SYSTEM IDENTIFICATION IN TUMOR GROWTH MODELING USING SEMI-EMPIRICAL EIGENFUNCTIONS

Thierry Colin, Angelo Iollo, Damiano Lombardi, Olivier Saut

Mathematical Models and Methods in Applied Sciences
Vol. 22, No. 6 (2012) 1250003 (30 pages)

DOI: 10.1142/S0218202512500030

World Scientific
www.worldscientific.com

SYSTEM IDENTIFICATION IN TUMOR GROWTH MODELING USING SEMI-EMPIRICAL EIGENFUNCTIONS

THIERRY COLIN*, ANGELO IOLLO†, DAMIANO LOMBARDI‡
and OLIVIER SAUT§

Institut de Mathématiques de Bordeaux,
UMR 5251 Université Bordeaux 1
and INRIA Bordeaux-Sud Ouest,
équipe-projet MC2, 351, Cours de la Libération,
Talence 33405, France
**thierry.colin@math.u-bordeaux1.fr*
†*angelo.iollo@math.u-bordeaux1.fr*
‡*damiano.lombardi@math.u-bordeaux1.fr*
§*olivier.saut@math.u-bordeaux1.fr*

Received 15 February 2011
Revised 16 September 2011
Communicated by J. T. Oden

A tumor growth model based on a parametric system of partial differential equations is considered. The system corresponds to a phenomenological description of a multi-species population evolution. A velocity field taking into account the volume increase due to cellular division is introduced and the mechanical closure is provided by a Darcy-type law. The complexity of the biological phenomenon is taken into account through a set of parameters included in the model that need to be calibrated. To this end, a system identification method based on a low-dimensional representation of the solution space is introduced. We solve several idealized identification cases corresponding to typical situations where the information is scarce in time and in terms of observable fields. Finally, applications to actual clinical data are presented.

Keywords: Tumor growth modeling; data assimilation; inverse problems.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

Mathematical modeling of tumor growth can be a useful tool to improve the understanding of cancer treatment in terms, for example, of prognosis, drug effect modeling,[25,29] and clinical protocols definition.[24,26] In the literature, different approaches have been proposed, ranging from individual models, such as cellular automata,[1] agent-based models,[20] continuous models describing the motion of the tumor

†Corresponding author

boundary,[19] models based on reaction-diffusion equations,[16,27] and models that treat the tumor within a mixture theory framework.[2,8,10] All these approaches are phenomenological in the sense that they are not obtained from first principles, trying rather to mimic experimental observations.

Classically the mathematical models used for clinical applications are based on sets of Ordinary Differential Equations (ODEs). These models do not consider the spatial aspect of the tumoral growth, yet they have proved to be of great interest in realistic applications.[12,28] Typically, they are parametrized using statistical methods and may provide a prognosis on the tumoral volume but neither the shape nor the location of the tumor can be inferred. We have chosen to add this spatial aspect to our modeling approach and to consider models based on Partial Differential Equations (PDEs). For this matter, we could have used complex models as in Refs. 5 and 17. These models are well adapted to study qualitatively the interplay between the various biological phenomena involved in tumoral growth. Yet their complexity and a huge number of free parameters make their use in clinical application difficult. Here, a simpler spatial model is considered. This choice is a trade-off between simplicity and accuracy of the phenomenological description but we believe that it is an improvement over ODE models as the present approach may yield much more information than scalar quantities (e.g. the localization, shape or even composition of the tumor). Contrary to an ODE description, our model takes the spatial dimension into account and its simplicity makes the parameter identification possible using a limited set of data.

In the following we will refer to continuous-type models based on mixture theory. In general they rely on a system of nonlinear coupled parametric partial differential equations, in which a set of parameters accounts for the complexity of the different tissues attacked by the tumor as well as for the variability from one individual to another. The tumor growth is investigated at a macroscopic scale and therefore all the microscopic and mesoscopic scale phenomena that we do not model directly are lumped in such parameters.

In order to apply such models in practical situations, these parameters need to be identified, i.e. a realistic value has to be estimated. One way to determine their values is by means of inverse problems, exploiting data coming from medical imagery, as achieved for example in Refs. 13 and 16. The main difficulty is that the amount of data for system identification is scarce. Although medical scans allow a quite accurate localization of the tumor in space, little information can be inferred about its cellular nature or nutrient distribution. In addition, usually only two scans are available before treatment making estimation of time evolution a challenging problem. On the other hand, retrieving the evolution of the tumor shape provides indirect information thanks to the fact that the models are spatially distributed.

The aim of this work is to set up an efficient identification procedure to estimate the parameters of a two-dimensional tumor growth model, with the final objective of obtaining a prognostic model. There is a wide literature of methods to solve inverse

problems concerning diffusion and propagation phenomena. In particular, two main classes of methods were developed: deterministic and stochastic approaches. In the latter a random process is considered and the parameters as well as the variable fields of a given model are inferred once their statistical properties are given,[21] and for tumor growth modeling.[23] Here we opt for a deterministic approach based on an *a priori* sampling of the solution space.

In the cases we will deal with, one of the most challenging problems will not be only to identify the parameters, but also to find fields that are not observable. That in general makes our problems greatly under-determined, but it has a great interest from a medical point of view since it allows one to have information about quantities that will determine the tumor evolution. For example it is clinically meaningful to reconstruct the distribution of the oxygen field in the tissues, or the distribution of proliferating cell density. As a matter of fact, in realistic situations the source of relevant data is medical imagery, so the observations that can be retrieved are indirect, continuous in space but discrete in time.

One possible approach to formulate the inverse problem is by optimal control theory, as was done for instance in Ref. 13. A direct system and an adjoint one have to be solved forward and backward in time, respectively. In general this method is very robust but has a high computational cost: in the case of the simplified model used in the present work it would result in ten coupled nonlinear partial differential equations. With more complicated models in three-space dimensions, this could hardly be feasible.

A different and computationally affordable approach is described in the following. It consists in directly using the residuals of the model (detailed in Sec. 2) within a Newton method to solve the inverse problem. This identification procedure is based on proper orthogonal decomposition (POD[18]) and it is introduced in Sec. 3. Similarly to what we would like to do in real applications where we try to identify complex natural processes with models that are intrinsically much simpler than reality, in Sec. 4 we show that solution fields obtained by a Stokes-type model can actually be identified using a simpler model based on this method. Several applications to biological data are finally provided in Sec. 5. They concern the evolution of lung metastases of a thyroid tumor.

2. Direct Problem: The Darcy-Type 2D Model

In this section we briefly introduce a simplified 2D Darcy-type model describing a two-species saturated flow in a porous isotropic medium. This is a parametric model that is simple and able to take the main physical features of tumor growth into account. In the literature, several complex models have been proposed,[5,6,8,25,31] describing age-structured populations as multi-species saturated flow including the modeling of the cell cycle. Compared to those models, the mathematical description proposed in the following is simpler and as a consequence it disregards certain biological mechanisms. However, here our objective is twofold: to give a reasonable

description of the phenomenon and to define an affordable identification problem. Hence the model proposed is a compromise between these diverging objectives.

We consider the dynamics of two different cellular species, that we will denote by P and Q. The density P represents the proliferating cells (dividing cells, responsible for tumor growth) and Q is the density of necrotic cells that die because of lack of oxygen in the tissue. We make a passive motion assumption, so that the velocity field is equal for every cellular phenotype phase (see Ref. 25 for details). Under this hypothesis, the mass balance equations for P and Q are

$$\frac{\partial P}{\partial t} + \nabla \cdot (\mathbf{v}P) = (2\gamma - 1)P, \tag{2.1}$$

$$\frac{\partial Q}{\partial t} + \nabla \cdot (\mathbf{v}Q) = (1 - \gamma)P, \tag{2.2}$$

where the velocity $\mathbf{v}$ models the tissue movement due to the increase of the tumor volume and γ is the hypoxia threshold, a scalar function of the oxygen concentration that is more precisely defined later in Eq. (2.9). If enough oxygen is available then $\gamma = 1$ and Eq. (2.1) describes the proliferation of tumor cells and the quantity of necrotic cells is constant, thanks to Eq. (2.2). If there is a lack of oxygen, then $\gamma < 1$ and some proliferating cells die and enter the necrotic phase, thanks to Eq. (2.2). The function γ is a purely phenomenological description of a complex biological process, and hence it has to be identified since it cannot be deduced from experiments.

The density of healthy cells is denoted by S and, since their metabolism is not as fast as the metabolism of proliferating cells, the equation for S reduces to a homogeneous transport equation, as explained in Ref. 25:

$$\frac{\partial S}{\partial t} + \nabla \cdot (\mathbf{v}S) = 0. \tag{2.3}$$

We use a hypothesis of saturated flow,[2,6] that is, $P + Q + S = 1$ at every point of the space domain and for every time. Summing up Eqs. (2.1)–(2.3) lead to an equation for the divergence of the velocity field, namely:

$$\nabla \cdot \mathbf{v} = \gamma P. \tag{2.4}$$

We observe that, from a physical point of view, this is equivalent to state that the mitosis acts as volume source for the flow.

From a mechanical point of view it is not sufficient to fix the divergence of the velocity field; in order to close our system we have to assign at least a law for the curl of the velocity. Several kinds of closures have been proposed in the literature, see Refs. 31 and 2. We chose to use a Darcy-type law, that describes quasi-steady flows in porous media, with a variable permeability:

$$\mathbf{v} = -k(P, Q)\nabla \Pi. \tag{2.5}$$

The scalar function Π plays the role of a pressure (or of a potential), and k is a permeability field, that is a function of P and Q. The most simple, phenomenological law is a linear mapping of the sum $(P+Q)$, so that we have:

$$k = k_1 + (k_2 - k_1)(P + Q), \tag{2.6}$$

where k_1 represents the constant permeability of the healthy tissue and k_2 is the permeability of the tumor tissue.

After defining the mechanics of the system, we have to specify the nutrient evolution that in this case reduces to a reaction-diffusion equation for the oxygen concentration. We make the assumption of a quasi-steady state:

$$-\nabla \cdot (D(P,Q)\nabla C) = -\alpha P C - \lambda C, \tag{2.7}$$

where α is the oxygen consumption rate for the proliferating cells, λ is the oxygen consumption coefficient of healthy tissue and $D(P,Q)$ is the diffusivity. Again, the diffusivity can be written as a linear mapping of $P+Q$:

$$D = D_{\max} - K(P + Q). \tag{2.8}$$

This phenomenological law reflects the fact that the diffusion of oxygen is different in the healthy or tumor tissues. The hypoxia function γ simply states that, when the concentration of oxygen is under a certain threshold the cells become necrotic. The definition of γ is a regularization of the unit step:

$$\gamma = \frac{1 + \tanh(R(C - C_{\text{hyp}}))}{2}, \tag{2.9}$$

where R is a coefficient and C_{hyp} is the hypoxia threshold.

According to the physics of the system, reflecting different clinical cases, Dirichlet boundary conditions or Neumann boundary conditions can be imposed for both the oxygen and the pressure fields. For example, in order to mimic the presence of a blood vessel, we impose that the oxygen concentration is constant on a given subdomain. Imposing Neumann conditions on the pressure field is equivalent, from a physical point of view, to imposing that there is no mass leaving our domain. In order to have a well-posed problem the equation for the divergence of the velocity is modified. In particular, the divergence must be a zero average scalar quantity, so that we can write:

$$\nabla \cdot \mathbf{v} = \gamma(C)P - \frac{\int_\Omega \gamma P \, d\Omega}{\int_\Omega 1 - P - Q \, d\Omega}(1 - P - Q). \tag{2.10}$$

From a mechanical point of view this is equivalent to imposing that the growth of the tumor causes a compression of the healthy tissue. Therefore the healthy tissue equation can no longer be considered, in this case, a homogeneous transport equation.

This model can be seen as a radical simplification of that of Ref. 5. For example, the cell cycle described therein is modeled by the coefficient $\gamma(C)$ and the result of the angiogenesis process is summarized in Eq. (2.5). This could be perceived as an

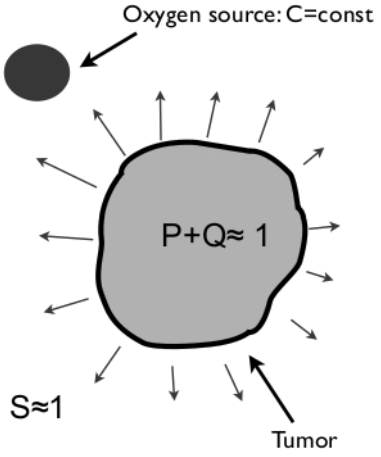


Fig. 1. Schematic computational setup. Boundary conditions are imposed at the square borders for pressure and oxygen.

oversimplification of the model. But, as noticed in the Introduction, our model is still much more complex than those based on ODEs since it involves spatial scales. In view of system identification, this is a trade-off between accuracy and complexity.

A sketch of the computational setup for some of the test cases presented hereafter is given in Fig. 1. The various equations appearing in the mathematical model were discretized on a Cartesian mesh using a finite volume approach. Advection equations are solved numerically with a WENO scheme.[9,14] To improve numerical accuracy, solutions of the diffusion equations are computed with a Ghost-Fluid method.[11]

3. Regularization by Means of Semi-Empirical Eigenfunctions

In clinical practice we do not have access to both P and Q. Rather, what is known to a certain extent is the total tumor cell density field $Y = P + Q$. This decrease of information makes the inverse problem greatly underdetermined. The main idea is then to include additional information using a classical regularization approach. In the literature, regularization techniques in inverse problems were proposed within a stochastic framework,[15,21,32] by making assumptions on the statistics of what is observed.

Here we define a regularization in such a way that we do not change the deterministic structure of the problem: the solution is sought in a given low-dimensional functional space.

3.1. *Regularization based on empirical eigenfunctions*

In the sequel, the observed quantity Y is known at two time instants. The unknowns are the concentration of oxygen C, the velocity field $\mathbf{v}$, the proliferating cells density

P and all the parameters (k_2, $D_{\max}$, K, α, λ, R, C_{hyp}) at a given observation time. The objective is now to write the model as a function of Y whenever possible.

Summing up Eqs. (2.1) and (2.2) we get:

$$\dot{Y} + \nabla \cdot (Y\mathbf{v}) = \gamma(C)P. \tag{3.1}$$

For the velocity field and for the oxygen concentration field, we have:

$$\nabla \cdot \mathbf{v} = \gamma(C)P - \frac{\int_\Omega \gamma P \, d\Omega}{\int_\Omega 1 - Y \, d\Omega}(1 - Y), \tag{3.2}$$

where the expression relative to Neumann boundary condition for the pressure field was retained; in the case of Dirichlet boundary conditions the second term of the right-hand side of this equation simply vanishes.

We take the curl of the Darcy law:

$$k(Y)\nabla \wedge \mathbf{v} = \nabla k(Y) \wedge \mathbf{v}, \tag{3.3}$$

and add the equation for the oxygen concentration field:

$$\nabla \cdot (D(Y)\nabla C) = \alpha PC + \lambda C. \tag{3.4}$$

The definition of the hypoxia function, γ, is unchanged.

The basis of the solution functional space is constructed using POD. This approach was already used in population dynamics to set up reduced-order model[34] and it is widely used in fluid mechanics to obtain reduced-order models for optimization and control.[4,33] This basis, given the space dimension, maximizes the energy representation of a previously computed solution set. In other words, the solution of the inverse problem is sought in the space spanned by a basis that gives an optimal representation of a sufficiently large number of solution samples. For each of the following tests, we have built a parametric space in which all the parameters vary in such a way that a large set of different solutions are included in the database.

In particular, the parameter space is sampled varying the permeability constants ratio $\chi = k_2/k_1$, the diffusivity ratio $\eta = K/D_{\max}$, and the oxygen consumptions α and λ. The details on the database construction are provided below. For each solution we saved enough time snapshots to accurately resolve all the relevant frequencies. In principle the database should be such that all the possible different biological behaviors are represented.

The Sirovich method[30] is used to build the low-dimensional space. Each POD mode for a given variable is written as a linear combination of snapshots of that variable. The maximum energy norm problem results in an eigenvalue problem for the time auto-correlation matrix. We describe the method for P, pointing out that for all the other variables, i.e. $\mathbf{v}$, C and γP, it is exactly the same. The time auto-correlation matrices are defined in the following manner:

$$A_{ij}^{(P)} = \langle P_k(t_h), P_m(t_n) \rangle \quad h, n = 1, \ldots, N; \quad k, m = 1, \ldots, M, \tag{3.5}$$

where N is the number of snapshots for each simulation, M is the total number of simulations: as a consequence $i = (k-1)N+h$ and $j = (m-1)N+n$ and $\langle \cdot, \cdot \rangle$ is the discrete L^2 scalar product on the space domain. The eigenvalues and eigenvectors of the above matrix are computed: let $\mathbf{b}^i$ denote the ith eigenvector and λ_i the corresponding eigenvalue. It can be shown, see Ref. 30, that the ith eigenmode of the space auto-correlation matrix

$$A_{hn}^{(P)} = \sum_k P_k(x_h) P_k(x_n) \quad h, n = \text{number of grid points}; \quad k = 1, \ldots, M \tag{3.6}$$

can be written as linear combination of snapshots:

$$\phi_i = \frac{\sum_j b_j^i P_j}{\lambda_i^{1/2}}, \tag{3.7}$$

where we have called $P_j = P_m(t_n)$, $j = (m-1)N + n$, and b_j^i is the jth component of the ith eigenvector. The set of eigenfunctions ϕ_i is orthonormal.

The dimension of the empirical functional space, i.e. the number of POD modes we use to reconstruct the solution, is chosen such that if additional POD modes are included, the reconstruction of a given field does not vary up to a certain error value that, in this work, was fixed at 10^{-4} in L^2 norm.

The repeated index summation convention is used from now on. We write the variables $P, C, \mathbf{v}, \gamma P$ as eigenmode expansions:

$$P = a_i^{(P)} \phi_i^{(P)}, \quad i = 1, \ldots, N_P, \tag{3.8}$$

$$C = a_i^{(C)} \phi_i^{(C)}, \quad i = 1, \ldots, N_C, \tag{3.9}$$

$$\mathbf{v} = a_i^{(v)} \phi_i^{(v)}, \quad i = 1, \ldots, N_v, \tag{3.10}$$

$$\gamma P = a_i^{(\gamma P)} \phi_i^{(\gamma P)}, \quad i = 1, \ldots, N_{\gamma P}, \tag{3.11}$$

where $a_i^{(\cdot)} = a_i^{(\cdot)}(t)$ are scalar functions of time, $\phi_i^{(\cdot)} = \phi_i^{(\cdot)}(\mathbf{x})$ are functions of spatial coordinates. Substituting these expressions in the system Eqs. (3.1) and (3.4) we obtain:

$$\dot{Y} + a_i^{(v)} \nabla \cdot (Y \phi_i^{(v)}) = a_i^{(\gamma P)} \phi_i^{(\gamma P)}, \tag{3.12}$$

$$a_i^{(v)} \nabla \cdot \phi_i^{(v)} = a_i^{(\gamma P)} \phi_i^{(\gamma P)} - \frac{\int_\Omega a_i^{(\gamma P)} \phi_i^{(\gamma P)} \, d\Omega}{\int_\Omega 1 - Y \, d\Omega} (1 - Y), \tag{3.13}$$

$$a_i^{(v)} k(Y) \nabla \wedge \boldsymbol{\phi}_{\mathbf{i}}{}^{(v)} = a_i^v \nabla k(Y) \wedge \phi_i^{(v)}, \tag{3.14}$$

$$a_i^{(C)} \nabla \cdot (D(Y) \nabla \phi_i^{(C)}) = \alpha a_j^{(P)} a_i^{(C)} \phi_j^{(P)} \phi_i^{(C)} + \lambda a_i^{(C)} \phi_i^{(C)}. \tag{3.15}$$

The hypoxia function γ, Eq. (2.9), is multiplied by P, in such a way that the product γP is:

$$a_i^{(\gamma P)} \phi_i^{(\gamma P)} = a_j^{(P)} \phi_j^{(P)} \frac{1 + \tanh(R(a_i^{(C)} \phi_i^{(C)} - C_{\text{hyp}}))}{2}. \tag{3.16}$$

We solve system (3.12)–(3.15) by minimization of the residuals under certain constraints that are introduced below. The first constraint is linked to the fact that Eq. (3.15) is a homogeneous equation with respect to the coefficients $a_i^{(C)}$. As a direct consequence, one possible solution of the oxygen diffusion equation is the trivial one. Indeed, if $C_{\text{hyp}} < 0$ such a solution would also be a solution for the whole system, Eqs. (3.12) and (3.16). In order to prevent the identification of a system with unphysical solutions we can proceed in two different ways. We can discretize the boundary conditions for oxygen, getting one scalar constraint, exactly as in the Petrov–Galerkin method. In the case of Dirichlet boundary conditions $C = C_0$ on $\partial\Omega_C$ where Ω_C is a blood vessel domain, we obtain one scalar equation of the form:

$$\sum_i \left(\frac{\sum_j b_j^i}{\lambda_i^{1/2}} \right) a_i^{(C)}(t) = 1, \quad \forall\, t. \tag{3.17}$$

Another option is to lift the solution and transform Eq. (3.15) into a nonhomogeneous equation, with source terms. Both these approaches yield similar results in terms of inverse problem solution and hence in this work we simply lift the solution.

At a given time (say t_0), the snapshot $Y(t_0)$ and a subsequent snapshot $Y(t_1)$ are used to perform the computation of the time derivative. Let the residue of the lth equation be R_l. We write $F = \sum_l R_l^2$ and

$$(a_i^{(\cdot)}(t_0), \pi_j) = \text{argmin}(F), \tag{3.18}$$

where $a_i^{(\cdot)}$ are the expansion coefficients for the variables $P, C, \mathbf{v}, \gamma P$ and π_j are the parameters to be identified.

The second constraint to be imposed in the minimization results from the observation that, since in the inverse problem we do not solve the equation for the variable P, the latter does not automatically satisfy: $0 \leq P \leq 1$ and therefore this is a constraint we have to impose. In order to impose this constraint, that is fundamental for the point of view of the population dynamics, the residuals are penalized as follows:

$$\tilde{F} = F + c_1(\max\{a_i^{(P)}\phi_i^{(P)}\} - 1) + c_2(-\min\{a_i^{(P)}\phi_i^{(P)}\}), \tag{3.19}$$

where c_1, c_2 are positive numbers, that we have set in such a way that penalization does not affect the stability of the procedure. The inverse problem finally takes the form of a nonlinear algebraic optimization problem, that is solved using a Newton trust region method.

In order to decrease the computational cost of the procedure, a third constraint is imposed to define a feasible set of solutions. In particular, let $a_i^{(P)}(t_0)$ be the ith POD coefficient relative to the variable P evaluated at the time t_0. The maximum and the minimum values that the coefficient $a_i^{(P)}(t)$ reaches in the database

simulations can be calculated on the basis of the auto-correlation matrix of the variable. Indeed, the definition of the ith POD mode implies that:

$$\phi_i^{(P)} = \frac{b_j^i P_j}{\lambda_i^{1/2}} \Rightarrow a_{ik}^{(P)} = \langle P_k, \phi_i^{(P)} \rangle = \frac{1}{\lambda_i^{1/2}} \langle P_k, b_j^i P_j \rangle, \tag{3.20}$$

where $a_{ik}^{(P)}$ denotes the projection of the kth snapshot of P on the ith eigenvector.

From the definition of the eigenvector we finally obtain:

$$a_{ik}^{(P)} = \frac{1}{\lambda_i^{1/2}} \langle P_k, P_j \rangle b_j^i = \lambda_i^{1/2} b_k^i \Rightarrow \max_k \{a_{ik}^{(P)}\} = \lambda_i^{1/2} \max_k \{b_k^i\}. \tag{3.21}$$

We can conclude the same for the minimum of the coefficient. Thanks to this relationship we can estimate the interval of excursion I_k^{db} of the projection coefficients

$$I_k^{db} = [\min\{a_{ik}^{(P)}\}, \max\{a_{ik}^{(P)}\}]. \tag{3.22}$$

We ask the solution to be not too different from the simulations of the direct problem, since we assume that we built a database in which the biological behaviors of the variables were represented. Thus we restrict the admissible values of the POD coefficients to an interval I_k that is obtained from I_k^{db} by a stretching factor $1 + \delta$ where δ is a suitable positive number. In all the following simulations the value $\delta = 0.1$ was adopted. It should be noted that this choice still allows the procedure to identify solutions that are very different with respect to the solution of the database. This procedure is repeated for all the variables included in the database.

3.2. *Time interpolation*

The hypothesis that two subsequent snapshots are close in time, or, in other words, that the time between two snapshots is small if it is compared with the characteristic evolution time of the phenomenon, is very optimistic. In order to relax this hypothesis, instead of using first-order finite differences, that is equivalent to perform a linear interpolation between snapshots, a different kind of interpolation can be used. However, a higher-order finite difference scheme, equivalent to a polynomial interpolation between the snapshots, would require a large number of snapshots. As an alternative, still assuming that only two images are available, an additional hypothesis about the growth rate could be retained. Here, we consider two cases. In the case of exponential growth we write:

$$\dot{Y} \approx A \exp\{\zeta t\} + B \exp\{-\zeta t\} = f(\zeta), \tag{3.23}$$

where A, B are chosen in such a way that the two available snapshots are interpolated. One parameter, ζ, is free and enters the residual minimization process. The first equation of the system (3.1)–(3.4) becomes:

$$f(\zeta) + \nabla \cdot (a_i^{(v)} \phi_i^{(v)} Y) = a_i^{(\gamma P)} \phi_i^{(\gamma P)}. \tag{3.24}$$

In the case of a logistic-type growth we proceed in a similar way. We take

$$Y \approx AG(\omega, \sigma) + BG(-\omega, -\sigma), \tag{3.25}$$

where

$$G(\omega, \sigma) = \frac{\omega e^{\omega t}}{\omega - \sigma e^{\omega t}}. \tag{3.26}$$

As before A and B are adjusted such that the snapshots are interpolated. In this case, however, we are left with two free parameters (ω and σ) that are found within the residual minimization process.

4. Identification of a Stokes-Type Flow Using 2D Darcy-Type Model

The objective of this section is to give a quantitative *a posteriori* error analysis of the identification procedure. The framework is as similar as possible to an actual clinical situation. In this sense, we consider the identification of a Stokes-type flow using a Darcy-type model, in two different physical situations. This is a more realistic situation compared to what one would like to do in clinical practice. The equations describing a Stokes flow are the following ones:

$$\nabla \cdot \mathbf{v} = 0, \tag{4.1}$$

$$-\nabla \cdot \sigma + \nabla p = 0, \tag{4.2}$$

$$\sigma = \nu(\nabla \mathbf{v} + \nabla^T \mathbf{v}), \tag{4.3}$$

where ν is the kinematic viscosity, σ is the stress–tensor, p is the pressure field, and $\mathbf{v}$ the velocity field.

For the direct simulations the following parameters were used: $\nu = 2.0 - (P + Q)$, so that kinematic viscosity is a non-uniform isotropic field depending on the tissue, $\alpha = 2.0$, $K = 0.8$, $\lambda = 0.01$.

In order to perform the identification, we use two subsequent snapshots taken from the numerical simulations of a Stokes-type flow. The Stokes-type model and its discretization are fully described in Ref. 7. These snapshots are considered as if they were part of the evolution of a Darcy-type flow. In order to solve the inverse problem we follow exactly the procedure described in the previous section. To this end we build a database of solutions of the Darcy-type flow, starting with the same initial conditions of the Stokes-type flow. We construct the eigenfunction basis, we regularize the inverse problem and we solve it by minimizing the residuals.

We point out that in the Stokes-type flow the cellular species obey the same population dynamics as in the Darcy-type model. What changes is the mechanical closure. As pointed out in Ref. 3 a viscoelastic flow in a two-dimensional limit can be well-described by a Darcy-type law with a suitable definition of permeability. Thus we expect that the 2D simplified Darcy model represents in a satisfactory manner

all the main physical features of the more complex Stokes-type flow. Nevertheless we point out that there are also differences between the models (mainly the role of diffusivity) so that this can be considered a good test approaching realistic system identifications. Let us note that in this case we try to perform our regularization using the eigenfunctions that have been extracted from the simple model. This is realistic since it is what we can actually do in real applications.

As a preliminary test we investigate how the density of proliferating cells P determined using a Stokes mechanical closure is approximated by empirical eigenmodes obtained for a Darcy-type flow with the same initial conditions. The initial conditions and the computational setup is identical to that of the next section. We consider the L^2 relative projection error, see Table 1. This table shows that the relative projection error is acceptable and that it decreases with the number of modes and increases with time, as expected.

Next, we present the results of two different inverse problems, in which both the behavior of the oxygen concentration and the mechanical behavior of the healthy tissue vary. In the first case the oxygen concentration on the boundary of the blood vessel is given, whereas in the second test case we assume that the oxygen is provided through the boundary of the computational domain and the value of the oxygen concentration on this boundary has to be identified. In the second test case we also impose a geometrical constraint corresponding to the fact that the tumor cannot leave the computational domain. This is done by modifying the boundary condition on the velocity, which in turn affects the dilatation rate and therefore the mechanics of all the tissues, including the healthy one. This is a model for the tumor growth inside an organ before the metastatic process.

4.1. *Database and POD representation*

In the numerical examples presented in this section the database was built as follows: 576 simulations were completed, taking 20 snapshots of each. The parameters used were:

- $\chi = k_2/k_1$: $[0.50, 1.0, 1.5, 2.0]$, $k_1 = 1$ for all the simulations,
- $\eta = K/D_{\max}$: $[0.10, 0.45, 0.80]$, $D_{\max} = 2$ for all the simulations,
- α: $[0.1, 1.0, 3.0, 5.0]$,

Table 1. Relative projection error as a function of the number of POD modes N_P and of the time instant considered T.

N_P	$T = 0$	$T = 5$	$T = 10$	$T = 15$
5	5.18e-2	9.64e-2	15.10e-2	15.68e-2
10	3.85e-2	5.12e-2	5.81e-2	9.13e-2
15	2.64e-2	3.62e-2	3.12e-2	4.45e-2
20	1.53e-2	2.38e-2	2.42e-2	3.51e-2
25	1.12e-2	1.44e-2	1.91e-2	2.52e-2

- C_{hyp}: $[0.05, 0.045, 0.085, 0.125]$, $\max_{\Omega}\{C\} = 0.15$,
- λ: $[0.001, 0.1, 0.5]$.

For the Darcy-type model the unknowns are P, Y, C, γ, Π (or $\mathbf{v}$). There are two possibilities to build the regularization basis: the first one consists in extracting a unique vector basis, taking all the variables as components of the same vector field into account. The second one consists in looking for a separated expansion for each variable. This second option was chosen as the involved variables have different meanings and they are inhomogeneous from a physical standpoint. We show the first POD modes of C and P for Case I presented in the next section. In Fig. 2(a), the first POD mode is shown, representing the average of the oxygen fields. The blood vessel is recognized. In the other modes the presence of the tumor

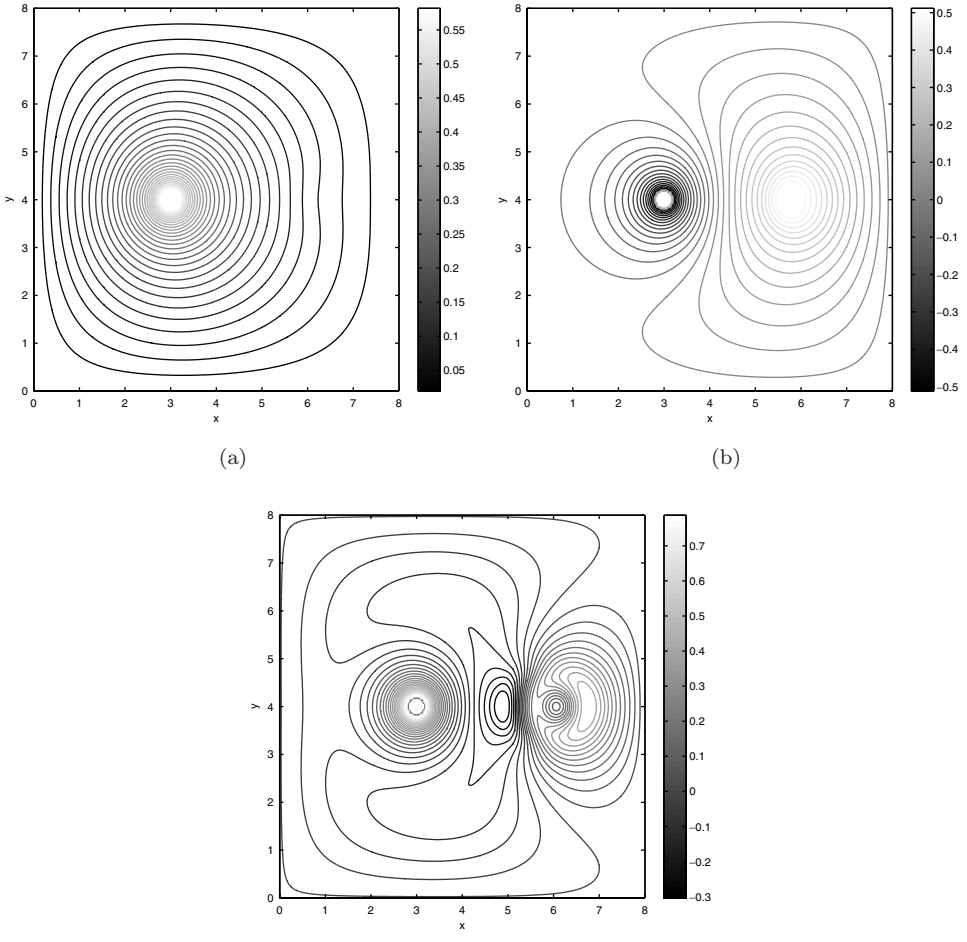


Fig. 2. POD modes for the oxygen field: (a) first, (b) second and (c) third mode.

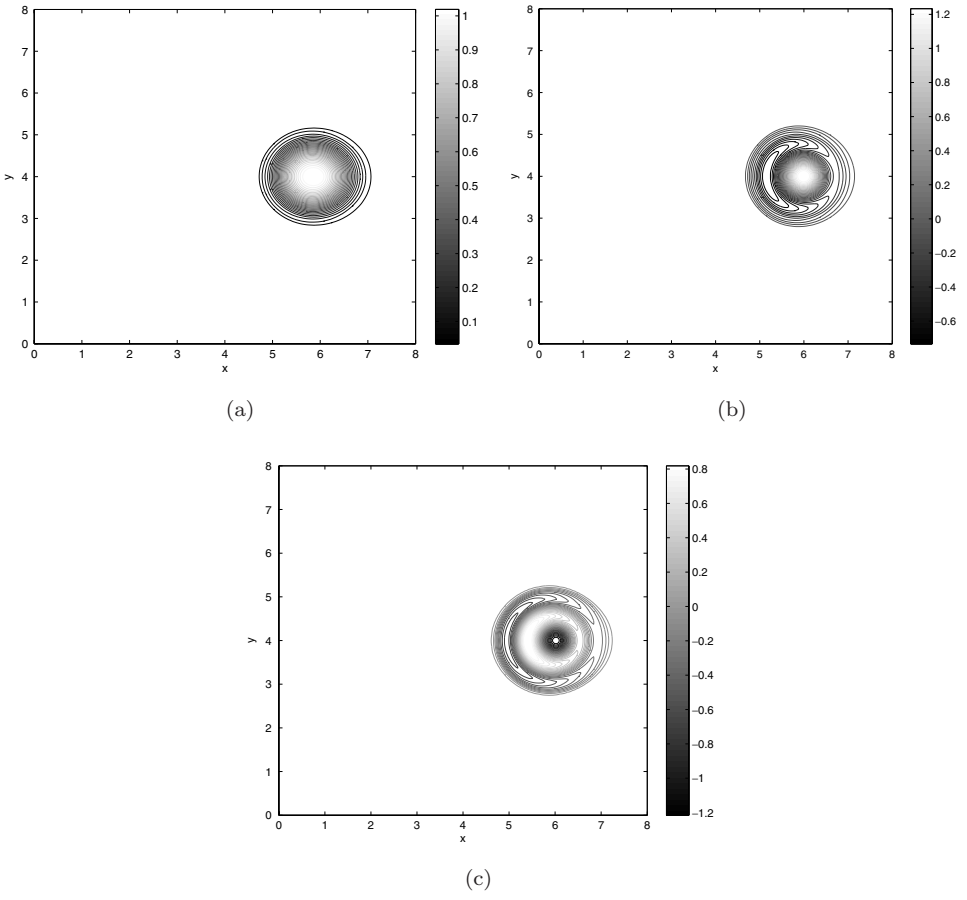
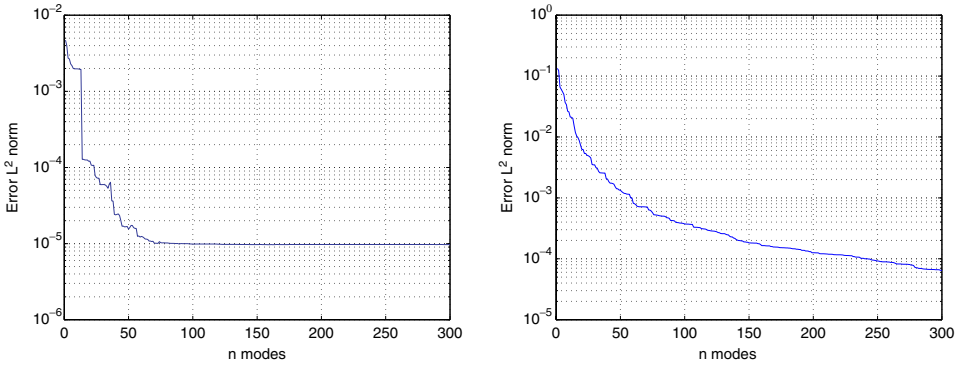


Fig. 3. POD modes for the proliferating cells density: (a) first, (b) second and (c) third mode.

becomes clearer as it induces variations due to oxygen consumption. In Fig. 3, the POD modes for the proliferating cell density are shown. They are compactly supported. P is moving toward the blood vessel and the movement is rendered by the wavy structure highlighted by the second and third modes. The quantitative properties of representation of the modes have been investigated. In Fig. 4, the L^2 representation error for the oxygen (a) and for proliferating cells density (b) is represented as function of the number of POD modes, when the setting presented in Case I, detailed below, is adopted. This error is relative to a growth simulation that does not belong to the database (initialized with random parameters). The error for oxygen decreases faster. This is due to the fact that oxygen is governed by a diffusion equation, while proliferating cells evolve driven by a transport and their support evolves during the simulation. More POD modes are needed to take this non-global behavior for proliferating cells into account.



(a) (b)

Fig. 4. L^2 representation error with respect to the number of used modes for (a) oxygen and (b) proliferating cells density; the database for the Darcy model was used.

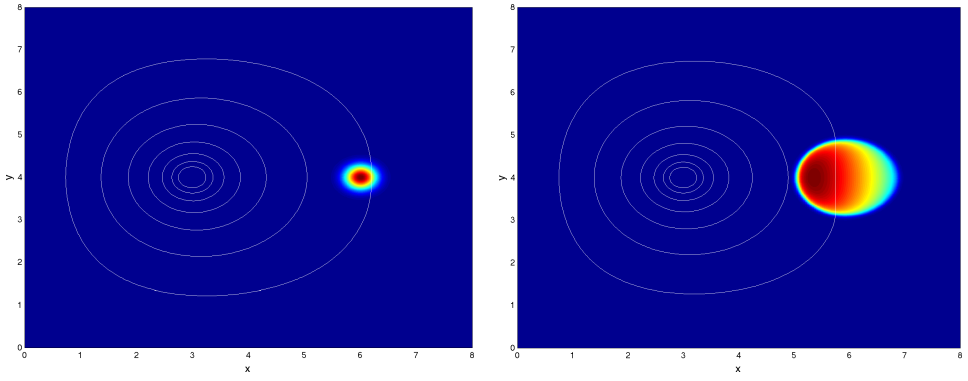
4.2. *Case* I : *Distant blood vessel*

The tumor is a spheroid that starts growing alimented by a single blood vessel of known position and source intensity. In particular, the computational domain is the box $L_x = 8$, $L_y = 8$, the tumor is initially located at $x = 6, y = 4$, the blood vessel at $x = 3, y = 4$.

Dirichlet boundary conditions for both the oxygen and the pressure fields are imposed. In particular, we have:

$$\Pi \text{ and } C = 0 \quad \text{on } \partial\Omega, \quad C = C_0 \quad \text{on } \partial\Omega_C, \tag{4.4}$$

where Ω_C represents the blood vessel and $C_0 = 0.15$. In Fig. 5, two snapshots of the solution of the Stokes-type flow for this case are represented. In



(a) (b)

Fig. 5. (Color online) Solution of the Stokes-type flow in the case of a distant blood vessel, at: (a) $T = 0$; (b) $T = 20$; isolines represent oxygen concentration in the tissue. Color scale goes from 0 (dark) to 1 (clear).

Fig. 5(a) the initial condition is plotted: at the right-hand side the tumor (the proliferating cell density is represented) is initialized as a spheroid with an exponential distribution of proliferating cell density. The contour lines represent the isosurfaces of the oxygen concentration in the tissue. The same quantities are represented in Fig. 5(b) for a subsequent time. It can be seen that the tumor has grown, it has started moving towards the blood vessel, and that due to the oxygen consumption the tumor changes the oxygen distribution in the tissue.

In this case the system of Eqs. (3.1) and (3.4) reduces to:

$$\dot{Y} + \nabla \cdot (a_i^v \phi_i^v Y) = a_i^{\gamma P} \phi_i^{\gamma P}, \tag{4.5}$$

$$a_i^{\gamma P} = \frac{1 + \tanh(R(a_i^c \phi_i^c - C_{\text{hyp}}))}{2}, \tag{4.6}$$

$$a_i^v \nabla \cdot \phi_i^v = a_i^{\gamma P} \phi_i^{\gamma P}, \tag{4.7}$$

$$k(Y) \nabla \wedge a_i^v \phi_i^v = \nabla k \wedge a_i^v \phi_i^v, \tag{4.8}$$

$$\nabla \cdot (D(Y) a_i^c \nabla \phi_i^c) = \alpha a_j^P a_i^C \phi_j^P \phi_i^C + \lambda a_i^C \phi_i^C. \tag{4.9}$$

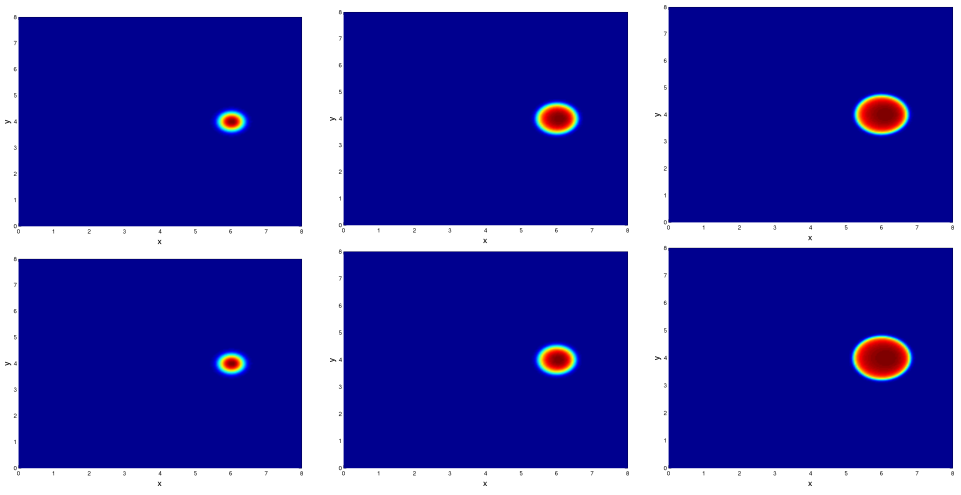
Constraints to the oxygen field are imposed in order to prevent unphysical solutions to arise, as explained in the previous section. Linear interpolation, exponential interpolation and logistic interpolation are used to obtain an estimate of $\dot{Y}$. According to the sensitivity analysis on the representation properties of POD modes, we used the following number of POD modes: for the variable P, $N_P = 10$, for C, $N_C = 5$, for $\mathbf{v}$, $N_v = 30$ and for γP, $N_{\gamma P} = 10$. We present the main results of our numerical tests.

In this section the numerical results of the procedure are discussed. Before analyzing in detail the errors from a quantitative point of view we present briefly the qualitative behavior of the solution obtained by simulating with a Darcy-type flow the same tumor we have simulated using a Stokes-type model, when the parameters of the Darcy model are found by system identification. In Fig. 6, we present three snapshots taken from the two simulations. We performed the identification at $T = 5$ using the second snapshot at $T = 10$ to approximate the time derivative with an exponential-type interpolation. The qualitative behavior of the reconstruction is very similar.

Since the medical data usually involves only the sum of $P + Q$, we choose to concentrate the investigation on the tumor volume error and the tumor center of mass error. In particular, we compute the volume as:

$$V_Y(t) = \int_\Omega \mathbf{1}_\mathbf{Y} \, d\Omega, \tag{4.10}$$

where $\mathbf{1}_\mathbf{Y}$ is the indicatrix of the variable Y, and so that the volume is a measure of the support of that variable. In an equivalent manner we consider the position



(a) (b) (c)

Fig. 6. (Color online) Cell density Y for the Stokes flow (upper row) and for the Darcy flow (below): (a) $T = 5$, (b) $T = 10$, (c) $T = 15$; Color scale ranges from 0 (blue) to 1 (red).

of the mass center as being:

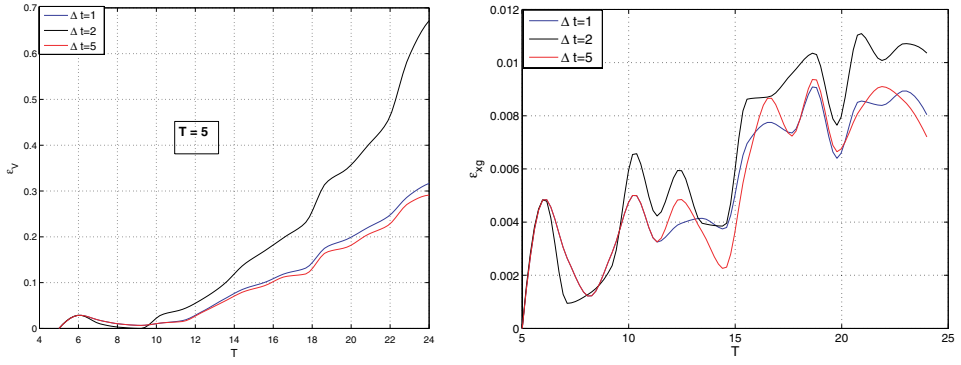
$$\mathbf{x_G} = \frac{\int_\Omega \mathbf{1_Y x}\, d\Omega}{\int_\Omega \mathbf{1_Y}\, d\Omega}. \tag{4.11}$$

The procedure to evaluate the error is the following: two snapshots of the Stokes-type flow are taken and the identification using a Darcy-type flow is performed. We simulate the Darcy flow system using the parameters we have identified and taking as initial condition the first snapshot of the Stokes-type flow. The volume and the center-of-mass position are evaluated. If we denote as $V_Y^{(e)}$ and $x_G^{(e)}$ the volume and the center-of-mass position of the tumor in the Stokes-type flow, we can define the relative errors as follows:

$$\epsilon_V(t) = \frac{V_Y(t) - V_Y^{(e)}(t)}{V_Y^{(e)}(t)}, \tag{4.12}$$

$$\epsilon_X(t) = \frac{(X_G(t) - X_G^{(e)}(t))^2}{(V_Y^{(e)}(t))^{(1/2)}}. \tag{4.13}$$

The system is identified at different times, varying the time interval between the snapshots. In Fig. 7(a), the relative error in volume is plotted. The reference snapshot is taken at $T = 5$ and a linear time interpolation is used. The errors are computed varying the time at which the second snapshot is taken. The error remains below the value of 0.1 for a large part of the time history of the simulation of the Darcy-type flow. All the identification are practically equivalent.

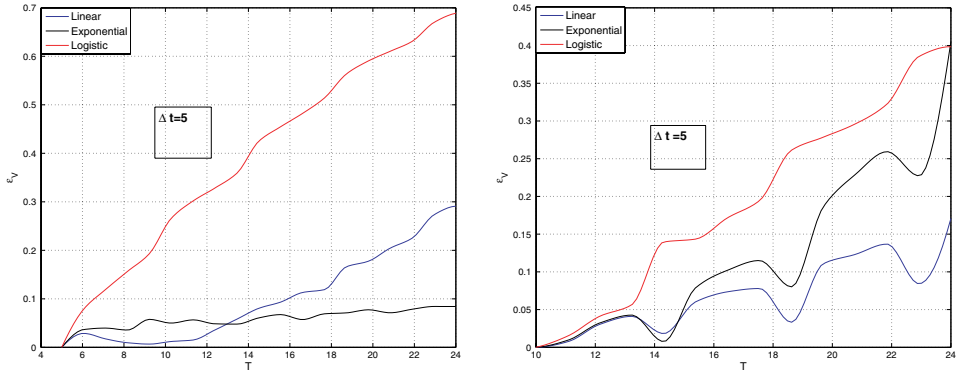


(a) (b)

Fig. 7. Relative error when identification time is $T = 5$, with linear interpolation: (a) volume error and (b) center-of-mass position error.

In Fig. 7(b), we show the error on the center-of-mass position scaled with the square root of the volume. We adopt this normalization in order to have a dimensionless error. Indeed it is reasonable to assume that the higher the dimension of the tumor the higher could be the absolute error in the center-of-mass position. In other words, the relative error obtained represents the absolute error of the center-of-mass position per unit length of the mean tumor radius. All the simulations are equivalent and the error is particularly low, so that we can conclude that the transport approximation is rather good. We point out that the time scales considered are rather realistic, in terms of the increase of the tumor size.

In Fig. 8(a), we show the relative volume error with $\Delta t = 5$ and for different time interpolations. We see that the logistic-type interpolation is less accurate. At $T = 10$, the linear interpolation is better than the exponential one. In contrast,



(a) (b)

Fig. 8. Relative volume error when $\Delta t = 5$, varying interpolations: (a) $T = 5$ and (b) $T = 10$.

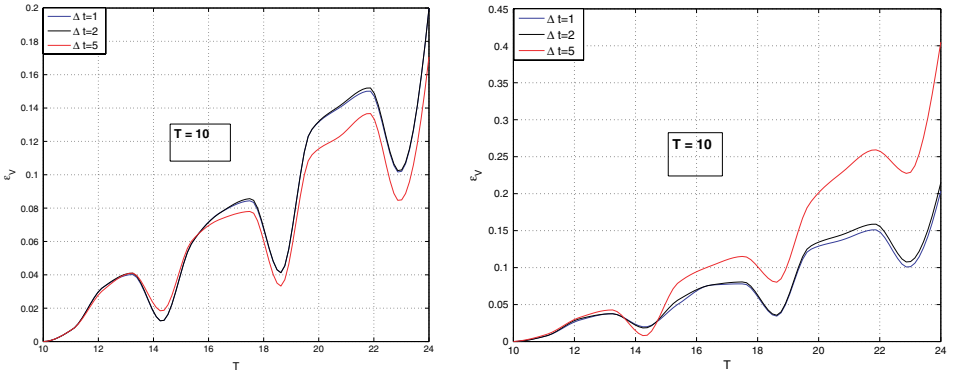


Fig. 9. Relative volume error when $T = 10$, varying Δt with: (a) linear interpolation and (b) exponential interpolation.

when $T = 5$, the exponential interpolation provides a rather good approximation of the time derivative, so that the error remains under 0.1 for all the rest of the simulation.

Let us analyze the error when $T = 10$ for both interpolations, when we vary the time at which we take the second snapshot. We can see in Fig. 9 that the three types of interpolation yield similar results, although the exponential interpolation performs slightly better for $\Delta t = 2$. When we perform the identification at $T = 10$, the error is small for a large part of the growth history, but in general tends to grow faster with respect to the error we make when we identify our system at $T = 5$. This can be due to the fact that the tumor has started moving towards the blood vessel so that the derivative can no longer be well approximated by the derivative of an exponential-type system.

When the Darcy-type flow is integrated, a snapshot taken from the Stokes-type flow is used as an initial condition: this is equivalent in practical simulations to assume that the region corresponding to the active part of the tumor is known. Therefore, the errors we previously computed are solely due to the parameter reconstruction and the differences between the models.

Since in the identification only the sum $P + Q$ is known at two given times, it can be useful to reconstruct also the initial condition for the variables $P, C, \mathbf{v}$ and hence analyze the results when we use, as initial condition, the identified initial condition. Indeed, this reflects the situations occurring in reality. Moreover, we get estimates about the composition of a tumor in terms of phenotype phases as this turns out to be valuable clinical information.

When the identified initial condition is used as a starting point for the simulation, the results worsen and there is a certain sensitivity to the first guess of the Newton solver. Since the inverse problem is underdetermined and quite ill-conditioned, plenty of local minima exist. One way to circumvent this problem is

to take a snapshot of Y from the database as the initial guess. We have obtained in this way relative errors in volume of 25–30%.

4.3. *Case* II : *Irregular tumor shape*

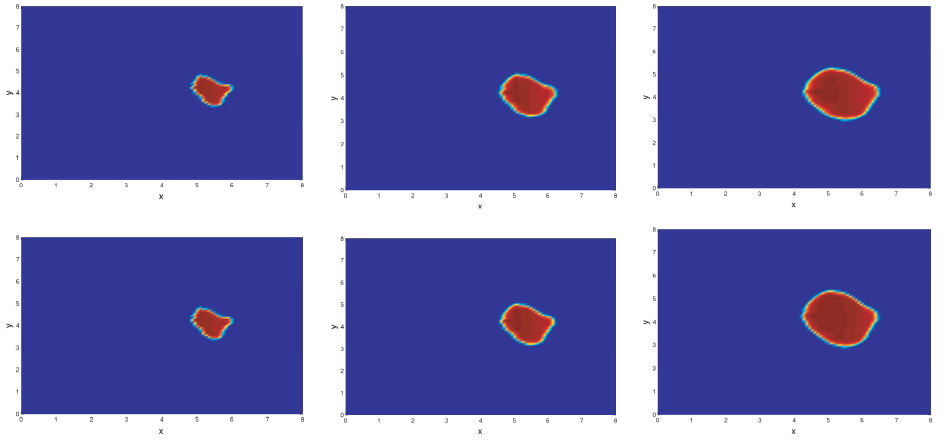
In this subsection the dynamics of an irregular-shaped tumor is identified. The overall problem setting is the same as before. The aim is to check if the procedure is accurate enough to identify complex geometries. Spatial accuracy is in fact the main motivation for using a distributed model based on PDE.

In Fig. 10, the solution of a Stokes flow (upper row) is compared with the solution of the Darcy flow when the identification procedure has been applied with an exponential type of interpolation. The initial tumor distribution was taken from a scan image of a lung cancer (courtesy of Sarrut, Centre Léon Bérard, Lyon, France). The identification is able to take into account the effects of complex geometries, even if the models are quite different. In particular, in the Stokes-type flow we can expect that the tumor boundary becomes less irregular as time increases, since there is a diffusion process, while in the Darcy-type flow there is no diffusivity.

As before, the error analysis is performed, and the same qualitative and quantitative behavior is found with respect to the case of regular shapes. In particular, the relative error stays under the value of 0.1 for a relevant portion of the growth history, corresponding to the doubling of the tumor volume.

5. Applications to Biological Data

In collaboration with Institut Bergonié (the cancerology institute of Bordeaux), we study thyroid cancer metastases in lungs. These metastases are a therapeutic



(a) (b) (c)

Fig. 10. Cell density Y for the Stokes flow (upper row) and for the Darcy flow (below): (a) $T = 5$, (b) $T = 10$ and (c) $T = 15$.

challenge because some are fast evolving and are good candidates for trials with molecular targeted therapies, but others are slowly evolving and it is difficult to decide when to treat them. For the oncologists the development of predictive tools could be of interest in the planning and in the evaluation of an anti-tumoral treatment. For example with slowly evolving tumors, a prediction of growth could reinforce the decision of waiting without specific treatment or on the contrary to help in the decision of starting radiofrequency thermal ablation or molecular targeted therapy.

Several scans may hence be available for a slowly evolving nodule and this occurrence is important in our case since it allows a validation of the assimilation technique on a long time horizon. We make the choice of keeping a two-dimensional setting. The main reason for that is computational feasibility, in the sense that computationally intensive identification procedures in clinical practice seem out of reach for the moment. Another motivation for this choice is that for practical reasons physicians tend to interpret the scans on the largest section of the lesion, even though three-dimensional data are available.

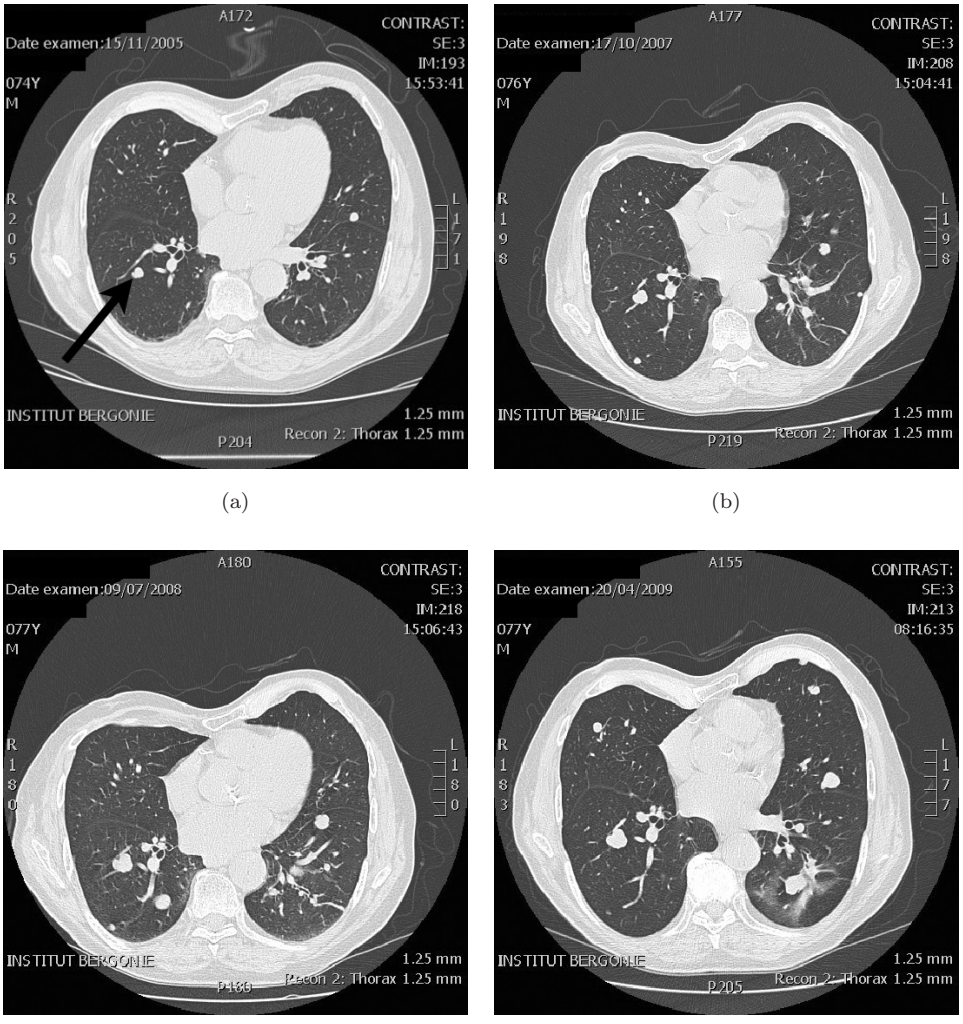
The intervals of the parameters sampled for the database construction is the same as in the synthetic examples detailed in the previous section. The number of simulations may vary, and it will be specified later on.

5.1. *Case* I : *Slow rate growth*

In Fig. 11 four scans covering an evolution over 45 months are presented. Even though this patient is affected by several metastases, the evolution of the one marked in Fig. 11(a) will be studied here. Using only the first two scans, we recover the parameters and the initial conditions that allow us to perform a forward simulation beyond the time corresponding to the second scan. Therefore, starting from the scan corresponding to October 2007, the growth rates obtained are actual model predictions.

In order to determine the POD modes, two databases were constructed by integrating in time the Darcy-type model. The first database consists of 128-parameter configurations that result in growth rates of the order of the ones observed between the first and the second scans on a conventional time scale of 1. For each of the 128 configurations, 20 time snapshots were recorded. 15 POD modes were used for $\mathbf{v}$, P, and γP, 10 modes for C. In order to check the stability of the identification with respect to the solution space sampling, a second database of 768 simulations was also built. Again 20 time snapshots per simulation were considered. The results shown in the following do not significantly vary as a function of the database used.

Let us show the POD modes extracted from the database of 128 simulations for the concentration of proliferating cells and the oxygen distribution in the tissue. In Fig. 12 some eigenmodes of the oxygen field are represented. The modes are regular and the structures in the oxygen field due to the consumption of the tumor may be recognized. In Fig. 13, the proliferating cells density modes are represented. In



(a) (b)

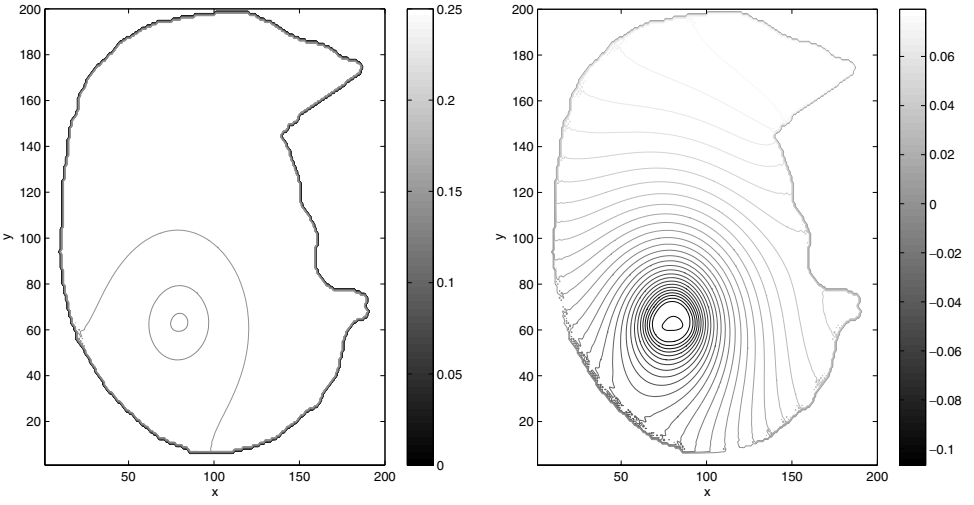
(c) (d)

Fig. 11. Scans: (a) November 2005, (b) October 2007, (c) July 2008 and (d) April 2009.

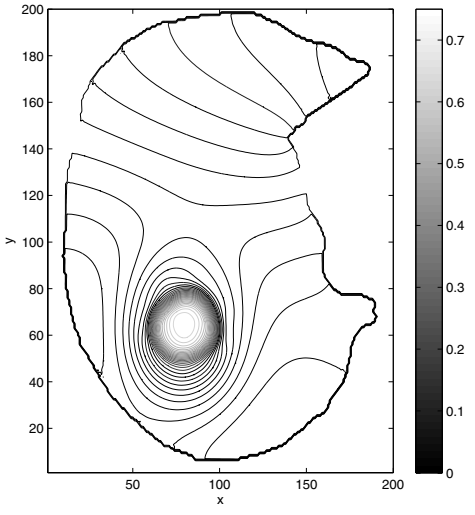
all the modes the shape of the initial tumor may be recognized and a wave-kind of behavior appears to render the growth and the tumor invasion of the surrounding tissue.

Initially the proportion of proliferating cells is fixed to $P = 1$ on the tumor support, that is, at the beginning the tumor is totally proliferating. This value is of course not always realistic, but the results of the identification proved to be weakly-dependent on this assumption.

In Fig. 14 we present the simulated nodule growth compared to the actual nodule size resulting from the scan of April 2009. The support of the Y distribution has



(a) (b)



(c)

Fig. 12. POD modes for the oxygen field, Case I: (a) first mode, (b) third mode and (c) fifth mode.

approximately the same area as the real tumor. However, in the simulation the nodule is more isotropic than in reality. Indeed in Fig. 14(c) a zoom of the space error is shown: the error is localized on the tumor boundary and concerns essentially the shape.

In order to give a more quantitative evaluation of the results obtained, we focus on the overall growth history. To this end, the predicted area of the tumor is compared to the actual one. In Fig. 15, the solid line represents the simulation

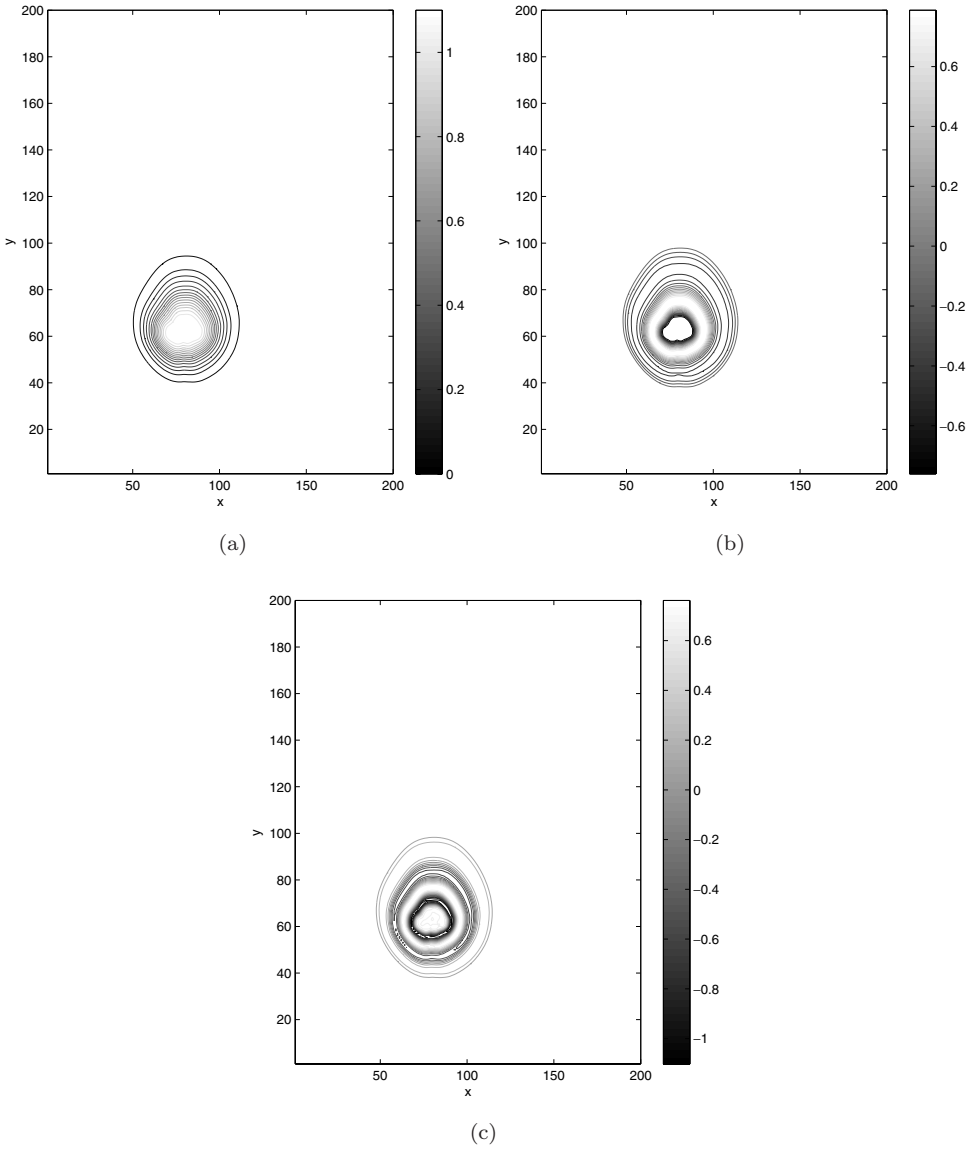
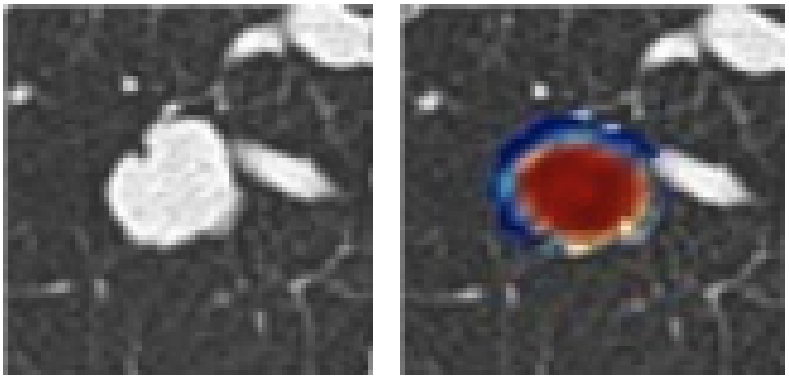


Fig. 13. POD modes for the proliferating cells density, Case I: (a) first mode, (b) third mode and (c) fifth mode.

while the circles are the computed areas of the scans. The square is the result of the last exam, performed recently, for which we provided the physicians a prediction of the size before the actual scan. Similar results on area or volumes can be obtained by identifying ODE-based models. Nevertheless, the present approach has the advantage of retrieving a precise spatial localization of the tumor as well as an indication of its cellular-type and nutrient distribution.



(a) (b)

Fig. 14. (Color online) Zoom of the superposition of image and computation for the data taken at November 2008: (a) fourth scan, (b) simulation. The color scale represents Y, from 0 (blue) to 1 (red).

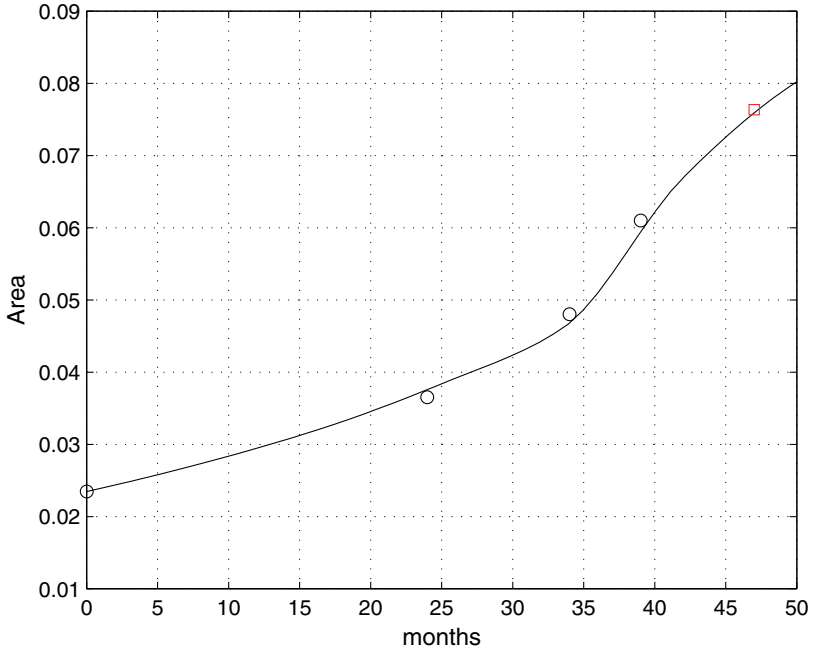


Fig. 15. Area as function of time.

We point out that only the first two scans were used in the identification. Nonetheless, the procedure is able to correctly detect the changes in trend at month 35, without having data near this point.

Let us comment this aspect in greater detail, since it is of interest for a deeper understanding of the procedure. The database of simulations built contains

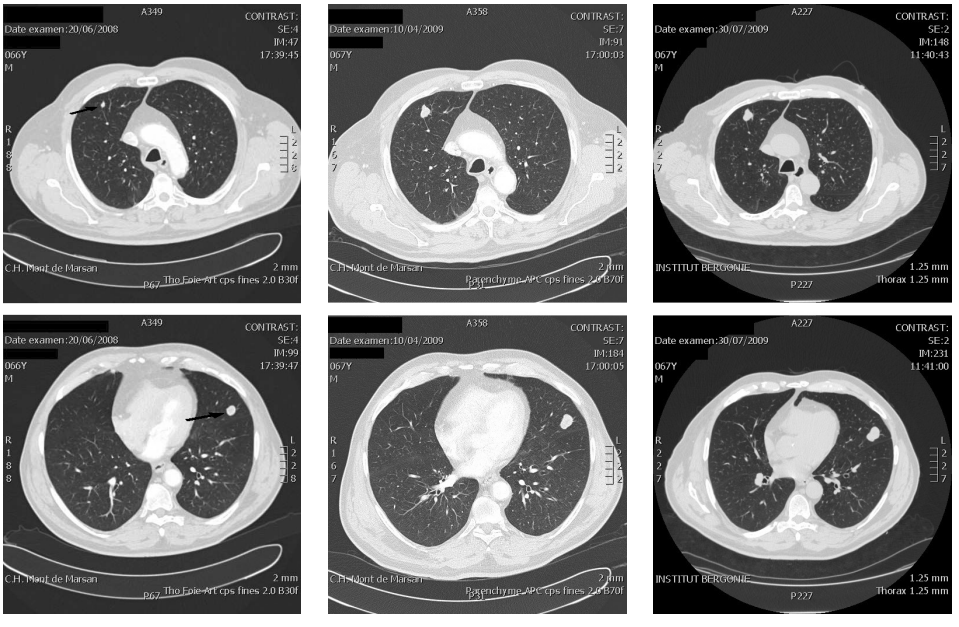
solutions of the PDE model commented in the above sections, and no solutions have been explicitly added of a logistic-type or exponential-type growth. The PDE model mimics a cellular cycle, whose rhythm may vary both in time and space. This kind of model is able to produce plateau-type solutions as well as exponential (and very aggressive) growths, when parameters are varied. In the database a wide range of solutions are represented, with compatible growth rates, and such a behavior is retained in the regularization via the POD modes. Once calibrated, the model has the solution (in terms of volume) represented in Fig. 15.

The volume error on the fourth scan corresponds to 10 days on a time of 39 months, and, for the last exam, to 20 days over 47 months.

5.2. *Case* II : *Lung nodules*

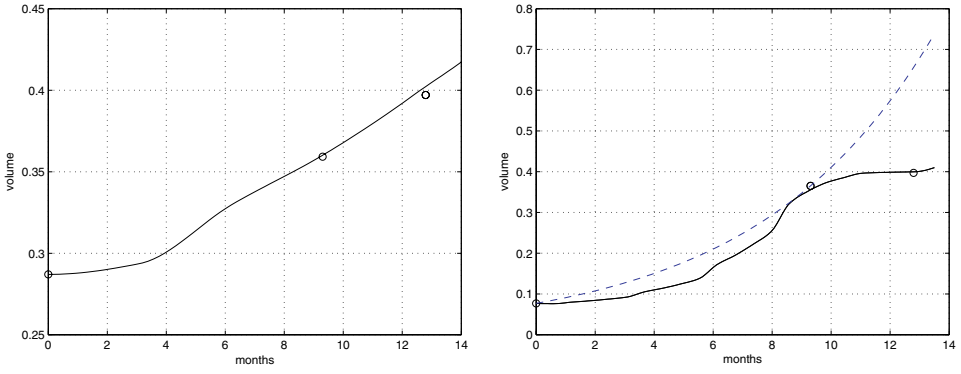
In this section the growth of two different metastatic nodules belonging to the same patient is considered. Their evolution is shown in Fig. 16. The nodules exhibit different dynamics: while the first one is characterized by a rapid phase of growth followed by a plateau-type solution, the second one has a regular growth.

As before, we try to recover the third scan by taking the first two images as assimilation data. The database used for the identification consists in 512 simulations for both nodules, varying the model parameters so that the growth rates are comparable to those observed between the first and the second scans, on a



(a) (b) (c)

Fig. 16. Nodule 1 (upper row) and Nodule 2 (lower row) at: (a) June 2008, (b) April 2009 and (c) July 2009.



(a) (b)

Fig. 17. Volume curve for: (a) Nodule 2 and (b) Nodule 1.

conventional time scale of 1. The low-order representation details are the same as those of the previous example.

In Fig. 17(a) the volume curve is shown for the Nodule 2. The prediction is correct and it substantially confirms what was obtained for the case discussed in the previous section. In Fig. 17(b) the prediction (in terms of volume) for Nodule 1 is shown. The identification is more difficult since two different growth patterns match the data with comparable residuals: the first one is an exponential-type of growth (dashed line in the figure) and the second one is a plateau-type evolution. Here the difficulty lies in the rapid growth of the nodule (its volume increased by a factor of 4): the configuration corresponding to the second image is too far in time from the first one. The phenomenon is undersampled and the approximation of the time derivative becomes poor. Local minima of the functional have comparable residual values and different families of solutions are possible. The correct curve has a residual norm slightly smaller than the exponential-type solution. However, in such a case, a third image is mandatory in order to get reliable predictions.

This case shows that the present technique is able to detect with a certain approximation different behaviors. Nevertheless there are certain reliability limitations linked to the relationship between the time scale on which the exams are performed and the proliferation speed of the tumor. These examples of applications are not yet conclusive from the clinical viewpoint. An appropriate experimental protocol is under definition with our partners at Institut Bergonié in order to systematically investigate the tumor growth prediction error.

6. Conclusions and Perspectives

An efficient procedure to perform identification of tumor growth models has been setup. The only part of the procedure that needs high performance computing is the

construction of the database and the computation of the auto-correlation matrices. However, this step of the procedure is easily parallelized and performed once for all for each patient, due to the different initial conditions.

The identification examples presented have been performed in a realistic setting where the observables are limited and the information scarce in time. Scalar fields that are not observables, such as proliferating cell density and oxygen concentration could be reconstructed. These fields play a crucial role in the tumor evolution determining the degree of aggressivity and eventually the way the tumor undergoes a metastatic process. In this sense the procedure can be interesting in a context of diagnosis, prognosis, clinical protocol definition and therapy optimization.

In this perspective, parametric models are interesting since they allow one to account for physical and biological effects that are not directly modeled. Several types of models can be analyzed, focusing on the capability of simple models to represent the solution of more complex ones.

Additional information coming from this study concerns *a posteriori* analysis of the reconstructed parameters. This analysis may allow to estimate parameters values that are invariant or weakly changing for different patients, and moreover which parameters affect more the tumor evolution. In other words, reference values of parameters and confidence intervals can be found for different models.

In the near future we will investigate the clinical relevance of the preliminary results obtained using biological data relative to a specific patient. More nodules of different biological-type will be considered in order to systematically assess the results of the identified models. In this respect the data assimilation method employed here can be improved to take into account uncertainty associated with the resolution of the scans, properties like initial tumor phenotype, spatial inhomogeneities, etc. To this end, a possible path of investigation is the minimization of the expected residuals with respect to the input data probability distributions, exploiting the ideas of polynomial chaos.[22] On the other hand, additional information of integral nature can be obtained from PET scans or even blood markers and more in general from functional imaging.

The extension to three-dimensional images will allow a more precise description of non-isotropic tumor growths and of the corresponding nonhomogeneous spatial distributions of the oxygen (nutrient) distribution C. In perspective, the possibility of modeling a nonhomogeneous spatial distribution of the nutrient may turn out to be important for example in the phenomenological modeling of cancer recurrence after chemotherapy.

Metastases to the lung seem to be the most adapted to be described by the Darcy-type system, thanks to the homogeneity of the tissues and of the vascularization. For other types of tumors (glioblastoma and breast tumor) the model has to be complexified to take the heterogeneity of tissues (breast tumor) and the presence of invasive cells (brain tumor) into account.

Acknowledgments

The authors wish to thank Drs. Jean Palussière, Françoise Bonichon and Michèle Kind (Institut Bergonié) for their invaluable help in the definition of a clinical test case and for their medical expertise.

References

1. T. Alarcón, H. M. Byrne and P. K. Maini, A cellular automaton model for tumor growth in inhomogeneous environment, *J. Theor. Biol.* **225** (2003) 257–274.
2. D. Ambrosi and L. Preziosi, On the closure of mass balance models for tumor growth, *Math. Models Methods Appl. Sci.* **12** (2002) 734–754.
3. G. Bayada, L. Chupin and S. Martin, Visco elastic fluid in a film domain, *Quart. Appl. Math.* **65** (2007) 625–651.
4. M. Bergmann, C. H. Bruneau and A. Iollo, Enablers for robust pod models, *J. Comput. Phys.* **228** (2009) 516–538.
5. F. Billy, B. Ribba, O. Saut, H. Morre-Trouilhet, T. Colin, D. Bresch, J. P. Boissel, E. Grenier and J. P. Flandrois, A pharmacologically based multiscale mathematical model of angiogenesis and its use in investigating the efficacy of a new cancer treatment strategy, *J. Theor. Biol.* **260** (2009) 545–562.
6. D. Bresch, T. Colin, E. Grenier, B. Ribba and O. Saut, A viscoelastic model for avascular tumor growth, *Disc. Cont. Dynam. Syst. Suppl.* (2009) 101–108.
7. D. Bresch, T. Colin, E. Grenier, B. Ribba and O. Saut, Computational modeling of solid tumor growth: The avascular stage, *SIAM J. Scientific Comput.* **32** (2010) 2321–2344.
8. H. Byrne and L. Preziosi, Modeling solid tumor growth using the theory of mixtures, *Math. Med. Biol.* **20** (2003) 341–366.
9. R. P. Fedkiw, B. Merriman and S. Osher, Simplified discretization of systems of hyperbolic conservation laws containing advection equations, *J. Comput. Phys.* **157** (2000) 302–326.
10. A. Friedman, A hierarchy of cancer models and their mathematical challenges, *Disc. Cont. Dynam. Syst. Ser. B* **4** (2004) 147–159.
11. F. Gibou, R. P. Fedkiw, L. T. Cheng and M. Kang, A second-order-accurate symmetric discretization of the Poisson equation on irregular domains, *J. Comput. Phys.* **176** (2002) 205–227.
12. P. Hahnfeldt, D. Panigrahy, J. Folkman and L. Hlatky, Tumor development under angiogenic signaling: A dynamical theory of tumor growth, treatment response, and postvascular dormancy, *Cancer Res.* **59** (1999) 4770–4775.
13. C. Hogea, C. Davatzikos and G. Biros, An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects, *J. Math. Biol.* **56** (2008) 793–825.
14. G.-S. Jiang and D. Peng, Weighted eno schemes for Hamilton–Jacobi equations, *SIAM J. Scientific Comput.* **21** (2000) 2126–2143.
15. J. Kaipio and E. Somersalo, Statistical inverse problems: Discretization, model reduction and inverse crimes, *J. Comput. Appl. Math.* **198** (2007) 493–504.
16. E. Konukoglu, O. Clatz, H. Bondiau, P.-Y. Delingette and N. Ayache, Extrapolating glioma invasion margin in brain magnetic resonance images: Suggesting new irradiation margins, *Med. Image Anal.* **14** (2010) 111–125.

17. J. S. Lowengrub, H. B. Frieboes, F. Jin, Y.-L. Chuang, X. Li, P. Macklin, S. M. Wise and V. Cristini, Nonlinear modelling of cancer: Bridging the gap between cells and tumours, *Nonlinearity*, **23** (2010) R1.
18. J. L. Lumley, The structure of inhomogeneous turbulent flows, in *Atmospheric Turbulence and Radio Wave Propagation*, eds. A. M. Yaglom and V. L. Tatarski (Nauka, 1967), pp. 166–178.
19. P. Macklin and J. Lowengrub, Evolving interfaces via gradients of geometry-dependent interior Poisson problems: Application to tumor growth, *J. Comput. Phys.* **203** (2005) 191–220.
20. Y. Mansury, M. Kimura, J. Lobo and T. S. Deisboeck, Emerging patterns in tumor systems: Simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model, *J. Theor. Biol.* **219** (2002) 343–370.
21. Y. M. Marzouk and H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, *J. Comput. Phys.* **228** (2009) 1862–1902.
22. H. N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Ann. Rev. Fluid Mech.* **41** (2009) 35–52.
23. J. T. Oden, A. Hawkins and S. Prudhomme, General diffuse-interface theories and an approach to predictive tumor growth modeling, *Math. Models Methods Appl. Sci.* **20** (2010) 477–517.
24. B. Ribba, T. Colin and S. Schnell, A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies, *Theor. Biol. Med. Model.* **3** (2006) 1–19.
25. B. Ribba, O. Saut, T. Colin, D. Bresch, E. Grenier and J. P. Boissel, A multiscale mathematical model of avascular tumor growth to investigate the therapeutic benefit of anti-invasive agents, *J. Theor. Biol.* **243** (2006) 532–541.
26. R. Rockne, E. C. Alvord, J. K. Rockhill and K. R. Swanson, A mathematical model for brain tumor response to radiation therapy, *J. Math. Biol.* **58** (2009) 561–578.
27. J. A. Sherratt and M. A. J. Chaplain, A new mathematical model for avascular tumour growth, *J. Math. Biol.* **43** (2001) 291–312.
28. M. Simeoni, P. Magni, C. Cammia, G. De Nicolao, V. Croci, E. Pesenti, M. Germani, I. Poggesi and M. Rocchetti, Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents, *Cancer Res.* **64** (2004) 1094–1101.
29. J. P. Sinek, S. Sanga, X. Zheng, H. B. Frioboes, M. Ferrari and V. Cristini, Predicting drug pharmacokinetics and effect in vascularized tumor using computer simulation, *J. Math. Biol.* **58** (2009) 485–510.
30. L. Sirovich, Low dimensional description of complicated phenomena, *Contemp. Math.* **99** (1989) 277–305.
31. P. K. Maini, T. Roose and S. J. Chapman, Mathematical models of avascular tumor growth, *SIAM Rev.* **49** (2007) 179–208.
32. L. Tenorio, Statistical regularization of inverse problems, *SIAM Rev.* **43** (2001) 347–366.
33. J. Weller, E. Lombardi and A. Iollo, Robust model identification of actuated vortex wakes, *Physica D* **238** (2009) 416–427.
34. Y. Zhang, M. A. Henson and Y. G. Kevrekidis, Nonlinear model reduction for dynamic analysis of cell population models, *Chem. Eng. Sci.* **58** (2003) 429–445.