted using the Gomperzian model along with the linear model of the radiation response. The different behavior of the two indices can be explained considering that chemotherapy is a localized treatment because drugs are delivered by means of blood vessels which are more densely localized on the outer surface of the tumor (Figure 36). This peculiarity could cause a major relevance of the events occurring to the most superficial layer of cells.

The Linear sensitivity analysis (LSA) showed how for all the patients, both for area and volume, the most critical parameter is the uncontrolled growth rate, at least for the first period of time while in a second moment it is the coefficient that weighs the therapy that takes the lead. This might imply that a certain period of time is needed before the therapy affects the evolution of the mass, anyway, in some cases, especially as far as it concerns the area of the second class of patients, the growth parameter is the one that affect most the outcome during all the evolution time. This apparently unpredictable behavior can be explained noticing that the growth-rate coefficient is actually negative, this probably means that the effects of the therapy overcame the natural growth tendency of the tumor so early that the model was not able to divide the two dynamics and the parameter that weights the first term, which includes also the therapy, turns out to be the most critical. The outcomes achieved through the validation leave-one-out encouraged the idea that even with a limited pool of subject analyzed the parameters met have quite a general usability among the patients of the same class although their data were not used to find the coefficients themself.

A number of issues in this work can be highlighted.

1. the tumor growth and therapy response was represented using continuous equations (ODE), limiting the description to the tissue level and macro-scale mechanisms. Micro-scale processes at cellular level, as for example cell cycle, cell-to-cell interaction, and tumor cell invasion, were modeled using cellular automata but not validated;
2. the spatial dimension of the mass and its location were neglected and only scalar variables (area and volume) were considered, approximating the tumor shape as a spheroid. This allowed to avoid registration across CBCT studies but not to collocate the tumor progression in a common coordinate system;

3. the patient cohort was limited in size and this prevented us from getting any meaningful statistical evidence about the quality of the prediction results and model performance;
4. the uncertainty on the data used for the fitting and validation was not evaluated in a rigorous way which is fundamental in order to assess the quality of the prediction obtained;
5. the nominal value of the dose delivered to the patient was used despite the potential occurrence of errors due to tumor shape modification, soft tissue deformation, physical interaction with the surrounding organs and the effects of the gravity.

Nonetheless, despite such shortcomings, the presented work showed how even a not complex macrosopic model can be effective in predicting the interaction between tumor progression and the effects of radiation therapy. In conclusion, while being a preliminary study, the results are meaningful and promising for extending the methodology to include more sophisticated representation of the biological processes and therapy interaction, and making the approach suited for more extensive validation and future clinical translation.

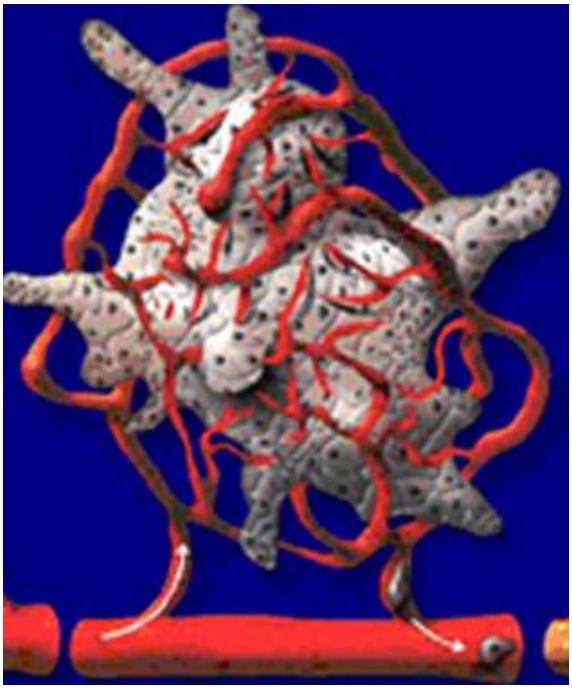

Figure 36
Superficial vascularization of the tumor

5.2 Future developments

In order to overcome the previously quoted drawbacks we would like to propose some solutions:

1. optimize and evaluate the performance of the cellular automata model presented in paragraph 3.4 or to use a third approach able to model the cellular level, and to compare its results with the one already achieved. Anyway the different level of detail of the representation would require a more complex set of data;

2. move to a different kind of model could also allow the 3D representation of the tumor, overcoming the necessity of spherical approximation made using the ODE, but the registration of all the CBCT images to the CT image would be needed;
3. collect a larger amount of clinical data (e.g. contouring all the acquired images or involving a greater number of patients) in order to obtain meaningful statistics;
4. set a protocol for assessing the error made by the physician while contouring the region of interest, in order to get an objective uncertainty value;
5. calculate the actual dose delivered to the patient considering tissue deformation;

An interesting idea could be to use the ODE models as a patient specific tool, adjusting the coefficients from the most generic ones computed for the whole class of patients to a more accurate set tuned on the tumoral evolution of a single patient while the therapy is delivered. In this way the planning can be modified accordingly to the specific needs and characteristics of the subject.

Bibliography

American Pathologists (2011). Cervical Cancer - Cervical Adenocarcinoma.

Barazzuol L. et al. (2009). A mathematical model of brain tumour response to radiotherapy and chemotherapy considering radiobiological aspects. *JournalofTheoreticalBiology*.

Bellomo N., et al. (1999). *Modelling and Mathematical Problems Related to Tumor Evolution and Its Interaction with the Immune System*. Mathematical and Computer Modelling.

Boondirek A. (2010). *A Review of Cellular Automata Models of tumor growth*. International Mathematical Forum.

Boondirek A., et al. (2006). *A Stochastic Model of Cancer Growth with Immune Response.* Journal of the Korean Physical Society.

Borkestein K. et al. (2004). *Modeling and computer simulation of tumor growth and tumor responce to radiotherapy.* Radiation research

Burton A. C. et al. (1966). *Rate of growth of solid tumours as a problem of diffusion.*

Dolinsky C. et al.. (2011). *Cervical Cancer: The Basics*. OncoLink: Abramson Cancer Center of the University of Pennsylvania.

Clatz O. et al. (2005). *Realistic simulation of the 3d growth of brain tumors in mr images coupling diffusion with biomechanical deformation*. Medical Imaging

Cobb K. et al. (2007). *Modeling Cancer Biology.* Biomedical Computation Review

Cruywagen G. C. et al. (1995). *The modelling of diffusive tumours. J. Biological Systems*.

Deroulers C. et al.(2009). *Modeling tumor cell migration: From microscopic to macroscopic models*. Physical review

Dordal M.S. et al. (2005). *Flow cytrometric assesment of the cellular pharmacokinetics of fluorescent drugs*. Cytometry.

Duchting W.et al. (1985). *Recent progress in modelling and simulation of three-dimensional tumor growth and treatment*. Biosystems.

Fister R. K. et al. (2000). Optima control applied to cell-cycle specific cancer therapy.

Folkman J.. et al. (1966). *Tumor behavior in isolated perfused organs: in vitro growth and metastases of biopsy material in rabbit thyroid and canine intestinal segment.* Annals of surgery

Frisch U. et al.(1986). Lattice-gas automata for the Navier-Stokes equation.

Garbey M. et al.(2011). *Multi-scale modeling and computational surgery: application to breastconservative therapy.* Journal of the Serbian Society for Computational Mechanics

Gatenby R. A. et al. (1995). *The potential role of transformation-induced metabolic changes in tumor-host interaction*. Cancer research

Hausen, H. z. (2002). Papillomaviruses and cancer: from basic studies to clinical application. *Nature Reviews Cancer*, 342-350.

Hill A. et al.(1928). *The diffusion of oxygen and lactic acid through tissues.* Proceedings of the Royal Society of London

Huang Z et al.. (2010)Predicting outcomes in cervical cancer kinetic model of tumor regression during radiation therapy. *Cancer research*.

Hwang M et al. (2009). Rule-Based Simulation of Multi-Cellular Biological Systems—A Review of Modeling Techniques. *Cellular and Molecular Bioengineering*, 285–294.

Konukoglu E. et al.(2008). *Modeling Glioma Growth andPersonalizing Growth Models in Medical Images* (Ph.D Thesis).

Ledzewicz U.et al. (2004) Structure of Optimal Controls for a Cancer Chemotherapy Model with PK/PD. 43rd IEEE Conference on Decision and Control.

Levy G. et al. (1966). Kinetics of pharmacologic effects. Clinical Pharmacology Therapeutics.

Lyng H. et al.(2000). Disease Control of Uterine Cervical Cancer: Relationships to Tumor Oxygen Tension, Vascular Density, Cell Density, and Frequency of Mitosis and Apoptosis Measured before Treatment and during Radiotherapy. *Clinical cancer research*.

Mager D.E. et al. (2003). *Diversity of Mechanism-Based Pharmacodynamic Models*. Drug Metabolism and Disposition

Mayneord W. V. et al (1932). *On a law of growth of Jensen's rat sarcoma.* The American Journal of Cancer

Moreira J. et al. (2002). *Cellular Automation Models of Tumor Development: a Critical Review*. Advances in Complex Systems

Murray J. et al.(1994). *Optimal drug regimens in cancer chemotherapy for singledrugs that block progression through the cell cycle*. Mathematical Biosciences.

Ribba B. et al. (2005). *A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies*. Theoretical Biology and Medical Modeling

Stamatakos G et al. (2009). *Introduction of Hypermatrix and Operator Notation into a Discrete Mathematics Simulation Model of Malignant Tumour Response to Therapeutic Schemes In Vivo*. Cancer informatics.

Swanson K. R. et al. (2002). *Quantifying efficacy of chemotherapy of brain tumors with homogeneous and heteregeneous drug delivery*. Acta biotheoretica.

Swanson K. R. (2009). *A mathematical model for brain tumor response to radiation therapy*. Mathematical Biology.

Wagner J.G. et al. (1968). *Kinetics of pharmacologic response: I. Proposed relationships between response and drug concentration in the intact animal and man*. Journal of theoretical biology.