Journal of Computational Physics 492 (2023) 112449

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Can cancer cells inform us about the tumor microenvironment?

Geir Nævdal [a], Steinar Evje [b,*]

[a] NORCE Norwegian Research Centre AS, Bergen, Norway
[b] University of Stavanger, Stavanger, Norway

ARTICLE INFO

Article history:
Received 4 October 2022
Received in revised form 14 April 2023
Accepted 17 August 2023
Available online 23 August 2023

Keywords:
Multiphase model
Reservoir modeling
Parameter estimation
Ensemble Kalman filter
Cell migration
Interstitial fluid pressure

ABSTRACT

Characteristics of the tumor microenvironment (TME) such as the leaky intratumoral vascular network and the density and composition of the desmoplastic extracellular matrix (ECM) contain essential information that determine the possibly heterogeneous interstitial fluid (IF) velocity field and interstitial fluid pressure (IFP). This information plays an important role for how anticancer drug that is delivered through the blood vasculature will distribute and possibly affect the tumor. The main question we deal with in this work is: Can we lure the cancer cells to reveal such information to us? By means of an in silico tumor model we demonstrate that subject to the condition that the tumor progression behavior is dominated by a cancer cell phenotype which moves by fluid-sensitive migration mechanisms as reported from experimental works, such information about the TME can be acquired by measuring the change in the cancer cell volume fraction distribution between two times T_0 and T_1, e.g., based on MRI images. We demonstrate this principle by using a continuum based multiphase model for tumor progression combined with assimilation of observed data through an ensemble Kalman filter approach which has been extensively and successfully used for updating advanced multiphase flow models in the context of reservoir simulation. Our results based on a synthetic dataset demonstrate how the methodology can be used to extract valuable quantitative information (e.g., interstitial fluid velocity field and fluid pressure, tissue conductivity reflecting ECM status, and effective vasculature conductivity) for which direct measurements may not be possible or impractical.

1. Introduction

1.1. Data-driven modeling

There is considerable interest in the use of computational methods to gain insight into tumor progression and treatment. Research groups have incorporated imaging data into macroscopic models of tumor cell growth in order to include elements of patient specific information when simulating the effect of a therapeutic drug on the tumor progression [70,71]. These works have elucidated the significant potential of using computational methods in a patient-specific and predictive framework to forecast the response of tumors to therapy. By relying on assimilation methods previously employed in the setting porous media reservoir flow [1,13,14,41], we seek in this work to demonstrate the forecasting power of biophysical

* Corresponding author.
E-mail addresses: geir.naevdal@norceresearch.no (G. Nævdal), steinar.evje@uis.no (S. Evje).

https://doi.org/10.1016/j.jcp.2023.112449

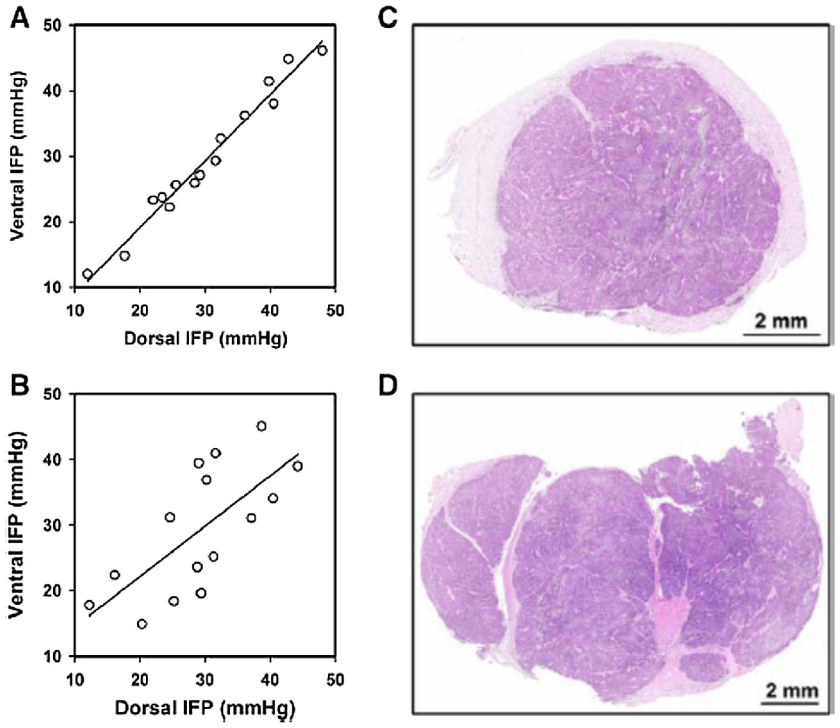

Fig. 1. IFP measured at two different positions inside the tumor for (A) 15 tumors with a homogeneous histology. (B) 15 tumors with a heterogeneous histology. The points represent single tumors. (C) Example of a tumor with homogeneous histology. (D) Example of a tumor with heterogeneous histology, showing that the tissue was divided into compartments by thick bands of connective tissue. Reproduced from [22][DOI: 10.1016/j.tranon.2019.05.012] with the courtesy of E. Rofstad.

and biomathematical modeling in the context of tumor growth and therapy [73]. This study aims to investigate two related issues: (i) whether a parameter estimation technique together with a partial differential equation (PDE) based model is able to gain valuable information about the tumor microenvironment (TME); (ii) and if this information can be used to predict further development of the tumor. TME heavily impacts how the tumor is developing through the different stages [69,58], in addition to the evolution of malignancies. The penetration distance and distribution of drugs, i.e. the effectiveness of the drug, is hindered by the tumor microenvironment [60,28]. A motivation for the current work has been a recent work on preclinical models of cervical carcinoma cancer [22]. The authors investigated interstitial fluid pressure in HL-16 cervical carcinoma xenograft tumor models. The study showed that IFP could differ substantially within the central region of the tumor as shown in Fig. 1. Significant IFP heterogeneity inside the tumor was associated with the presence of thick connective tissue bands that divided the tumor into compartments. Fibroblasts that are recruited into tumor masses, called tumor-associated fibroblasts (TAFs) or cancer-associated fibroblasts (CAFs), are the main cellular components of the surrounding stroma of many solid cancers. Evidently, these CAFs provide a range of different cytokines, growth factors, tissue remodeling enzymes and extracellular matrix (ECM) components, all of which regulate the tumor stroma [11,31]. Gaggioli et al. [20] discovered that fibroblasts may act as leader cells of a collective group of following carcinoma cells. This would allow cancer cells to retain their epithelial features, which is observed in solid tumors *in vivo* and also *in vitro* [75,16], while having a mesenchymal-like cell to lead them to invade the adjacent stroma. Furthermore, Gaggioli propose that cancer cells move within paths in behind fibroblasts. These paths are created by the fibroblasts both through proteolytic activity and force-mediated matrix remodeling [6]. Fibroblast-enhanced tumor cell migration was reported by Shieh et al. in the presence of interstitial flow where, similar to Gaggioli et al., it was found that fibroblasts create tracks within the ECM for cancer cells to follow [55]. Later, Labernadie et al. identified another possible mechanism to which fibroblasts lead cancer cells through a heterotypic cell-cell adhesion between the two types of cell [36]. This mechanism represents a direct fibroblast-cancer cell interaction through N-cadherin/E-cadherin junctions. In a recent work a cell-fibroblast-fluid multiphase model [66] was developed and explored to shed light on the preclinical findings reported in [22]. A special feature of the model is that it can explore possible interrelations, in a continuum based setting, between characteristics of TME which affect IF and IFP and fluid-sensitive migration mechanisms for fibroblasts and cancer cells, which dictate the tumor progression and possible invasive behavior as explored experimentally in [3,22].

1.2. Research question/problem setup

The research question we focus on in this work is: Given that we know the tumor cell distribution in a 2D setting in terms of a volume fraction $\alpha_c(x,t)$ at a time T_0. By assuming that we have access to information about the tumor cell

distribution (through, e.g., MRI image) at a later time T_1, how can we extract quantitative information about the tumor microenvironment (TME) like tissue conductivity (which is related to ECM structure), IF velocity field and IFP as well as effective vascular conductivity? A more precise statement in terms of mathematical variables and parameters is made after the mathematical model with its different assumptions has been given in Section 2. The proposed methodology implies that we (indirectly) can extract useful information about TME like ECM bands and IFP and tumor vasculature while the tumor is still developing. Such information was collected and measured after the mice were anesthetized in [22]. A relevant question to address is under what assumptions this is possible as it seems by no means obvious that observation of cancer cell migration behavior only can reveal information about the IFP and IFP velocity field.

To put this question into perspective it seems instructive to refer to recent research where physics-informed deep learning methods are developed in the context of fluid mechanics [53,54,52]. The aim is to learn velocity and pressure fields from flow visualizations. More precisely, by means of the assumption that the underlying dynamic is described by the classical Navier-Stokes (NS) equations and observations (in terms of images) of a passive scalar, such as the transport of dye or smoke in physical systems and contrast agents in biological systems, one tries to extract information about corresponding velocity and pressure fields (which are much more challenging to measure directly). The key is the combination of many images of the passive agent concentration that follows the fluid flow and incorporating the NS equations in the loss function (objective function) used for training the physics-informed deep learning networks [54]. In the tumor setting, a main additional complicating factor is that the manner cancer cells make use of or relates to fluid is more involved. Our framework relies on a mathematical model for tumor progression where fibroblasts and tumor cells migrate driven by chemotaxis towards higher concentrations of chemical agents. The chemical agents are governed by the fluid velocity field [55,49,63,17]. The fluid velocity field and IFP is in turn a result of TME characteristics through tissue conductivity (which is a result of ECM structure) and tumor vasculature through internal hydraulic pressure and effective vascular conductivity. Hence, in theory there is room for detecting TME characteristics through observing changes in the cancer cell volume fraction, although it seems to the best of our knowledge as a relatively unexplored issue. The framework we use for implementing and testing this approach is to combine the cell-fibroblast-fluid model [67] for tumor progression with assimilation of observed cancer cell volume fraction at times T_0 and T_1, respectively through an ensemble Kalman filter approach.

The ensemble based approaches have been applied for a multitude of large-scale estimation problems [1,13–15,27,41,64], including within weather forecasting and reservoir engineering. It has a particular strength in updating spatial fields. The application investigated in this paper has clear similarities to updating multiphase models within reservoir engineering. In both cases one would update several spatially varying parameter fields based on available measurements. The current work utilizes the experience built up over a couple of decades on the use of ensemble based approaches within reservoir engineering. A main difference between this approach and the one mentioned above based on physics-informed deep learning, is that we only have observations of cancer cell distribution at two different times whereas the first make use of measurements scattered in both space and time. In that respect, our approach bears similarity to the calibration method that has been used more recently in a series of works [70,71]. On the other hand, their approach and the one we pursue differ in the sense that the cancer cell migration mechanisms they account for are not directly coupled to fluid flow but tissue mechanical properties and therefore does not give estimates of IFP and IF velocity.

Why are we interested in the TME characteristics like IFP and the corresponding IF velocity field? While it is a potent driver of invasion in brain [46,33] skin [23] and breast cancer [55,49] IF flow has been poorly measured and characterized in vivo. It is thereby pertinent to develop accurate means to quantify IF flow in the pre-clinical and clinical settings through noninvasive methods [34,68,59]. It is of interest to extract information pertaining to the TME with a special eye to information about the leaky intratumoral vascular network and characteristics of the desmoplastic ECM. The ECM possibly involves heterogeneous dense structures of stroma and presence of CAFs that play a central role in remodeling of ECM as well as tumor progression. This is relevant for the planning and performance of neoadjuvant therapies whose purpose is to shrink the primary tumor before it is removed through surgical operation. Typically, a therapeutic drug will reach the primary tumor through the intratumoral microvascular network by leaking through the aberrant blood vessel walls [58]. Depending on the amount of fluid that leaks and the ECM density and structures, an elevated possible heterogeneous, interstitial fluid pressure (IFP) is formed on the inside of the tumor with a rapid decrease at the tumor margin. This results in a corresponding fluid velocity field in the region close to tumor margin which will affect the drug transport. Hence, the efficiency of the drug most likely depends critically on the above mentioned parameters associated with TME, in addition, to other aspects like cancer cell resistancy to the drug in question [58].

1.3. Methodology

We explore our in silico tumor framework by using synthetic generated data. That is, we assign a stochastic model for the unknown parameter vector θ which typically involves spatial-dependent Gaussian random fields as well as constant parameters. A Gaussian random field is specified through the given mean, variance and correlation. The output of the model is the state vector $Y = F(X, \theta)$ where α_c (cell volume fraction) is one of them and X is the input vector of known input data (initial states and parameters that characterize the model system and assumed to be known). F represents the solution operator associated with the PDE model. The methodology is as follows:

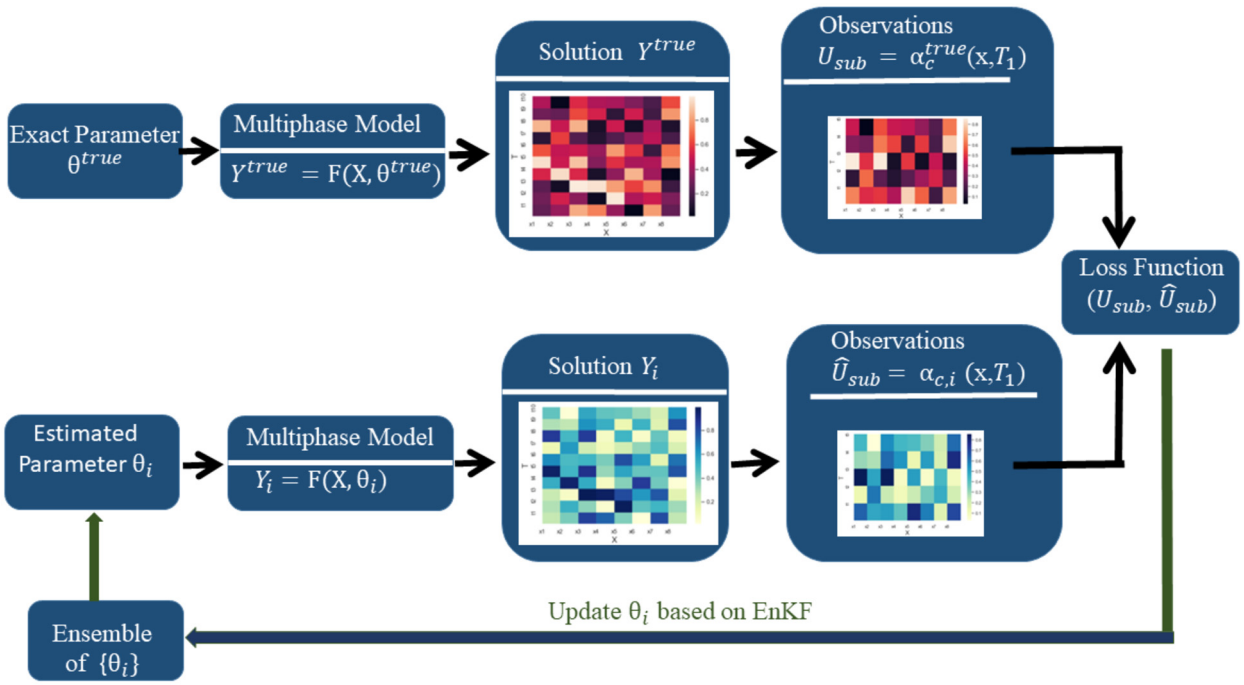

Fig. 2. Illustration of the methodology used in this work based on the EnKF approach. Note that Y refers to the full solution of the model (1) whereas U refers to a subset of Y representing the observation data which in our case is the cancer cell distribution $\alpha_c(\mathbf{x}, T_1)$.

(i) A parameter vector θ^{true} is generated from the specified stochastic model(s). This is used to compute the true tumor status represented by the state vector $Y^{\text{true}} = F(X, \theta^{\text{true}})$ through the multiphase tumor model [66]. From this we extract the observed tumor status denoted by U_{sub} which in our case is represented by the cell volume fraction $U_{\text{sub}} = \alpha_c^{\text{true}}(\mathbf{x}, T_1)$ at time T_1. See upper action line of the scheme in Fig. 2. The question is: To what extent are we able to identify a good approximation θ^{est} of θ^{true} by only making use of the observation $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ (including measurement noise)? By populating the computer model with the estimated parameter vector θ^{est}, how well can we predict the true behavior represented by Y^{true} at the later time T_2?

(ii) The first step is to span out the solution space by generating an ensemble $\{\theta_i\}_i^N$ of N different parameter vectors θ_i based on prescribed stochastic models. The solution vectors $\{Y_i\}_{i=1}^N$ where $Y_i = F(X, \theta_i)$ then represent the solution space. See the lower action line of the scheme in Fig. 2. We might expect to see a large variation in the computed tumor progression behavior, similar to the preclinical study shown in Fig. 1 (panel A and B), where IFP vary between 10 and 45 mmHg.

(iii) By using the information from $\widehat{U}_{\text{sub}} = \{\alpha_{c,i}(\mathbf{x}, T_1)\}_{i=1}^N$ extracted from $\{Y_i\}_{i=1}^N$, we can compute the error as compared to $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$. The ensemble Kalman filter approach then allows us to compute an updated ensemble of parameter vectors $\{\theta_i^*\}_{i=1}^N$ (possible through several steps to deal with the nonlinearity in the problem) designed to minimize this error. See the scheme in Fig. 2.

(iv) The final step is to compute an updated parameter vector $\overline{\theta}^*$ as an average of $\{\theta_i^*\}_{i=1}^N$, from which we can compute the tumor state $Y^* = F(X, \overline{\theta}^*)$ where the error between $\alpha_c^*(\mathbf{x}, T_1)$ and $\alpha_c^{\text{true}}(x, T_1)$ has been minimized and where $\alpha_c^*(\mathbf{x}, T_2)$ hopefully represents a good prediction of $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$.

We explore this approach subject to certain constraints. We first assume that the tumor progression in the simulated period $[0, T_2]$ is dominated by cancer cell migration and not growth. Then, we evaluate how inclusion of growth, whose parameters also must be estimated, may have an impact on the ability of detecting information about TME through the parameter vector θ. In order to make this work focused, parameters that characterize the cancer cell and fibroblast phenotype are assumed to be known. However, we test the robustness of the method by considering different phenotypes. Our conclusion is that the methodology seems to offer a promising general approach for data-driven simulation of tumor progression, at least as far as we can assess from the testing based on synthetic data.

2. Method

We will use the multiphase model that was explored in [66] to shed light on the preclinical data in [22] to represent the in silico tumor model. This model bears many similarities with previous models as described in [7,26,72,44,12,39,45,4,70, 29,24,47]. Motivated by recent results which emphasize the possible role played by fluid flow and related forces as a means

Table 1
Variables used in the model (1).

Variable	Description
$\phi, \alpha_c, \alpha_f, \alpha_w$	tissue porosity, volume fraction of cell, fibroblast and fluid
S_c, S_f	cell growth/death
$\mathbf{u}_c, \mathbf{u}_f, \mathbf{u}_w$	interstitial cell, fibroblast and fluid velocity
C, H	chemokine and TGF (transforming growth factor)
P_w, P_c, P_f	IF pressure, solid stress associated with cancer cell, fibroblast
$\Delta P_{fw}, \Delta P_{cw}, \Lambda_C, \Lambda_H$	CAF-CAF, cell-cell stress, chemokine and TGF chemotactic stress
$\zeta_c, \zeta_f, \zeta_w, \zeta_{cf}$	cell-ECM, fibroblast-ECM, fluid-ECM and cell-fibroblast coefficients
Q_v, Q_l	transvascular flux related to blood vessels, lymphatic vessels
T_v, T_l	conductivity of vascular vessel wall, conductivity of lymphatic vessel wall
$\tilde{P}_v^*, \tilde{P}_l^*$	effective vascular pressure, effective lymphatic pressure
D_C, D_H	diffusion coefficients
Q_C, Q_H	production/decay rates associated with C, H

for cancer cells to seed distant metastases [17], we will inform the model with experimentally observed fluid-sensitive migration mechanisms, both for fibroblasts and cancer cells. In the following we give a compact summary of the model and refer to [66] for more details.

2.1. Underlying assumptions and principles for the mathematical multiphase model

1. We consider three separate phases in terms of cancer cells α_c, fibroblasts (CAFs) α_f and interstitial fluid α_w which are represented through separate mass balance equations and momentum balance equations which account for the stress tensor and external forces for each individual phase [26,39].
2. Cancer cells and CAFs are able to generate friction forces and active traction forces [2]. To represent these mechanisms the tissue is considered as a porous medium composed of a stagnant, rigid solid matrix that essentially represents the ECM structures whereas the porous space is available for the cancer cells, CAFs and fluid to move through. The momentum balance laws, which are a generalization of the standard Darcy equation commonly used [72], explicitly represent interaction forces between cancer cells and CAFs as well as resistance forces (friction) between the different phases and underlying matrix [50]. This allows to explore how elevated intratumoral IFP is a result of the tissue conductivity and aberrant intratumoral vasculature caused by solid stress [45,4,12].
3. The intratumoral blood and lymphatic vasculature are modeled by using a continuum approach based on Starling's law [72,45,12,59].
4. We explicitly account for solid stress through P_c and P_f, respectively, associated with cancer cells and fibroblasts, similar to what is done in [18,72,26,39,44]. More precisely, $P_c = P_w + \Delta P + \Lambda$ where P_w is IFP and ΔP and Λ represent stress generated by the moving cells/fibroblasts as they interact with the surrounding ECM structure through diffusive spreading and chemotactic migration, respectively,
5. Solid stress is reported to be a driver for the development of the aberrant intratumoral vasculature with high intravascular pressure and high vasculature conductivity reflecting a high degree of leaky walls [28,48]. Solid stress may also suppress proliferation and promote cancer cell phenotype with higher degree of viability [18,28,43].

2.2. Tumor cell-fibroblast-fluid model

Let ϕ be the volume fraction occupied by the pore space through which the cells and fluid can move whereas $1-\phi$ represents the remaining matrix volume fraction. The mathematical model takes the following form using variables as summed up in Table 1:

$$
\begin{aligned}
&(\phi\alpha_c)_t + \nabla\cdot(\phi\alpha_c\mathbf{u}_c) = S_c \\
&(\phi\alpha_f)_t + \nabla\cdot(\phi\alpha_f\mathbf{u}_f) = S_f, \qquad\qquad \alpha_c+\alpha_f+\alpha_w = 1 \\
&(\phi\alpha_w)_t + \nabla\cdot(\phi\alpha_w\mathbf{u}_w) = -S_c - S_f + Q, \qquad Q = Q_v - Q_l \\
&\alpha_c\nabla(P_w + \Delta P_{cw} + \Lambda_C) = -\zeta_c\mathbf{u}_c + \zeta_{cf}(\mathbf{u}_f - \mathbf{u}_c) \\
&\alpha_f\nabla(P_w + \Delta P_{fw} + \Lambda_H) = -\zeta_f\mathbf{u}_f - \zeta_{cf}(\mathbf{u}_f - \mathbf{u}_c) \\
&\alpha_w\nabla P_w = -\zeta_w\mathbf{u}_w \\
&(\phi\alpha_w C)_t = \nabla\cdot(D_C\nabla C) - \nabla\cdot(\phi\alpha_w C\mathbf{u}_w) + Q_C \\
&(\phi\alpha_w H)_t = \nabla\cdot(D_H\nabla H) - \nabla\cdot(\phi\alpha_w H\mathbf{u}_w) + Q_H
\end{aligned}
\tag{1}
$$

where $\mathbf{u}_i = (u_i^x, u_i^y)$ for $i = c, f, w$. Since ϕ is here assumed to be constant it can be incorporated in the dimensionless time and grouped with the velocities $\mathbf{u}_w, \mathbf{u}_f, \mathbf{u}_c$ [67] (Appendix A). Mass balance equations for cancer cells, fibroblasts

Table 2
Source terms in $(1)_{7,8}$ accounting for production/decay of chemokine and growth factor.

Function		Description
Q_C	=	$\alpha_c\alpha_f\lambda_{11}\left(1-\left(\frac{C}{C_M}\right)^{\nu_C}\right)-\lambda_{12}\alpha_c C - M_C Q_l C$
Q_H	=	$\alpha_f\lambda_{21}\left(1-\left(\frac{H}{H_M}\right)^2-\left(\frac{H}{H_M}\right)^{\nu_H}\right)-\lambda_{22}\alpha_f H-\lambda_{23}H - M_H Q_l H$

and interstitial fluid are accounted for through $(1)_{1\text{-}3}$. The next three equations, $(1)_{4\text{-}6}$, are the corresponding momentum balance equations [26,39]. The RHS of $(1)_4$ is separated into two terms: $\zeta_c\mathbf{u}_c$ representing the resistance against migration felt by cancer cells from the ECM, $\zeta_{cf}(\mathbf{u}_f-\mathbf{u}_c)$ represents the drag force exerted by fibroblasts on cancer cells. $\Delta P_{cw}(\alpha_c)$ and $\Delta P_{fw}(\alpha_f)$ are functions quantifying the elevated pressures seen in the cell phase compared to interstitial fluid pressure P_w. This yields a diffusive-like cell migration, i.e., a random migration [72,70,71]. The potential function $\Lambda_C(C)$ represents the additional phase pressure of cancer cells due to chemotaxis toward chemokine (C). $\Lambda_H(H)$ represents the additional phase pressure associated with the fibroblast phase due to chemotaxis toward growth factor (H). The two remaining equations, $(1)_{7\text{-}8}$, are transport-reaction equations for the chemical components chemokine and TGF. Rate terms for production/decay of C and H are given by Q_C and Q_H and are specified in Table 2. The values used for different rate terms λ_{ij} are given in Table 6 (Appendix A) and are largely based on previous calibration [55,63,66].

2.3. Cell-matrix and cell-fibroblast interaction coefficients

The interaction coefficients which are used in the model are specified as follows:

$$\zeta_w = I_w k_w \alpha_w^{r_w}, \qquad \zeta_c = I_c k_c \alpha_c^{r_c}, \qquad \zeta_f = I_f k_f \alpha_f^{r_f}, \qquad \zeta_{cf} = I_{cf}\alpha_c^{r_{cf}}\alpha_f^{r_{fc}} \tag{2}$$

The parameters I_w, I_c and I_f represent static properties of the tissue, whereas k_w, k_c and k_f can account for dynamic properties related to for instance ECM remodeling and fiber alignment. The coefficients r_w, r_c, r_f, r_{cf} and r_{fc} play a similar role to the use of relative permeability functions in standard Darcy's equation approach extended to several phases [50,66]. The parameter I_w represents the fluid-ECM resistance force and is set such that a realistic interstitial fluid velocity and IFP are obtained, typically in the range of 0.1-1 μm/s and $5-50$ mmHg, respectively [56,49,23,22,17]. Through the parameter $k_w(\mathbf{x})$ we can account for local heterogenities like the bandlike structures reported in [22] and seen in Fig. 1. Moreover, I_{cf} is a constant determining the order of magnitude of the cell-fibroblast interaction. We assume that fibroblasts remodel and degrade the ECM, making it easier for the cancer cells to migrate in their path [20,55,6]. This is represented through the following equation:

$$k_c = 1 - A\big(1-\exp(-B\alpha_f)\big) = k_c(\alpha_f) \tag{3}$$

where A and B are specified constants. We refer to Table 5 (Appendix A) for values used.

2.4. Starling law

In nearly all tissue, plasma leaks out of blood capillaries, flows through the interstitium and drains into lymphatic vessels, where it passes through lymph nodes before being returned to the venous blood [28]. This circulation is expressed on the RHS of $(1)_3$ through the term $Q = Q_v - Q_l$. Starling Law is used for the flow of fluid into the interstitium given by [72,12]

$$Q_v = T_v\big(P_v^* - P_w - \sigma_T(\pi_v^* - \pi_w)\big) = T_v\big(\widetilde{P}_v^* - P_w\big) \qquad T_v = L_v\frac{S_v}{V} \tag{4}$$

where $\widetilde{P}_v^* = P_v - \sigma_T(\pi_v^* - \pi_w)$. Here L_v is the hydraulic conductivity (m^2 s/kg=m/Pa s) of the vessel wall, S_v/V (m^{-1}) the exchange area of blood vessel per unit volume of tissues V. P_v^* and P_w the vascular and interstitial fluid pressure, π_v^* and π_w the osmotic pressure in the vascular and interstitial space and σ_T the osmotic reflection coefficient for plasma proteins.

The lymphatic system drains excessive fluid from the interstitial space and returns it back to the blood circulation, as expressed by Q_l. The lymphatics are typically not functional in the intratumoral region [28,72,48,58]. The loss of functionality is caused by compressive solid stress that is developed in tumors. Through this stress the intratumoral lymphatic vessels collapse. Similar to the expression of Q_v in (4), we use an expression of the following form to express the absorption of fluid through the lymphatics

$$Q_l = T_l(P_w - P_l^*), \qquad T_l = L_l\frac{S_l}{V}. \tag{5}$$

Here L_l is the hydraulic conductivity of the lymphatic vessel walls whereas S_l/V is the surface area of the lymphatic vessel per volume unit of tissues V and P_l^* is the effective lymphatic pressure.

2.5. Cancer cell and CAF velocities

From (1) an expression for the interstitial cell velocity $\mathbf{u}_c$, CAF velocity $\mathbf{u}_f$, and IF velocity $\mathbf{u}_w$ can be derived [63,67]:

$$\mathbf{u}_c = \underbrace{\frac{\hat{f}_c}{\alpha_c}\mathbf{U}_T}_{\text{(i)}} - \underbrace{\frac{\hat{h}_1+\hat{h}_2}{\alpha_c}\nabla(\Delta P_{cw}(\alpha_c))}_{\text{(ii)}} - \underbrace{\frac{\hat{h}_1+\hat{h}_2}{\alpha_c}\nabla(\Lambda_C(C))}_{\text{(iii)}} + \underbrace{\frac{\hat{h}_2}{\alpha_c}\nabla\Lambda_H(H)}_{\text{(iv)}} + \underbrace{\frac{\hat{h}_2}{\alpha_c}\nabla(\Delta P_{fw}(\alpha_f))}_{\text{(v)}} \tag{6}$$

$$\mathbf{u}_f = \frac{\hat{f}_f}{\alpha_f}\mathbf{U}_T + \frac{\hat{h}_2}{\alpha_f}\nabla(\Delta P_{cw}(\alpha_c)) + \frac{\hat{h}_2}{\alpha_f}\nabla(\Lambda_C(C)) - \frac{\hat{h}_2+\hat{h}_3}{\alpha_f}\nabla\Lambda_H(H) - \frac{\hat{h}_2+\hat{h}_3}{\alpha_f}\nabla(\Delta P_{fw}(\alpha_f)) \tag{7}$$

$$\mathbf{u}_w = \frac{\hat{f}_w}{\alpha_w}\mathbf{U}_T + \frac{\hat{h}_1}{\alpha_w}\nabla(\Delta P_{cw}(\alpha_c) + \Lambda_C(C)) + \frac{\hat{h}_3}{\alpha_w}\nabla(\Delta P_{fw}(\alpha_f) + \Lambda_H(H)). \tag{8}$$

We refer to Table 4 (Appendix A) for more information related to the functions $\hat{f}_c$, $\hat{f}_f$, and $\hat{h}_1, \hat{h}_2, \hat{h}_3$. It follows that these functions depend directly on the correlations (2). The terms on the RHS of (6) represent different migration mechanisms, resulting in the overall velocity $\mathbf{u}_c$ [66]:

(i) Fluid-generated stress
(ii) Diffusive spreading
(iii) Chemotaxis of cells toward higher concentration gradients in chemokine
(iv) Migration due to fibroblast chemotaxis towards higher concentration gradients in TGF
(v) Migration due to fibroblast diffusion.

2.6. Summary of the model

Below is a short summary of main characteristics of the model which has been designed to be consistent with migration mechanisms observed through *in vitro* studies [55,49,46,63,67] and used to shed light on preclinical studies [22,67]:

1. Fluid is leaking from the vascular system residing near the periphery of the initial tumor and is absorbed by lymphatics in the peritumoral region, characterized by $Q = Q_v - Q_l$ where Q_v is given by (4) and Q_l by (5). The resulting fluid flow field is directed from the vascular to the lymphatic system [23,12]. Tumors develop elevated IFP because they show high resistance to blood flow (i.e., $\widetilde{P}_v^*$ in (4) is high), low resistance to transcapillary fluid flow (i.e., T_v in (4) is high), and impaired lymphatic drainage (i.e., Q_l in (5) is located to the peritumoral region) [3].
2. Chemical components in terms C (chemokine) and H (growth factor TGH) are secreted and produced by cancer cells and fibroblasts. These chemical components diffuse and advect with the fluid flow, creating chemical gradients downstream of the flow $(1)_{7-8}$. Cancer cells α_c and fibroblasts α_f migrate towards positive gradients of their respective chemotactic chemical, chemokine and transforming growth factor (TGF).
3. A large resistance force is imposed on the cancer cells, making it difficult for them to migrate. I.e., the cell-ECM interaction coefficient ζ_c in (2) takes a large value I_c. Fibroblasts, on the other hand, are considered much more mobile. They migrate driven by chemotaxis. Cancer cells may attach themselves onto fibroblasts through the interaction term ζ_{cf} and/or follow tracks where fibroblasts have migrated through the term (3) due to the diminished resistance in their wake. This yields a largely fibroblast-dependent cancer cell migration [20,55,36].

2.7. Estimation of parameters that characterize TME

The aggressiveness of the cancer cells through the model (1)–(5) lies in their interaction with CAFs combined with fluid-sensitive migration mechanisms. The model is armed with a fixed set of parameters that determines the cancer cell and fibroblast phenotype as specified in Table 4, 5, and 6, as used in [67]. On the other hand, parameters that characterize the TME, such as the leaky intratumoral vascular network through $Q_v(\mathbf{x}, t)$ given by (4) and the density and composition of the desmoplastic ECM through $k_w(\mathbf{x})$ in (2) and initial fibroblast distribution $\alpha_f(\mathbf{x}, t=0) = \alpha_f^0(\mathbf{x})$, are unknown and must be estimated. More precisely, the parameters that we estimate are:

(a) $k_w(\mathbf{x})$ which accounts for the inverse of the interstitial hydraulic conductivity through which the possible compartment structure (see Fig. 1) can be represented.
(b) $\alpha_f^0(\mathbf{x})$ which represents the initial fibroblast distribution assumed to be placed in a heterogeneous bandlike region that surrounds the primary tumor [65,37].
(c) the intratumoral vascular filtration parameter $T_v(\mathbf{x})$ involved in (4) which accounts for the area and leakiness of blood vessel walls.

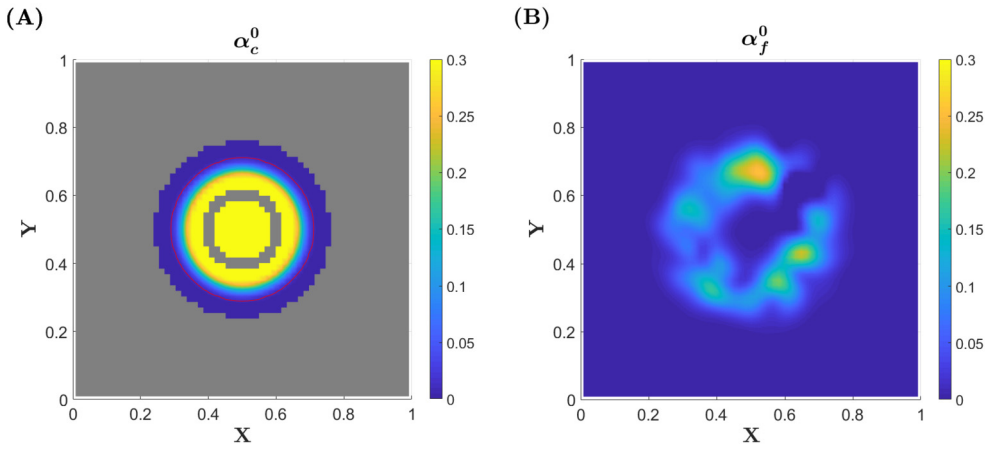

Fig. 3. (A) The characteristic length is $L^* = 1$ cm (Table 3, Appendix A). Hence, the diameter of the initial primary tumor is around 0.3 (dimensionless), i.e., 0.3 cm, in the domain of size 1×1 (dimensionless), i.e., 1 cm $\times$ 1 cm. Initial cell volume fraction, $\alpha_c^0(\mathbf{x})$. The red ring surrounding the tumor has a volume fraction of $\alpha_c = 0.01$. The grey area within the tumor is a superimposed illustration of where the vascular system is located. The grey peritumoral area shows where the fluid is drained through the functioning lymphatic system. **(B)** Example of initial fibroblast distribution $\alpha_f^0(\mathbf{x})$ in a bandlike region surrounding the initial tumor. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

As observed data we shall only use the cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ assumed to be observed at some time T_1. We perform an *in silico* simulation study to assess the ability of this methodology to capture the TME-related parameters. Then we populate the forward model with these key model parameters and use it to simulate the tumor progression in the time period $[0, T_2]$ with $T_2 = 2T_1$ where we can do a final assessment by comparing with observed data $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$. We refer to Appendix B for a description of the EnKF based method for how to compute the updated parameter vector $\theta = (\alpha_f^0, k_w, T_v)$.

2.8. Cancer cell and fibroblast phenotype

Important parameters that characterize the strength of the chemotactiv driven migration are Λ_{C1} for cancer cells and Λ_{H1} for fibroblasts. We refer to Table 4 (Appendix A) for the functional expressions and parameter values. Similarly, the parameter I_{cf} (see Table 5, Appendix A) which determines how strongly the more immobile cancer cells tend to attach to the more mobile fibroblasts, is important for the tumor progressing behavior. We keep these parameters fixed when we explore how we can identify the parameter vector θ that characterizes TME. However, we want to verify that the method we use for searching for an estimate of the parameter vector θ is not sensitive to the specific phenotype. Therefore, we will also in the next section explore simulation results when we vary these parameters.

3. Results

3.1. Initialization of the model

The unknown variables to solve for by using (1)–(5) are $\mathbf{Y} = (\alpha_c, \alpha_f, \alpha_w, C, H, \mathbf{u}_c, \mathbf{u}_f, \mathbf{u}_w, P_w) = F(X, \theta)$ where X is the input vector of known input data (initial states and parameters that characterize the model system and assumed to be known) and θ is the unknown parameter vector. The model (1)-(5) is combined with the boundary condition

$$P_w|_{\partial\Omega} = P_B^*, \quad \frac{\partial}{\partial \nu} C|_{\partial\Omega} = 0, \quad \frac{\partial}{\partial \nu} H|_{\partial\Omega} = 0, \quad t > 0 \tag{9}$$

where ν is the outward normal on the domain $\partial\Omega$ with $\Omega = [0, 1] \times [0, 1]$ (dimensionless) and P_B^* is atmospheric pressure. The corresponding initial data are

$$\alpha_c(\mathbf{x}, t = 0) = \alpha_c^0(\mathbf{x}), \quad \alpha_f(\mathbf{x}, t = 0) = \alpha_f^0(\mathbf{x}), \quad C(\mathbf{x}, t = 0) = H(\mathbf{x}, t = 0) = 0. \tag{10}$$

We consider a grid of 61×61 grid blocks similar to what was used in [67]. For more details related to the numerical solution method the interested readers are referred to [67].

The initial tumor cell volume fraction $\alpha_c^0(\mathbf{x})$ is assumed to be known and is shown in Fig. 3 (panel A). The tumor has a high concentration core of $\alpha_c = 0.3$ and will quickly decrease at the periphery, around 0.15 cm from the center of the primary tumor. Hence, we have developed a tumor with homogeneous tumor cell distribution and well defined margins.

The initial fibroblast volume fraction $\alpha_f^0(\mathbf{x})$ is one of the unknown parameters to be estimated. A typical example is shown in Fig. 3 (panel B). As mentioned above, the vector of unknown parameters θ is given by

$$\theta = (\alpha_f^0(\mathbf{x}), k_w(\mathbf{x}), T_v(\mathbf{x})). \tag{11}$$

The implementation of ECM bands through k_w is motivated from the study of cervical carcinoma tumor HL-16 models in [22], see Fig. 1. Herein, the bands are thought to affect the intratumoral pressure profile, yielding a heterogeneous IFP. We refer to the recent work [66] for a discussion of this issue in light of the model (1)-(5). In Fig. 3 (panel A) we can see the superimposed vascular system which is restricted to a bandlike region within the tumor (grey ring-shaped region). The filtration parameter T_v is assumed to take a constant value in this region. The intravascular pressure $\widetilde{P}_v^*$ involved in (4) is assumed known and set to be $\widetilde{P}_v^* = 6000$ Pa (45 mmHg) relatively atmospheric pressure.

The lymphatic network is restricted to the peritumoral grey region seen in Fig. 3 (panel A). We assume that the inner lymphatic pressure P_l^* is known and set to $P_l^* = -300$ Pa (-2.25 mmHg) while the filtration constant T_l is set to $T_l T^* = 0.0054$ [1/Pa] where T^* is the reference time (see Table 3) which typically gives IFP in the range [-3,3] mmHg in the peritumoral region [42,22]. As part of the EnKF methodology we first generate an ensemble of realizations for the parameter vector θ. For that purpose we shall restrict them to vary within the following ranges:

$$\alpha_f^0(\mathbf{x}) \in [0, 0.38], \quad k_w(\mathbf{x}) \in [1, 105], \quad T_v T^* \in [0.0005, 0.005] \text{ (1/Pa)}, \tag{12}$$

where T^* is the reference time (see Table 3). These intervals are set such that the resulting IFP lies within the range of 10-45 mmHg consistent with the results reported in [22], see Fig. 1B. Herein, $\alpha_f^0(\mathbf{x})$ and $k_w(\mathbf{x})$ are derived from Gaussian random fields [40]. The fields are generated with a spatial distribution based on a Gaussian random field with different practical ranges depending on the features of the parameter that is estimated [10,66]. For fibroblasts the practical range is 4 in the x and y direction. See Fig. 3 (panel B) for an example of $\alpha_f^0(\mathbf{x})$. For $k_w(\mathbf{x})$ we have developed an algorithm that creates thick bands within the primary tumor such that the tumor can be separated into several compartments [66]. Using this algorithm yields bands similar to the histological surveys shown in [22]. The vascular filtration coefficient T_v takes a constant value in a bandlike region but is randomly varying between each ensemble member.

3.2. What is the solution space spanned by the in silico model?

The different times that are involved in the following discussion are $T_0 = 0$, $T_1 = 25$, and $T_2 = 50$ (dimensionless) where $T_2 = 50$ amounts to 5.8 days. We define the total production of fluid $\overline{Q_v}$ from the leaky tumor vasculature to be

$$\overline{Q_v} = \int_\Omega T_v(\mathbf{x})(\widetilde{P}_v^* - P_w(\mathbf{x}, T_1))\, d\mathbf{x}, \tag{13}$$

while a representative expression for the IFP of one specific tumor is found from the average

$$\text{IFP} = \frac{1}{2}\Big(P_w(\mathbf{x}_1, T_1) + P_w(\mathbf{x}_2, T_1)\Big) \tag{14}$$

where $P_w(\mathbf{x}_i, T_1)$ $(i = 1, 2)$ refers to the numerical calculated P_w corresponding to position $\mathbf{x}_i$ at time T_1. Here $\mathbf{x}_1$ and $\mathbf{x}_2$ refer, respectively, to numerical grid block $(28, 28)$ and $(32, 32)$. These two positions are chosen randomly but such that it can mimic what is done experimentally in, e.g., [22]. The interstitial fluid pressure P_w is strongly correlated to the interstitial fluid velocity $\mathbf{u}_w$ through the tissue conductivity as represented through k_w. In turn, the fluid velocity field determines the distribution of chemical components C and H and, consequently, impacts the fibroblast and cancer cell migration patterns. Hence, IFP might play a vital role as to how the tumor develops over time. Associations have been reported between high IFP and metastatic propensity, i.e., dissemination of cancer cells to lymph nodes [19,42,23,22]. Based on the random fields generated for each member of the ensemble, we achieve a vast spectrum of different pressure profiles. This is illustrated in Fig. 4, which shows the mean IFP defined through (14) versus the fluid production from the vasculature expressed through $\overline{Q_v}$ given by (13), for an ensemble of 100 different realizations of $\theta = (\alpha_f^0(\mathbf{x}), k_w(\mathbf{x}), T_v)$.

In Fig. 5 we show results for 3 different samples (#16, #33, and #96) from the total ensemble of 100 samples. Rows A-C reflect the parameter vector $\theta_i = (\alpha_{fi}^0(\mathbf{x}), k_{wi}(\mathbf{x}), T_{vi})$ for $i = 16, 33, 96$. ECM bands represented through k_w for the three different members are shown in row A. The corresponding initial fibroblast distributions α_f^0 are shown in row B whereas T_v is illustrated in row C. Computed results at time T_1 are shown in rows D-F: cancer cell volume fraction $\alpha_c(\mathbf{x}, T_1)$ is shown in row D, IFP $P_w(\mathbf{x}, T_1)$ in row E, and IF velocity field $\mathbf{u}_w(\mathbf{x}, T_1)$ in row F. The location of the ECM bands largely determines the heterogeneous IFP and fluid velocity field. Since the tumor progression is driven by fluid-sensitive migration mechanisms, the invasive front is also correlated to IF.

In the following, we generate an ensemble of parameter vectors $\{\theta_i^{\text{true}}\}_{i=1}^{20}$ of 20 members with corresponding simulation results $Y_i^{\text{true}} = F(X, \theta_i^{\text{true}})$. We select a candidate from this ensemble as our "observed" data. The result in terms of θ_i^{true} corresponding to #7 is shown in Fig. 6 (left column, panel A-C). The question we now raise is: How can we obtain a best possible estimate of this parameter vector θ_7^{true} such that the error to the observed cancer cell observation $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ (left column, panel D), which is the only observed data we will make use of, is minimized and a best possible prediction of the distribution $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$ is obtained? Note that zero mean uncorrelated noise with standard deviation 0.03 is added to $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ before it is used to compute updated parameter vector θ (Appendix B). We also tested by choosing noise with

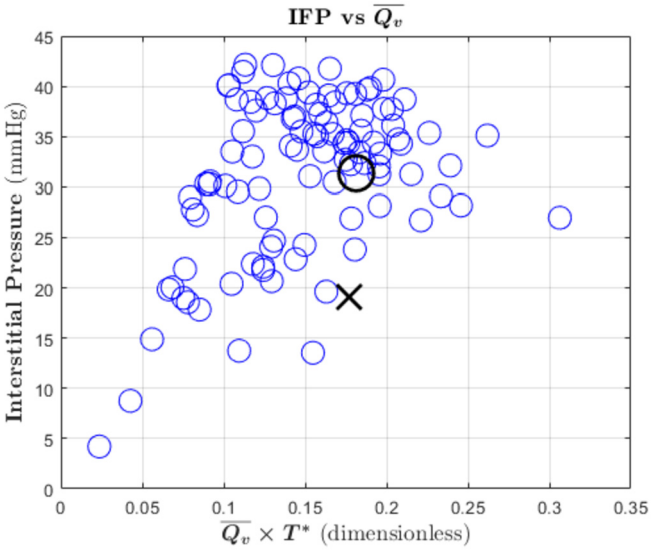

Fig. 4. Spanned solution space. Illustration of the spanned solution space measured in terms of $\overline{Q_v}$ vs IFP as expressed by (13) and (14), respectively, which corresponds to the a priori ensemble vector $\theta = \{\theta_i\}_{i=1}^{N_{\text{ens}}}$ with $N_{\text{ens}} = 100$. $\bigcirc = (\overline{Q_v}, \text{IFP})$ based on the mean initial parameter vector $\overline{\theta}$. $\times = (\overline{Q_v}, \text{IFP})$ based on θ^{true}.

standard deviation 0.05 and 0.07, respectively, which gave quite similar results. A main effect of higher level of noise in the observation data is that the uncertainty (standard deviation) in the updated parameter vector becomes higher. This is as expected.

3.3. Can the cancer cells reveal characteristics of the TME?

The starting point for searching for a good estimate of the parameter vector $\theta = (\alpha_f^0(\mathbf{x}), k_w(\mathbf{x}), T_v)$ is the a priori ensemble $\{\theta_i\}_{i=1}^{N_{\text{ens}}}$ generated above and illustrated in Fig. 4 where further details are shown for three of its members in Fig. 5. Each member of the ensemble has the same initial condition of tumor cell volume fraction $\alpha_c^0(\mathbf{x})$, seen in Fig. 3A, as we consider this to be observable data from an image of a tumor.

The ensemble members are a set of samples from the prior distribution. Each ensemble member i has a different parameter vector θ_i given by (11), i.e., $\theta_i = (\alpha_{fi}^0(\mathbf{x}), k_{wi}(\mathbf{x}), T_{vi})$ for $i = 1, \ldots, N_{\text{ens}}$ with $N_{\text{ens}} = 100$. Every member of the prior distribution θ_i is run through the computational model, i.e., we compute $Y_i = F(X, \theta_i)$ for $i = 1, \ldots, N_{\text{ens}}$, from initial condition at time $T_0 = 0$ until time $T_1 = 2.9$ days.

In particular, as explained in Appendix B, this allows us to extract $d_i = h(\theta_i) = \alpha_{ci}(\mathbf{x}, T_1)$ from which we obtain the ensemble mean $\overline{h(\theta)}$ defined by (26) to be used in combination with the observed data $d = \alpha_c^{\text{true}}(\mathbf{x}, T_1)$ (after including measurement uncertainty), as expressed by (29). Armed with the *updated* ensemble parameter vector $\theta_i^* = (\alpha_{fi}^{0,*}(\mathbf{x}), k_{wi}^*(\mathbf{x}), T_{vi}^*)$ with estimated fields obtained in accordance with (30), we utilize the corresponding *updated* ensemble mean $\overline{\theta}^* = (\overline{\alpha_f^{0,*}}(\mathbf{x}), \overline{k_w^*}(\mathbf{x}), \overline{T_v^*})$ from (29) to perform a forecast $Y^* = F(X, \overline{\theta}^*)$ at observation time T_1 where we have observation $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ as well as final time $T_2 = 5.8$ days where we have observable data $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$. In Fig. 6 we have visualized the different elements of the updated parameter vector $\overline{\theta}^*$ (middle column) which can be compared with the corresponding true parameter vector θ_7^{true} (left column). The difference is visualized in the right column of Fig. 6. In particular, by comparing the estimated and true parameter vector θ (panel A - panel C) we see that main trends are largely captured. This is also true for the observed and computed cancer cell volume fraction at time T_1 (panel D). However, the error plot (right column) illustrates that there are local variations in the estimated conductivity through $\log(\overline{k_w^*})(\mathbf{x}))$ (row C) which are not seen in the true. This error, in turn, gives rise to some voxel-to-voxel mismatch when comparing initial fibroblast distribution (row A) and cancer cell distribution (row D) up to the order of 0.1. Herein we have used, as described in Appendix B, the iterative approach called ensemble smoother - multiple data assimilation (ES-MDA) to better handle non-linearities in the predicted observed values $h(\theta)$ with $N = 4$ iterations.

What about the prediction at time T_2? The results of the predicted behavior at time T_2 are shown in Fig. 7 (middle column) and should be assessed in light of true data (left column). The difference is again visualized in the right column. Main observations from Fig. 6 and Fig. 7 are: (i) The essential trends of the true ECM band structure through k_w^{true} in Fig. 6 (row C, left) is captured by the estimated $\overline{k_w^*}$ in the same row, middle panel. Similarly, for the vasculature conductivity T_v^{true} and $\overline{T_v^*}$ in row B. (ii) Consequently, the predicted and true heterogeneous IFP in Fig. 7 (row C) and IF velocity field (row D), respectively, coincide fairly well. (iii) Finally, the predicted and true cancer cell migration pattern (row A) and fibroblast

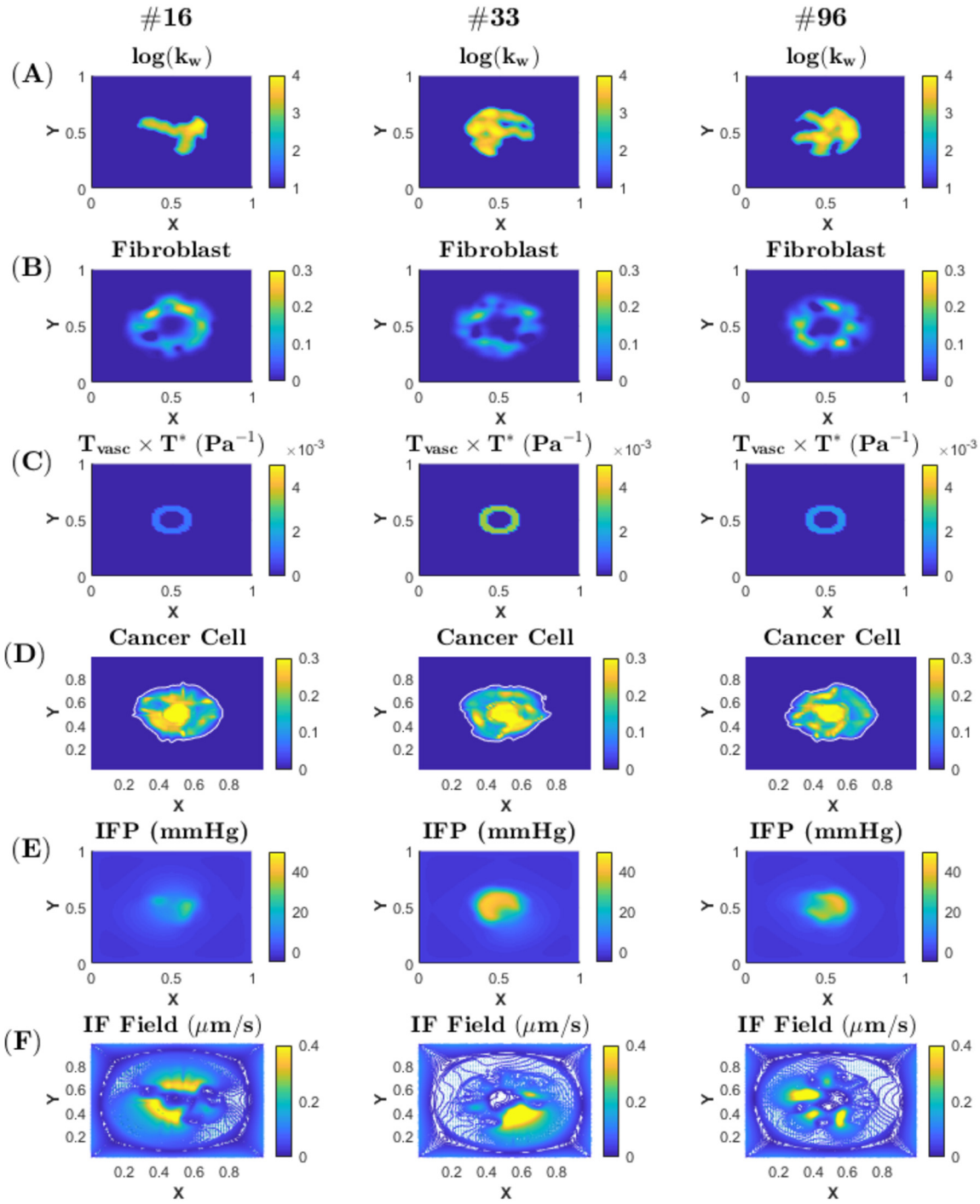

Fig. 5. Illustration of 3 different samples θ_i with $i = 16, 33, 96$ from the ensemble of $N_{\text{ens}} = 100$ and corresponding computed in silico tumors at time T_1. **Row A**: ECM bands through $k_{wi}(\mathbf{x})$; **Row B**: initial fibroblast $\alpha^0_{fi}(\mathbf{x})$; **Row C**: vascular conductivity $T_{vi}(\mathbf{x})$; **Row D**: computed cancer cell volume fraction $\alpha_{ci}(\mathbf{x}, T_1)$; **Row E**: computed IFP $P_{wi}(\mathbf{x}, T_1)$; **Row F**: computed fluid flow velocity $\mathbf{u}_{wi}(\mathbf{x}, T_1)$.

migration pattern (row B), respectively, in Fig. 7 shows good agreement. The right column of Fig. 7 shows that the difference in velocity field (row D) is responsible for some voxel-to-voxel mismatch in the position of the migrated fibroblasts (row B) which also gives rise to similar mismatch in the cancer cell distribution (row A).

We seek more insight into the 4-step iterative estimation procedure we have employed by plotting the difference between the true parameter vector θ^{true} and the updated parameter vector θ^*_i where $i = 1, \ldots, N_{\text{ens}}$. Fig. 8 illustrates the estimation error for the three fields that are being estimated: the fibroblast volume fraction α^0_f in panel A, the logarithm

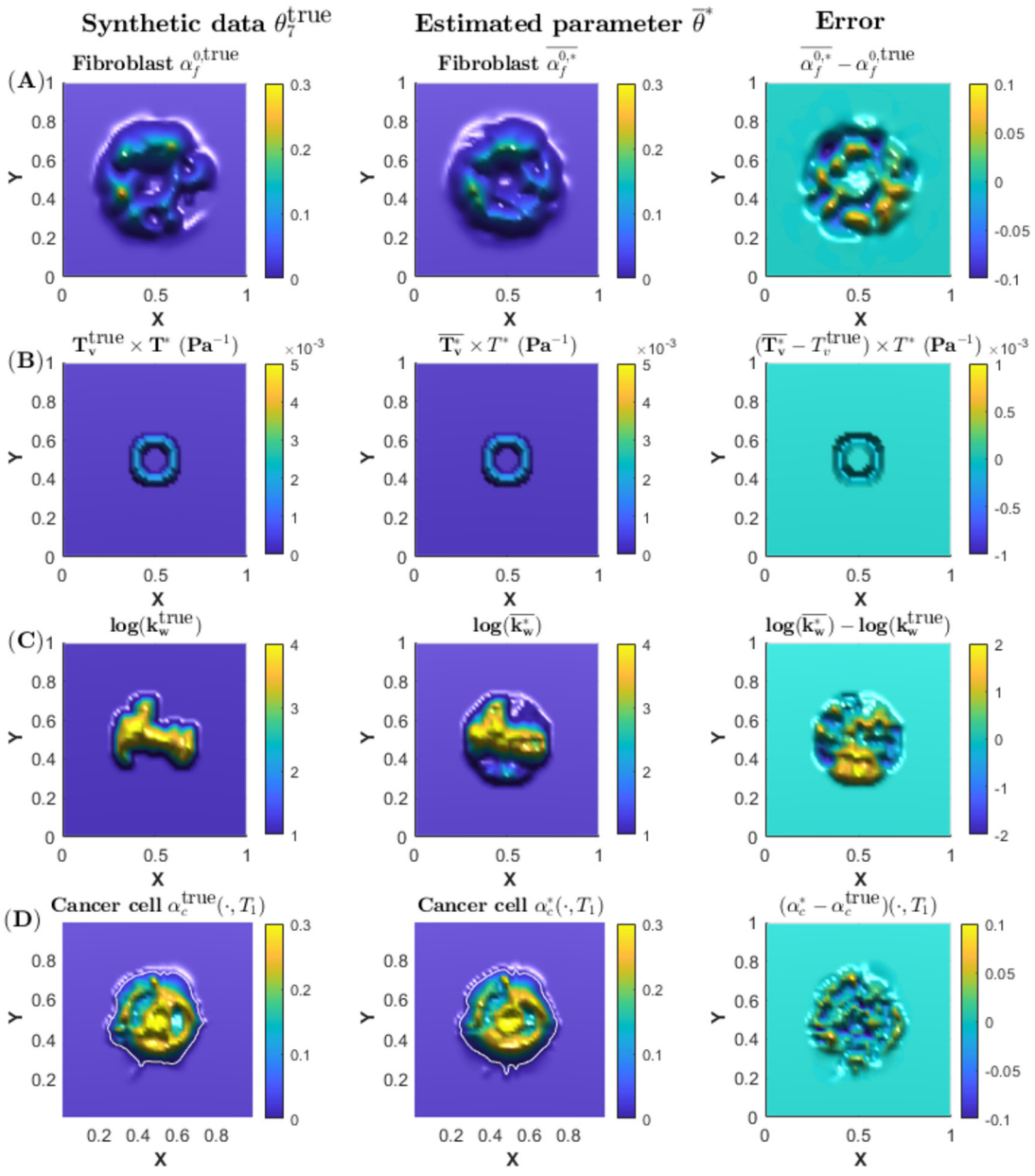

Fig. 6. Visualization of synthetic *true* parameter vector θ_7^{true} corresponding to case #7 (left column) and *estimated* parameter vector $\overline{\theta}^*$ (middle column) and the difference (right column). **Row A**. Initial fibroblast volume fraction: True $\alpha_f^{0,\text{true}}(\mathbf{x})$ (left) and estimated $\overline{\alpha_f^{0,*}}(\mathbf{x})$ (middle). **Row B**. Vascular conductivity: True $T_v^{\text{true}}(\mathbf{x})$ (left) and estimated $\overline{T_v^*}(\mathbf{x})$ (middle). **Row C**. True conductivity through $\log(k_w^{\text{true}}(\mathbf{x}))$ (left) and estimated $\log(\overline{k_w^*})(\mathbf{x}))$ (middle). **Row D**. True cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ (left) based on θ_7^{true} and computed $\alpha_c^*(\mathbf{x}, T_1)$ (middle) based on $\overline{\theta}^*$. We only use of $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ when we update the parameter vector $\overline{\theta}^*$ by means of EnKF.

of the inverse tissue conductivity k_w (panel B) and the vascular filtration constant T_v in panel C. The estimation error is calculated using root mean square error (RMSE) (i.e., L^2-norm) between the true parameter θ^{true} and the updated ensemble θ_i^* of each respective field:

$$RMSE_i = \left\| \theta^{\text{true}} - \theta_i^* \right\|_{L^2}, \qquad i = 1, \ldots, N_{\text{ens}} \tag{15}$$

where θ here refers to α_f^0, k_w, and T_v, respectively. The estimation error is monotonically decreasing as a function of iterations. The iterative approach relies on ES-MDA which increases the measurement uncertainty in each step, leading to a better linear approximation of the predicted observed values through many steps compared to only one step (Appendix B).

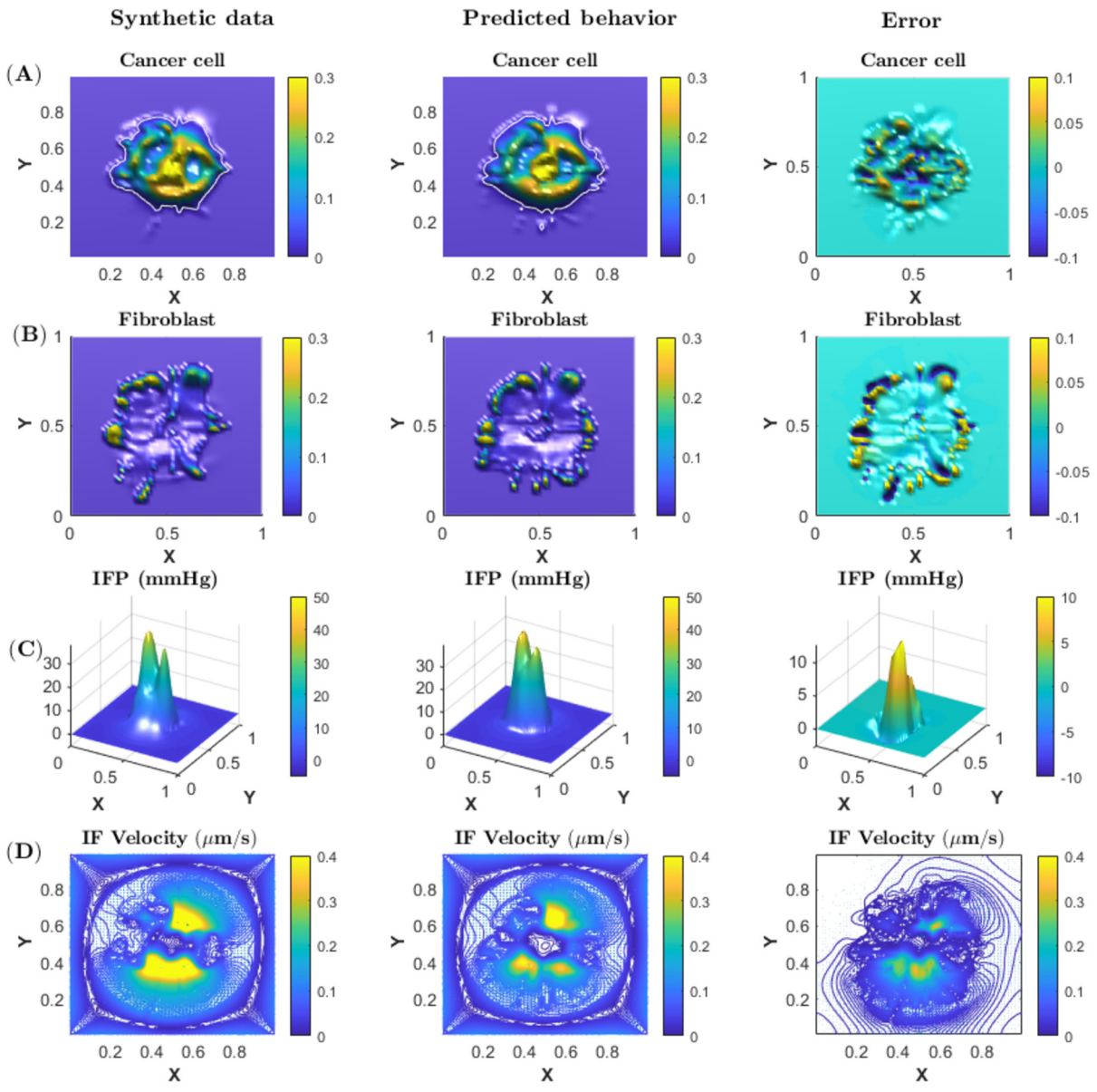

Fig. 7. The simulated result at time T_2 based on the true parameter vector θ_7^{true} (left column) and on updated parameter vector $\overline{\theta}^*$ (middle column), and the difference (right column). **Row A**. True cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_c^*(\mathbf{x}, T_2)$ (right). **Row B**. True fibroblast volume fraction $\alpha_f^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_f^*(\mathbf{x}, T_2)$ (right). **Row C**. True interstitial fluid pressure $P_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $P_w^*(\mathbf{x}, T_2)$ (right). **Row D**. True interstitial fluid velocity $\mathbf{u}_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\mathbf{u}_w^*(\mathbf{x}, T_2)$ (right).

The objective function O measures the mismatch between the observed data d and model prediction h with updated ensemble parameter vector and is written as

$$O_i = (h(\theta_i^*) - d)^T \frac{1}{N} C_D^{-1} (h(\theta_i^*) - d), \qquad i = 1, \ldots, N_{\text{ens}} \tag{16}$$

where $d = \alpha_c^{\text{true}}(\mathbf{x}, T_1)$ is the observed data and $h(\theta_i^*)$ is the model prediction $\alpha_{ci}^*(\mathbf{x}, T_1)$ when using the updated ensemble parameter vector θ_i^*. Moreover, N is the number of iterations and C_D is the covariance matrix of the measurement noise. The objective function shown in Fig. 8 (panel D) shows that the mismatch has a steep decline after one iteration then starts to even out.

In order to give an impression of the effect of using the updated parameter vector θ_i^*, we compute $Y_i = F(X, \theta_i^*)$ for $i = 16, 33, 96$. The results are presented in Fig. 9 and should be compared with the results in Fig. 5 which are based on the a priori parameter vector θ_i. The main observation is that all three cases show clear similarity to the behavior of the true data shown in Fig. 6 (left column) and Fig. 7 (left column). Further information about the behavior of the ensemble

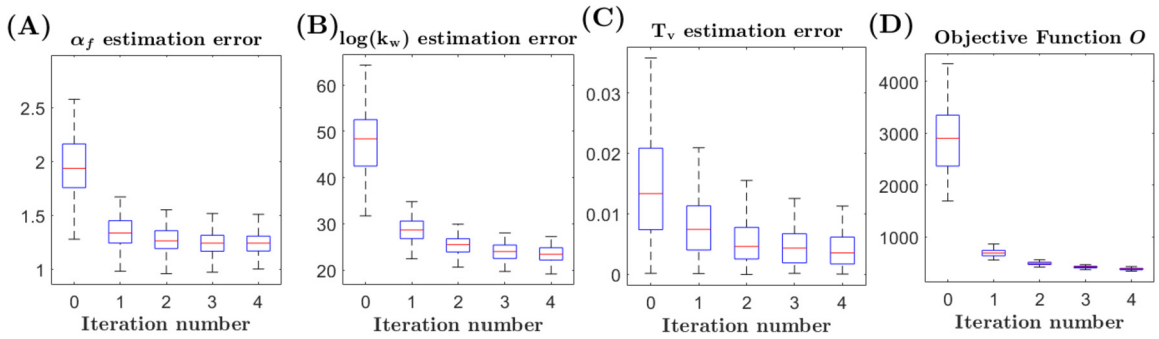

Fig. 8. Estimation error and objective function showing data mismatch: Estimation error of the three estimated fields, calculated through the $RMSE_i$ (15): **(A)** the initial fibroblast volume fraction α_f^0, **(B)** logarithm of the inverse tissue conductivity k_w and **(C)** the vascular filtration constant T_v. **(D)** Data mismatch as measured by the objective function O_i (16). The red horizontal lines show the median, the top and bottom blue edges indicate the 25th and 75th percentile, respectively, for the ensemble. The whiskers extend to the most extreme data points.

$Y_i = F(X, \theta_i^*)$ can be extracted by making a plot similar to the one in Fig. 4 in terms of $\overline{Q_v}$ and IFP but now based on the updated parameter vector θ_i^*. Results are shown in Fig. 10 and illustrates how the updated parameters and predictions as expressed by ($\overline{Q_v}$, IFP)-points tend to stay in a smaller region that includes "true" behavior (cross) whereas the behavior based on the a priori parameter vector shown in Fig. 4 is much more spread out.

This example confirms that since the tumor progression is driven by cancer cells and fibroblasts that depend on fluid-sensitive migration mechanisms (chemotaxis), information about TME is implicitly present and affects the change in the cancer cell volume fraction distribution over time. This information can be extracted from only observing the change in α_c from time T_0 till time T_1 and used to populate the in silico model with an updated parameter vector $\overline{\theta^*}$ such that a good prediction at the later time T_2 is obtained.

3.4. What is the effect of adding proliferation to the in silico model?

A natural question is what happens if tumor progression also involves growth and death of cancer cells. Is it so that these changes in the cancer cell volume fraction can "mask" for the role played by the TME characteristics that affects the cancer cell and fibroblast migration? We include a logistic growth term of the form [70,71]

$$S_c = \alpha_c(\mathbf{x}, t) K_p \Big(1 - \frac{\alpha_c(\mathbf{x}, t)}{\theta_p(\mathbf{x})}\Big) = \alpha_c(\mathbf{x}, t) K_p \Big(1 - \alpha_c(\mathbf{x}, t)\beta_p(\mathbf{x})\Big), \qquad \beta_p = \frac{1}{\theta_p} \tag{17}$$

where K_p is the growth rate and θ_p is the tumor carrying capacity (i.e., the maximum volume fraction of cancer cells that a region can support). The parameter K_p (proliferation rate) is assumed to be constant and vary within the interval $[0, 1] \times 10^{-5}$ s^{-1} from one tumor to another. The quantity $\theta_p(\mathbf{x})$ is here set to vary within the interval $[\frac{1}{6}, \frac{1}{2}]$ (i.e., $\beta_p \in [2, 6]$) and as a Gaussian random field. This gives room for interesting and more realistic tumor progression behavior within the simulated time period $[0, T_2]$. Heterogeneity in cancer cell density might be understood as a result of spatially varying carrying capacity which may occur due to physical limitations (e.g. decrease in available space to grow) and environmental limitations (e.g. poorly perfused, low nutrient concentration) [25,24]. In order to account for this situation we extend the parameter vector and consider θ_E given by

$$\theta_E = \Big(\alpha_f^0(\mathbf{x}), k_w(\mathbf{x}), T_v, K_p, \beta_p(\mathbf{x})\Big). \tag{18}$$

Similarly, as for the previous case we use sample #7 to represent synthetic data but where we now have augmented the parameter vector by using θ_E defined by (18) instead of (11). Fig. 11 shows the true data (left column) through θ^{true} versus the estimated parameter vector $\overline{\theta^*}$ (middle column). A quantification of the difference is shown in the right column of Fig. 11. In Fig. 11 we find the new rows D and E that represent, respectively, K_p and β_p. The main observation is that the estimated field $\overline{k_w^*}$ shown in row C (middle panel) is more diffusive and misses some of the characteristic structure of the true counterpart (left panel), see plot in right panel for a more precise quantification of the difference. On the other hand, the estimation of the growth related parameter K_p in row D is fairly good where the difference in value is around 0.1 (see right panel) and the essential spatial heterogeneities of β_p shown i row E is largely captured by the estimated parameter $\overline{\beta_p^*}$. This is also confirmed by the results in row F showing the observed cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ versus the computed $\alpha_c^*(\mathbf{x}, T_1)$ based on the updated parameter vector $\overline{\theta^*}$. The periphery of the tumor in row F is quite similar, as well as the distribution within the tumor, for the two cases. The error plot in the right column shows that there are local differences of the order of 0.1 which is similar to the case without proliferation. We also see from row F that the cancer cell volume fraction is considerably higher in regions where β_p (row E) takes lower values and therefore promotes proliferation.

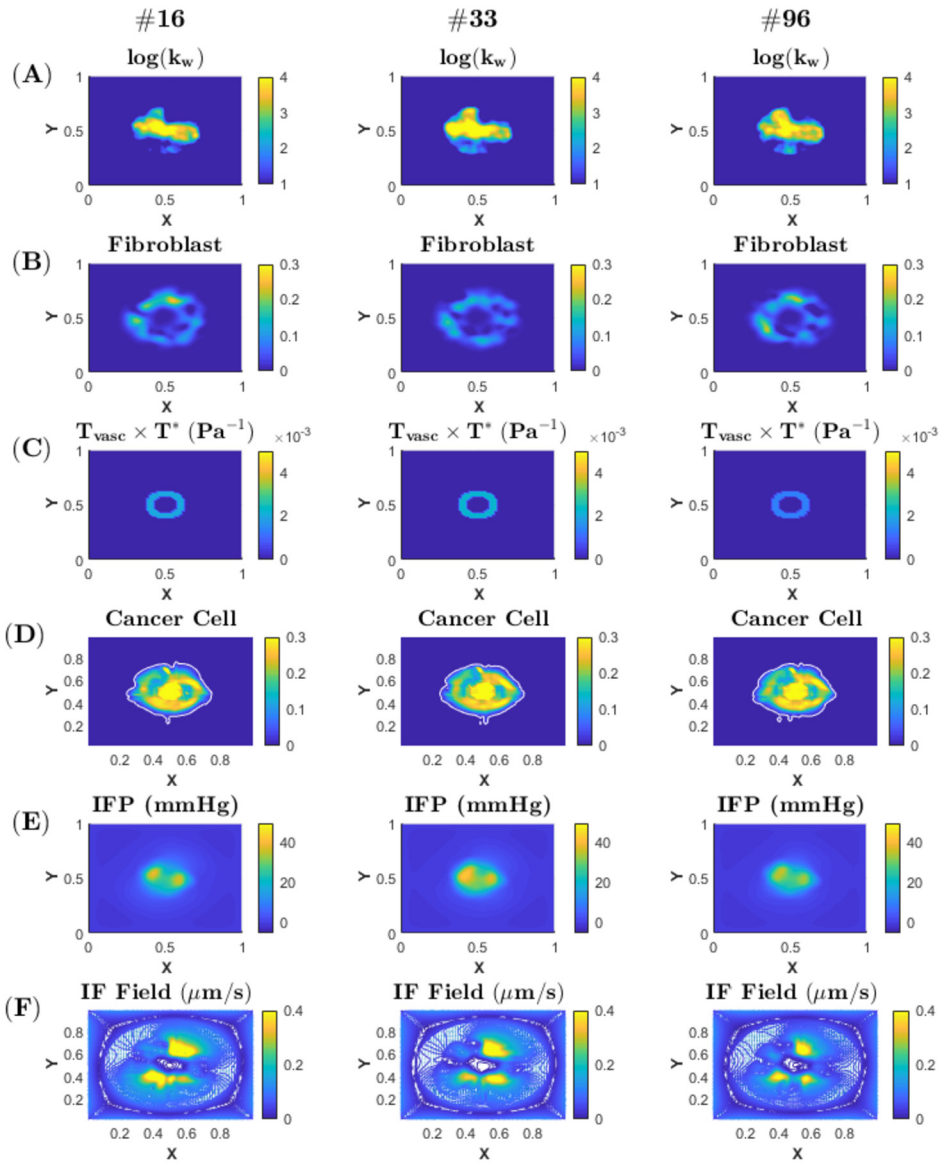

Fig. 9. Illustration of 3 different samples from Fig. 5 based on the updated parameter vector θ_i^* for $i = 16, 33, 96$. **Row A**: ECM bands through $k_{wi}^*(\mathbf{x})$; **Row B**: initial fibroblast $\alpha_{fi}^{0,*}(\mathbf{x})$; **Row C**: vascular conductivity $T_{vi}^*(\mathbf{x})$; **Row D**: computed cancer cell volume fraction $\alpha_{c,i}^*(\mathbf{x}, T_1)$; **Row E**: computed IFP $P_{wi}^*(\mathbf{x}, T_1)$; **Row F**: computed fluid flow velocity $\mathbf{u}_{wi}^*(\mathbf{x}, T_1)$.

On the other hand, the cell volume fraction is low in regions where β_p (row E) takes high values and indicate a low tumor carrying capacity.

What about the predicted tumor status at the later time T_2 based on the updated parameter vector $\overline{\theta^*}$? In Fig. 12 we find a comparison of true tumor status at time T_2 versus the predicted based on $\overline{\theta^*}$, respectively, in the left column and middle column, whereas the difference is shown in the right column. Row A shows that we still can obtain an estimate of the observed cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$ (left) through the predicted $\alpha_c^*(\mathbf{x}, T_2)$ (middle) on a level similar to the case without proliferation. In particular, the error associated with $\alpha_c^*(\mathbf{x}, T_2)$ is similar (~ 0.1). This is partly due to the fact that the main trends of the new quantities K_p (row D) and $\beta_p(\mathbf{x})$ (row E) in Fig. 11 are largely captured.

On the other hand, we see that estimated IFP (row C, middle panel) in Fig. 12 is higher than the true IFP (left panel). In fact, from the error plot in the right panel we see that there are local differences in the range of 20 mmHg. This is related to the difference in true and estimated fluid velocity field shown in row D, which is more significant for this example, as can be seen from a comparison with the error plot in Fig. 7 (row D, right panel). The inaccuracies in IFP and fluid velocity field are a consequence of the fact that the estimated $\overline{k_w^*}$ (row C, right panel) in Fig. 11 is more diffusive and is lacking some of the sharp transitions of the true k_w^{true} in left panel.

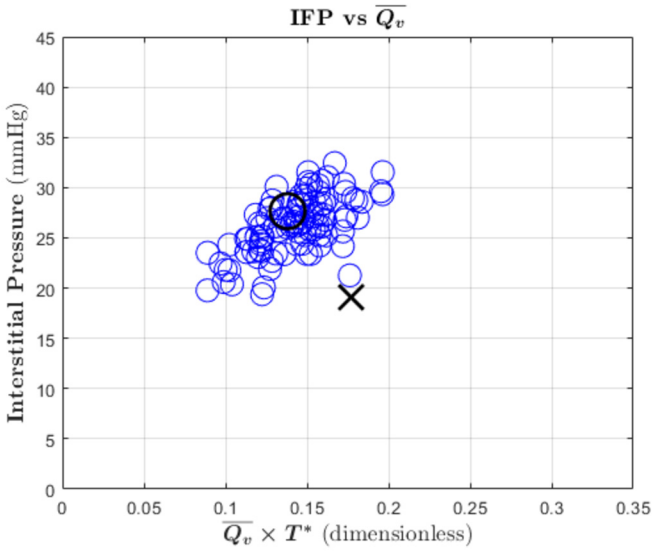

Fig. 10. Illustration of the spanned solution space based on the observation $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$. The updated ensemble parameter vector $\theta^* = \{\theta_i^*\}_{i=1}^{N_{\text{ens}}}$ with $N_{\text{ens}} = 100$ has been obtained through 4 iterations and used to generate updated simulations $Y_i^* = F(X, \theta_i^*)$ at time T_1. $\bigcirc = (\overline{Q_v}, \text{IFP})$ based on $\overline{\theta^*}$. $\times = (\overline{Q_v}, \text{IFP})$ based on θ^{true}.

This suggests that if the tumor progression is more dominated by proliferation the method we use may give a poorer approximation of the true parameter vector θ^{true} despite the fact that the prediction of the cancer cell volume fraction α_c is good. This reflects a situation where the information involved in the objective function (16) is insufficient to discern properly between growth and go behavior. All in all, the current version of the estimation method seems to handle fairly well inclusion of growth as long as it affects the tumor progression on a level as shown in Fig. 11 over the time period under consideration.

3.5. Does the finding depend on the cancer cell phenotype?

We recall that we have used a fixed set of parameters to represent the cancer cell and fibroblast phenotype. So what happens if we change these parameters? Is the calibration method sensitive to the cell phenotype? We focus on the parameters Λ_{C1}, Λ_{H1}, and I_{cf} mentioned in Section 2.8. As before, we assume that these parameters are known when we generate synthetic data as well as carry out the estimation procedure. We also test the EnKF approach for another data set (synthetic) to verify robustness with respect to variation in observation data.

Phenotype I. We consider as observed data sample #7 (i.e., we use θ_7^{true} from the ensemble of the 20 in silico TME characteristics) to generate the synthetic observed tumor statues $d = \alpha_c^{\text{true}}(\mathbf{x}, T_1)$, however, where we have changed cancer cell and fibroblast migration parameters as follows: Λ_{C1}, Λ_{H1}, and I_{cf} are all multiplied by 2 compared to default values (see Table 4 and Table 5). We ignore proliferation in this example. This will rise to a stronger outwardly directed chemotaxis-driven migration of fibroblasts and cancer cells through the higher values of Λ_{C1} and Λ_{H1} combined with a stronger cell-fibroblast interaction through increased I_{cf} which promotes collective migration. Based on this new $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ we use the EnKF with 4 iterations to generate an updated parameter vector $\overline{\theta^*}$ which in turn is used to predict the tumor progression in the time period $[0, T_2]$ In Fig. 13 we show the estimation error behavior of the parameter vector $\theta = \{\theta_i\}_{i=1}^{N_{\text{ens}}}$ as compared to the true θ_7^{true} (panel A - panel C). We see a convergence behavior similar to the first example, except from T_v (panel C) which does not show complete decreasing behavior. Then, in Fig. 14 the true and predicted behavior is shown. Comparing with the corresponding results in Fig. 7 reveals that the quality of this prediction is quite similar to the default case. The error plots shown in the right column of Fig. 7 and Fig. 14, respectively, also confirm this. This test suggests that the estimation method is robust with respect to changes in the parameters that characterize the chemotaxis-driven migration as well as cancer cell-fibroblast interaction.

Phenotype II. As a second test, we change the interaction coefficient I_{cf} to be half of the default value while Λ_{C1} and Λ_{H1} are twice as large as default values. This means that the cell-fibroblast coupling is weakened which gives more room for individual behavior. At the same time we also want to check sensitivity to the choice of observed (synthetic) data by choosing another sample. We choose sample #12 from the ensemble of 20 "true" in silico tumors with $S_c = 0$ (no proliferation). The "true" behavior is shown in Fig. 15 (left column) at time T_2 with the predicted behavior shown in the middle column and the difference in the right column. The "true" behavior shows a situation with stronger spreading of

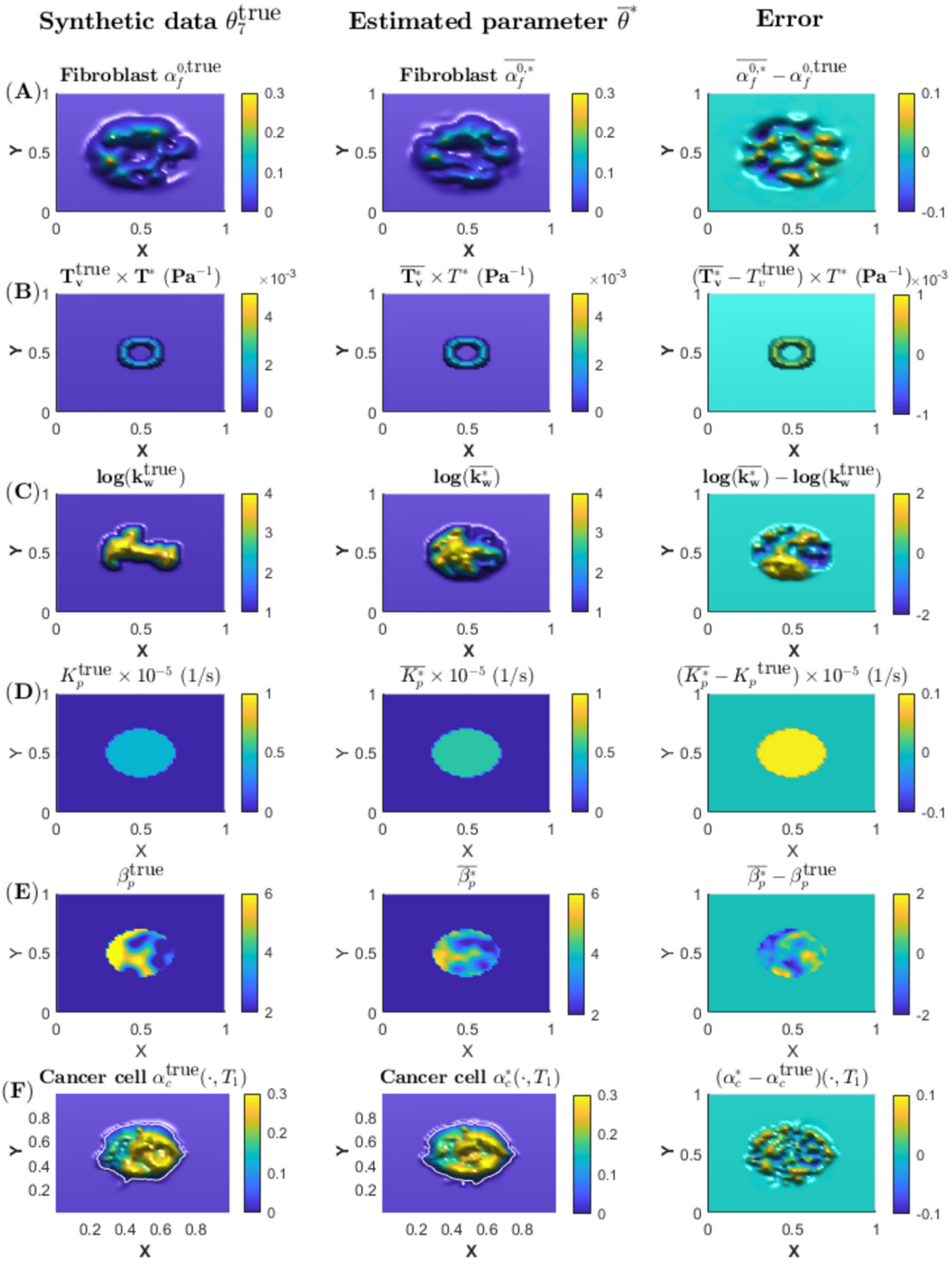

Fig. 11. Visualization of synthetic *true* parameter vector θ_7^{true} (left column), *estimated* parameter vector $\overline{\theta}^*$ (middle column), and the difference (right column) for the case with inclusion of proliferation through the growth term (17). **Row A**. Initial fibroblast volume fraction: True $\alpha_f^{0,\text{true}}(\mathbf{x})$ (left) and estimated $\overline{\alpha_f^{0,*}}(\mathbf{x})$ (middle). **Row B**. Vascular conductivity: True $T_v^{\text{true}}(\mathbf{x})$ (left) and estimated $\overline{T_v^*}(\mathbf{x})$ (middle). **Row C**. True conductivity through $\log(k_w^{\text{true}}(\mathbf{x}))$ (left) and estimated $\log(\overline{k_w^*})(\mathbf{x}))$ (middle). **Row D**. True growth factor K_p^{true} (left) and estimated $\overline{K_p^*}$ (middle). **Row E**. True carrying capacity through β_p^{true} (left) and estimated $\overline{\beta_p^*}$ (middle). **Row F**. True cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ (left) based on θ_7^{true} and computed $\alpha_c^*(\mathbf{x}, T_1)$ (middle) based on $\overline{\theta}^*$. We only make use of $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ when we search for an estimate of the parameter vector $\overline{\theta}^*$ by means of the EnKF approach.

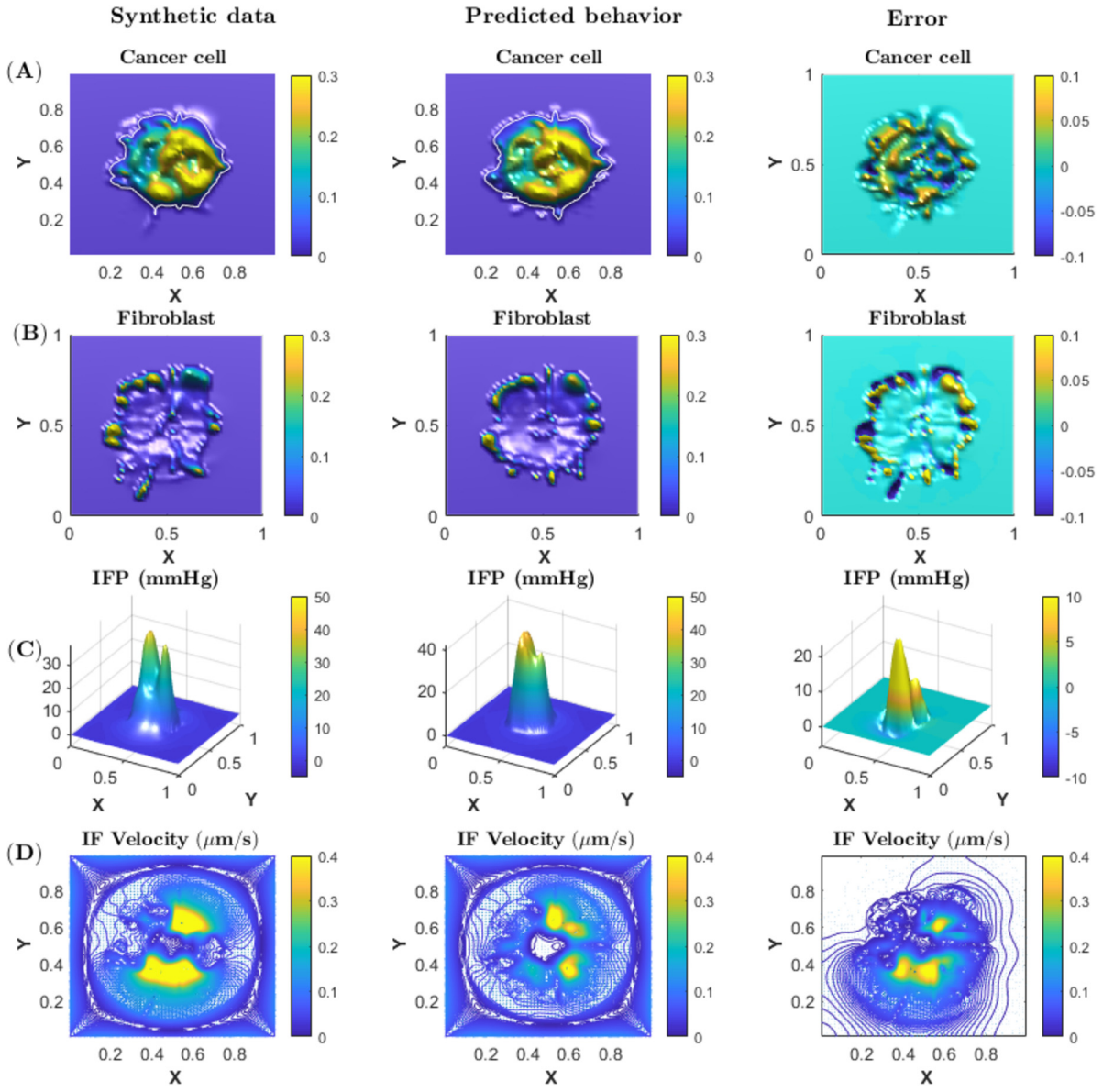

Fig. 12. Simulated result for the case with proliferation at time T_2 based on the true parameter vector θ_7^{true} (left column) and on updated parameter vector $\overline{\theta}^*$ (middle column). The difference is shown in the right column. **Row A**. True cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_c^*(\mathbf{x}, T_2)$ (middle). **Row B**. True fibroblast volume fraction $\alpha_f^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_f^*(\mathbf{x}, T_2)$ (middle). **Row C**. True interstitial fluid pressure $P_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $P_w^*(\mathbf{x}, T_2)$ (middle). **Row D**. True interstitial fluid velocity $\mathbf{u}_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\mathbf{u}_w^*(\mathbf{x}, T_2)$ (middle).

isolated clusters of fibroblasts (row B, left panel) and a more invasive, heterogeneous tumor margin with several isolated islands shown in row A, left panel. The corresponding IFP is shown in row C (left panel) and IF in row D (left panel). Clearly, it reflects a heterogeneous situation with high values of k_w in some regions which gives rise to a heterogeneous fluid velocity field as shown in row D where fluid flows in the direction of least resistance. The predicted behavior shown in the middle column at time T_2 reflects an overall match with the true data behavior in terms of main trends. The error plots in the right column reflect that there are local differences between true and predicted behavior on a similar level as for the base case. In particular, the more invasive and aggressive cancer cell migration pattern is largely captured (row A), as well as the fibroblast distribution (row B), which must be understood in view of the corresponding fit with the TME characteristics as expressed by IFP (row C) and IF velocity field (row D). More information about the parameter estimation error and data mismatch is given in Fig. 16 and behaves similarly to the previous case. This test case indicates robustness with respect to different type of observed (synthetic) data in terms of cancer cell spreading and corresponding TME characteristics, as well as robustness with respect to different phenotypes.

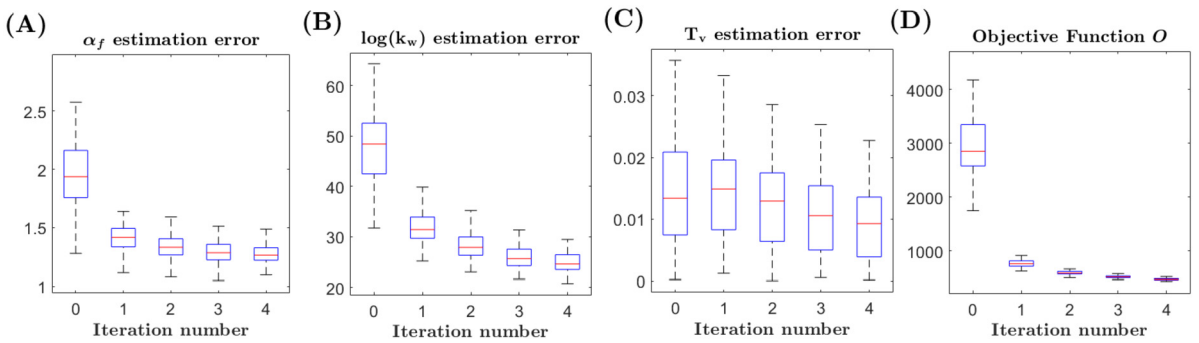

Fig. 13. Estimation error and data mismatch for phenotype I. Estimation error of the three estimated fields, calculated through $RMSE_i$ (15): **(A)** the fibroblast volume fraction α_f^0, **(B)** logarithm of the inverse tissue conductivity k_w, and **(C)** the vascular filtration constant T_v. **(D)** Data mismatch as measured through the objective function O_i (16).

4. Discussion

Experimental based research has revealed that cancer cells as well as fibroblasts make use of fluid sensitive migration mechanisms through autologous chemotaxis [56,55,49,46,17]. Patient data has shown that the greater the number of metastatic lymph nodes, the higher the levels of CCR7 expression [21]. Conversely, tumors lacking CCR7 expression have shown only low rates of lymph vessel invasion [57]. At the same time, other research groups have through preclinical and clinical studies searched for associations between metastatic propensity (lymph node metastasis) and various characteristics of the tumor microenvironment, features pertaining to the ECM structure and the aberrant blood vasculature. Both affect the drainage of fluid from the leaky vasculature and the collecting peritumoral lymphatics, which determines the tumor IFP and IF velocity field. A drawback is that such information is extracted through histological analysis and/or invasive methods. Several preclinical studies have found an association between lymph node metastases and high IFP [23,3,22]. High-tumor IFP was associated with poor disease-free survival independent of conventional prognostic factors, such as tumor size, stage, and lymph node status [74,19]. Taken together, this suggests that fluid-sensitive migration mechanisms in fact might be at work in real-life tumors and possibly play a role in tumor progression and invasive behavior, as well as metastatic dissemination [17,76]. On the other hand, research groups have demonstrated how MRI data which gives information about the tumor status in terms of cancer cell distribution during growth [25,24] as well as during anticancer treatment [29,30] at, e.g., two or several different times, can be used to calibrate continuum based models for tumor growth to fit the observed data, and then used to predict the outcome of the tumor progression/regression [70,71,29,24,30].

Motivated by this research, we have raised the question: Subject to the condition that tumor progression is driven by fluid-sensitive migration mechanisms, can the cancer cells inform us about TME characteristics? More precisely, given only information about the cancer cell distribution at an initial time T_0 and a later time T_1 (e.g., through MRI), can we detect information about TME like tissue conductivity, IFP and IF as well as estimate of the effective vascular conductivity involved in Starlings law? These are quantities that cannot easily be obtained by direct measurements. The tumor microenvironment is an important factor looking at the effectiveness of a drug [60,58]. Having more information available may help making better choices with regard to schemes for efficient drug delivery. Information about the presence of fibroblasts may also be useful in a treatment setting, however, such information seems not possible yet to extract without invasive methods.

We explore this question by means of a multiphase model (1)–(5) that accounts for fluid-sensitive migration mechanisms that have been reported from experimental work [56,55,36]. These mechanisms are indirectly fluid flow dependent, considering that the underlying chemical component of chemotaxis and growth factor are advected by the fluid flow. The model has been used previously to shed light on preclinical data which search for associations between TME characteristics and metastatic propensity [66,22]. The computational model is here used in combination with a data assimilation technique that uses only one observation of the cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_1)$ of a tumor, in addition to a known initial state, to estimate initial fibroblasts volume fraction $\alpha_f^0(\mathbf{x})$, inverse tissue conductivity $k_w(\mathbf{x})$ and vascular filtration constant T_v. These are field parameters that characterize important aspects of the tumor microenvironment that most likely play an important role in the evolution of the tumor as well as resistance to anticancer drug [42,23,28,58,3,22]. The EnKF based approaches have been applied in a large number of applications over the last couple of decades [1,13–15,27,41,64], but examples within cancer research are to the best of our knowledge scarce (an exemption is [35]). However, the EnKF based approaches are in general easy to implement and test as they can be implemented non-intrusively with respect to the simulation model. We use synthetic data obtained by simulation of the PDE model (1)–(5) which has been populated with a stochastic-generated input parameter vector $\theta = (\alpha_f^0, k_w, T_v)$ which characterizes TME. Main findings are:

(i) Since the tumor progression is driven by cancer cells and fibroblasts that depend on fluid-sensitive migration mechanisms (chemotaxis), information about TME is implicitly present and affects the change in the cancer cell volume fraction distribution over time. This information can be extracted by means of the EnKF approach by observing the change in α_c

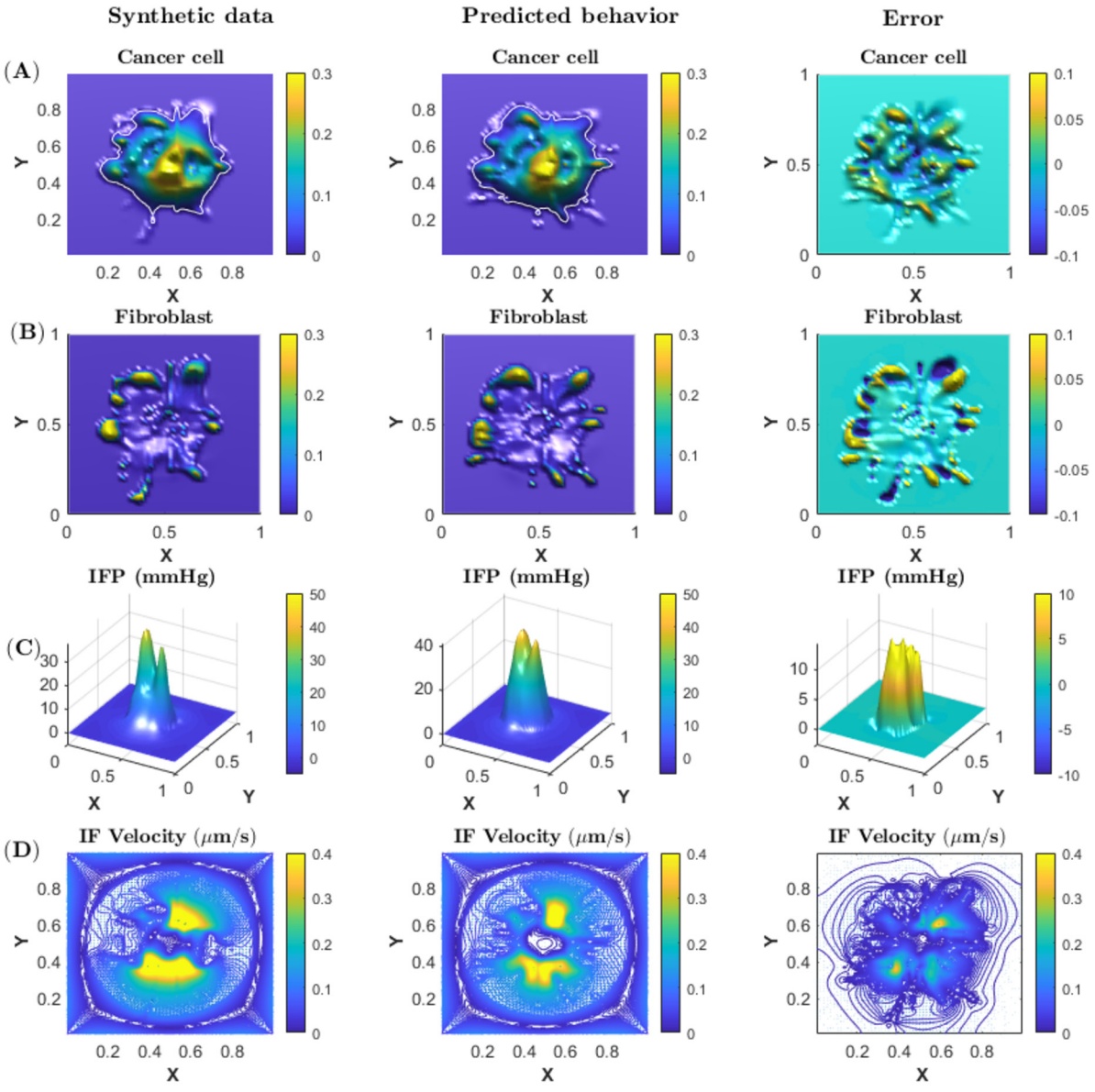

Fig. 14. Phenotype I: The simulated result at time T_2 based on the true parameter vector θ_7^{true} (left column) and on updated parameter vector $\overline{\theta}^*$ (middle column) and the difference (right column) where cancer cell and fibroblast parameters Λ_{C1}, Λ_{H1}, and I_{cf} have been multiplied by 2 (phenotype I). **Row A**. True cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_c^*(\mathbf{x}, T_2)$ (middle). **Row B**. True fibroblast volume fraction $\alpha_f^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_f^*(\mathbf{x}, T_2)$ (middle). **Row C**. True interstitial fluid pressure $P_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $P_w^*(\mathbf{x}, T_2)$ (middle). **Row D**. True interstitial fluid velocity $\mathbf{u}_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\mathbf{u}_w^*(\mathbf{x}, T_2)$ (middle).

from time T_0 till time T_1 and employed to compute an updated parameter vector $\overline{\theta^*}$. Populating the PDE model with this parameter vector results in a good prediction of the tumor status at the later time T_2.

(ii) By including growth of cancer cells through a constant growth factor K_p and a spatial varying carrying capacity term $\beta_p(\mathbf{x})$, estimation of the input parameter vector θ and corresponding prediction of tumor status at later time, gave slightly poorer result. If the effect of proliferation is dominating, the current version of our estimation method is more uncertain since these changes may mask the changes in cancer cell volume fraction due to fluid-sensitive migration. On the other hand, the estimation of the spatial varying carrying capacity term $\beta_p(\mathbf{x})$, which is responsible for spatial tumor heterogeneity, showed good result.

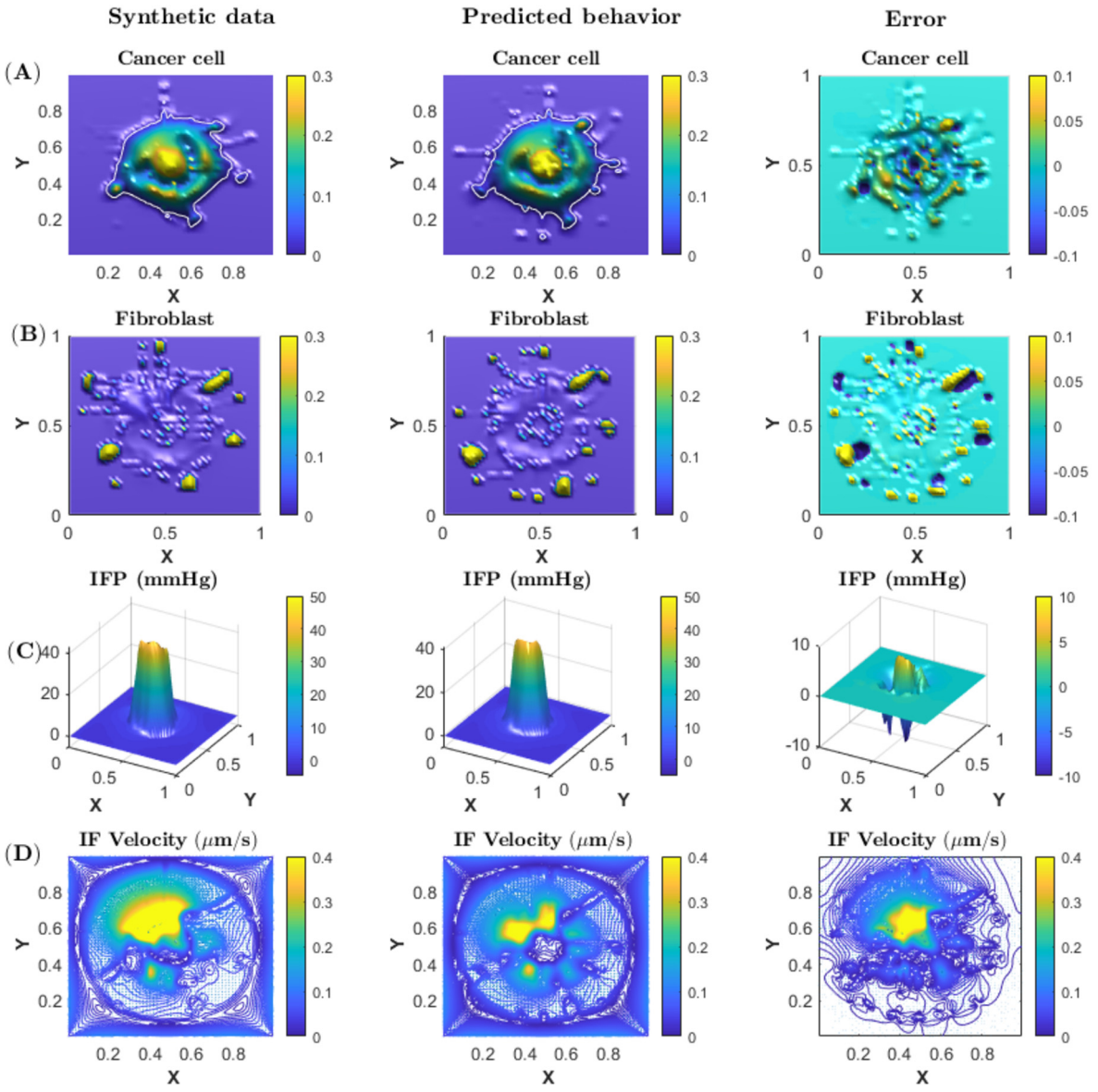

Fig. 15. Phenotype II: The simulated result at time T_2 based on the true parameter vector $\theta_{12}^{\text{true}}$ (left column) and on updated parameter vector $\overline{\theta}^*$ (middle column). The difference is shown in the right column, where cancer cell and fibroblast parameters Λ_{C1} and Λ_{H1} have been multiplied by 2 whereas and I_{cf} has been halved (phenotype II). **Row A**. True cancer cell volume fraction $\alpha_c^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_c^*(\mathbf{x}, T_2)$ (middle). **Row B**. True fibroblast volume fraction $\alpha_f^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\alpha_f^*(\mathbf{x}, T_2)$ (middle). **Row C**. True interstitial fluid pressure $P_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $P_w^*(\mathbf{x}, T_2)$ (middle). **Row D**. True interstitial fluid velocity $\mathbf{u}_w^{\text{true}}(\mathbf{x}, T_2)$ (left) and the predicted $\mathbf{u}_w^*(\mathbf{x}, T_2)$ (middle).

(iii) When the tumor progression is dominated by the fluid-sensitive migration mechanisms (which relate migration to tissue conductivity, IFP, and IF) the assimilation method appears to be robust with respect to variation in cancer cell and fibroblast migration/interaction characteristics.

This opens for the possibility that a more invasive tumor progression, while a challenge with regard to promoting lymph node metastasis, it might be possible to indirectly extract valuable quantitative information (e.g., interstitial fluid velocity field and fluid pressure, tissue conductivity reflecting ECM status, and effective vasculature conductivity), for which direct measurements may not be possible or impractical, that can be used for improved design of patient-specific treatment. A natural way to seek to deal more efficiently with the case when growth is more dominating might be to include more measurements in the objective (loss) function. For example, other researchers have used tumor metrics like maximum diameter of the primary tumor, blood volume, etc, and information about contrast agent delivery [24,30]. A computational model developed by Tektonidis et al. investigates the go or grow (GoG) hypothesis in gliomas [62]. The GoG hypothesis states that tumor cells have to either be migrating or proliferate. In other words, tumor cells are not able to migrate and

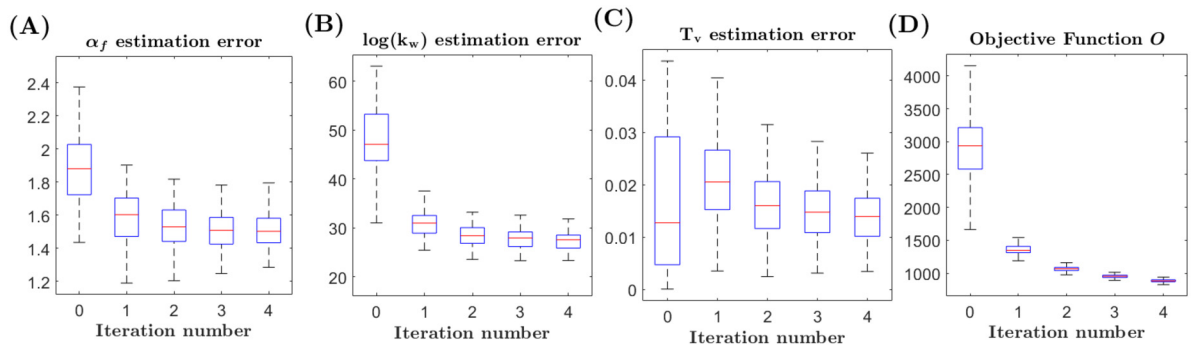

Fig. 16. Estimation error and data mismatch for phenotype II with synthetic data for case #12: Estimation error of the three estimated fields, calculated through $RMSE_i$ (15): **(A)** the fibroblast volume fraction α_f^0, **(B)** logarithm of the inverse tissue conductivity k_w, and **(C)** the vascular filtration constant T_v. **(D)** Data mismatch through the objective function O_i (16).

proliferate at the same time. Interestingly, the researchers found that a proliferative-motility dichotomy, cell-cell repulsion and density-dependent switching between a proliferative and motile phenotype were required to fit the experimental data. This observation might be exploited in the framework of the current manuscript. The model used in this work assumes that the primary tumor is already established and at a size as shown in Fig. 3. It would also be possible to adapt the model to a situation where we consider tumors grown "from seed". However, TME characteristics as expressed by T_v and k_w must then include a time-dependence to allow for evolution of these characteristics over time. This identification would also require access to MRI images sampled at different times over the time period of interest.

A main underlying assumption in our investigations is that tumor growth is dominated by cancer cell and fibroblasts migration where fluid flow directs the chemical agents. Future investigations that try to combine MRI based data with the model to test the ability of the model to fit preclinical and clinical data as explored by others [9,25,24] appear to be an interesting way forward. Another aspect and related issue is to explore extension of the parameter vector to include parameters pertaining to the cancer cell phenotype and not only parameters that characterize TME. An interesting and important direction for further work is to include delivery of an anti-cancer drug. How can the EnKF approach be used to calibrate the model when we study tumor progression under the impact of an anticancer drug [73,29,30]?

The current manuscript has exclusively dealt with synthetic data. An interesting question is how to possibly validate the method laid out in this work by other approaches. In a *clinical* setting it seems not realistic to do any validation without adding the effect of an anticancer drug. This approach has been explored by others [70,29,30]. It seems as an interesting and relevant task to explore the ensemble based approach when an anticancer drug is included. However, this will require some further investigations to see how the EnKF approach can be used to extract drug-related parameters, e.g, estimates of parameters pertaining to the delivery of the anticancer drug as well as cancer cell sensitivity to the drug.

In a *preclinical* setting (e.g., mouse model) one may envision to collect MRI images from experiments as reported in [3,22], motivated by other work, e.g. [24]. Based on such images one could use the method of the current manuscript to extract information about IFP, compartment structures through k_w and extracellular matrix components (dense versus sparse matrix), fluid flow through T_v, and desmoplastic structures (caused by fibroblasts α_f). This information could then be assessed by carrying out histological investigations as done in [3,22].

In this study we are limited to 2D simulation results where main purpose has been to demonstrate the concept: how to possibly extract information about TME from observing tumor progression extracted from MRI still images at different times. As indicated in Appendix C, the current version of the model seems possible to use on a scale relevant for preclinical data. However, for larger scales relevant for clinical data in 3D it seems necessary to rely on parallel computing methods where one takes advantage of the possibility to run many simulations in parallel. Note that the ensemble based approach of EnKF naturally allows for this as the multiphase model for each ensemble member is solved independently. This means that a significant reduction in computational time can be achieved by running the simulations on a cluster. There is also room for improvement of the numerical method itself by taking advantage of the fact that the IFP and IF velocity field involve solving a stationary problem which may not need to be updated with the time steps required by the cell migration [51].

Declaration of competing interest

Geir Nævdal: Methodology, Coding, Writing, Review & Editing
Steinar Evje: Methodology, Coding, Investigation, Writing, Review & Editing

Data availability

Data will be made available on request.

Appendix A. Input parameters

Table 3
Reference variables.

Variable	Description	Values
T^*	Reference time	10^4 s
L^*	Reference length	0.01 m
u^*	Reference velocity	10^{-6} m/s
D^*	Reference diffusion	10^{-8} m^2/s
P^*	Reference pressure	10^4 Pa
C^*	Reference chemokine density	10^{-4} kg/m^3
H^*	Reference TGF density	10^{-4} kg/m^3
C_M, H_M	Max chemokine, TGF density	$0.3C^*$, $0.5H^*$

Table 4
Functions involved in cell migration given by Eq. (6) and (7).

Function		Description	Values	Source
$\Delta P_{cw}(\alpha_c)$	=	$-\gamma_c \ln(\delta_c + [1-\alpha_c])$	(δ_c=0.01; γ_c=1 kPa)	[63,66]
$\Delta P_{fw}(\alpha_f)$	=	$\gamma_f \alpha_f^{\delta_f}$	(δ_f=25; γ_f=7 kPa)	
$\Lambda_C(C)$	=	$\frac{-\Lambda_{C1}}{1+\exp(-\xi_C(C-C_M))}$	(Λ_{C1}=12.5 kPa; $\xi_C = 8/C^*$ m^3/kg)	[63,66]
$\Lambda_H(H)$	=	$\frac{-\Lambda_{H1}}{1+\exp(-\xi_H(H-H_M))}$	(Λ_{H1}=25 kPa; $\xi_H = 16/H^*$ m^3/kg)	[63,66]
$\hat{f}_c(\alpha_c,\alpha_f)$	=	$\frac{\alpha_c[\alpha_f\zeta_{cf}+\alpha_c(\zeta_{cf}+\zeta_f)]}{(\alpha_c+\alpha_f)^2\zeta_{cf}+\alpha_c^2\zeta_f+\alpha_f^2\zeta_c+\frac{\alpha_w^2}{\zeta_w}(\zeta_c\zeta_{cf}+\zeta_c\zeta_f+\zeta_{cf}\zeta_f)}$		(Eqs. (2) and (3))
$\hat{f}_f(\alpha_c,\alpha_f)$	=	$\frac{\alpha_f[\alpha_c\zeta_{cf}+\alpha_f(\zeta_{cf}+\zeta_c)]}{(\alpha_c+\alpha_f)^2\zeta_{cf}+\alpha_c^2\zeta_f+\alpha_f^2\zeta_c+\frac{\alpha_w^2}{\zeta_w}(\zeta_c\zeta_{cf}+\zeta_c\zeta_f+\zeta_{cf}\zeta_f)}$		(Eqs. (2) and (3))
$\hat{h}_1(\alpha_c,\alpha_f)$	=	$\frac{\alpha_c\frac{\alpha_w^2}{\zeta_w}[\alpha_c(\zeta_f+\zeta_{cf})+\alpha_f\zeta_{cf}]\phi}{(\alpha_c+\alpha_f)^2\zeta_{cf}+\alpha_c^2\zeta_f+\alpha_f^2\zeta_c+\frac{\alpha_w^2}{\zeta_w}(\zeta_c\zeta_{cf}+\zeta_c\zeta_f+\zeta_{cf}\zeta_f)}$		(Eqs. (2) and (3))
$\hat{h}_2(\alpha_c,\alpha_f)$	=	$\frac{\alpha_c\alpha_f(\alpha_c\alpha_f-\frac{\alpha_w^2}{\zeta_w}\zeta_{cf})\phi}{(\alpha_c+\alpha_f)^2\zeta_{cf}+\alpha_c^2\zeta_f+\alpha_f^2\zeta_c+\frac{\alpha_w^2}{\zeta_w}(\zeta_c\zeta_{cf}+\zeta_c\zeta_f+\zeta_{cf}\zeta_f)}$		(Eqs. (2) and (3))
$\hat{h}_3(\alpha_c,\alpha_f)$	=	$\frac{\alpha_f\frac{\alpha_w^2}{\zeta_w}[\alpha_c\zeta_{cf}+\alpha_f(\zeta_c+\zeta_{cf})]\phi}{(\alpha_c+\alpha_f)^2\zeta_{cf}+\alpha_c^2\zeta_f+\alpha_f^2\zeta_c+\frac{\alpha_w^2}{\zeta_w}(\zeta_c\zeta_{cf}+\zeta_c\zeta_f+\zeta_{cf}\zeta_f)}$		(Eqs. (2) and (3))

Table 5
Parameters characterizing the mobility of tumor cells and fibroblasts by Eq. (2).

Variable	Description	Values	Source
I_w, r_w	fluid-ECM interaction	$2\cdot 10^{12}$ (Pa s/m^2), 0	[63,66]
k_w	fluid-ECM interaction	[1, 105]	calibrated
I_c, r_c	cell-ECM interaction	$5000I_w$ (Pa s/m^2), 0.8	[63,66]
k_c	cell-ECM interaction	≈ 1 (see Eq. (3))	[63,66]
I_f, k_f, r_f	fibroblast-ECM interaction	$100I_w$ (Pa s/m^2), 1.0, 0.6	[63,66]
I_{cf}, r_{cf}, r_{fc}	cell-fibroblast interaction	$1000I_w$ (Pa s/m^2), 0.5, 0.5	[63,66]
A, B	Reduced cell-ECM resistance	0.7, 10	[63,66]
$T_v T^*$	vascular conductivity	[0.0005,0.005] (1/Pa)	calibrated
$T_l T^*$	lymphatics filtration constant	0.0054 (1/Pa)	[66,72]

Table 6
Parameters for production/decay of chemical agents by Eqs. $(1)_{7,8}$.

Variable	Description	Values	Source
D_C	Diffusion coefficient chemokine	7×10^{-12} m^2/s	[63,66]
D_H	Diffusion coefficient TGF	8×10^{-12} m^2/s	
λ_{11}	Proteolytically freed chemokine	3×10^{-3} m^3/kg s	[63,66]
λ_{12}	Cell consumption rate chemokine	1×10^{-4} 1/s	
ν_C, M_C	Logistic rate exponent, absorption percentage	0.25, 50%	
λ_{21}	Proteolytically freed TGF	0.5×10^{-6} m^3/kg s	[63,66]
λ_{22}	Cell consumption rate TGF	5×10^{-3} 1/s	
λ_{23}	Natural decay of TGF	5×10^{-5} 1/s	
ν_H, M_H	Logistic rate exponent, absorption percentage	0.2, 50%	

Appendix B. Parameter estimation

There is clearly an uncertainty attached to the unknown parameters θ given by

$$\theta = (\alpha_f^0(\mathbf{x}), k_w(\mathbf{x}), T_v) \tag{19}$$

used to describe the tumor and its behavior. Each tumor will have its own spatially varying parameter fields, such as the initial fibroblast concentration, α_f^0, the tissue conductivity through k_w, and the vascular vessel wall permeability T_v as illustrated in Fig. 5. Initially, we assume that we have some knowledge of the typical (spatial) variation of these fields and their range as described by (12). The combination of our prior knowledge of the parameter fields with the information obtained from the data d, typically in terms of observed cell volume fraction $\alpha_c^{\text{obs},T_1}$ at some time T_1, can be done with a Bayesian approach for solving the model parameter estimation problem [5,61]. In the Bayesian approach, to solve the model parameter estimation problem, one starts out with a prior probability distribution, $p(\theta)$, of the uncertain parameters, θ. We assume that we can compute the likelihood of the data, $d = \left(\alpha_c^{\text{obs},T_1}\right)$, as $L(d|\theta)$. Combining these two quantities gives us the posterior distribution of the model parameters, $p(\theta|d)$, as

$$p(\theta|d) = \frac{L(d|\theta)p(\theta)}{\int L(d|\theta)p(\theta)\,d\theta} = \frac{L(d|\theta)p(\theta)}{c} \tag{20}$$

where the integration in the denominator is over the set of all possible model parameters θ. (Here $c = \int_{\text{all possible }\theta} L(d|\theta)p(\theta)\,d\theta$.)

To define the prior distribution of the parameter vector θ given by (19) we resort to describe the initial fibroblast concentration $\alpha_f^0(\mathbf{x})$ and the inverse tissue conductivity $k_w(\mathbf{x})$ as spatial varying fields, whereas we keep T_v as a value that is unknown, but fixed over the computational domain. Describing spatial varying fields is quite common within geosciences, see e.g. [10], and we will use terminology from that area. (An example of use of such terminology in the study of human tissue is [38].) Samples from the prior distribution can be used to predict the development of the tumor from the initial time to a time point of interest, here T_1, without constraining the model to any information from images of the tumor. This can be used to calculate statistics of interesting quantities, similar as done in the Monte-Carlo simulations provided in [8]. Finding samples of the posterior solution $p(\theta|d)$ in (20) could in principle be done with the Metropolis-Hastings algorithm, which is a Markov Chain Monte Carlo (MCMC) approach [61]. However, to solve the problem with this approach will require a large number of simulations. Therefore we suggest to use an approximative method based on the ensemble Kalman filter (EnKF). The EnKF approach [14] has become highly used in geoscience, in many fields, including weather forecasting [27,64], subsurface flow in reservoir engineering [1], and climate science [32, Chapter 20].

In the case that both the prior distribution and the measurement uncertainty follows a Gaussian distribution, and moreover there is a linear relationship between θ and the observed quantity d, a closed form solution for the posterior distribution can be obtained. This can be expressed as follows: Assume that the prior distribution $p(\theta)$ follows a Gaussian distribution $N(\mu, P)$, i.e. it has mean μ and covariance matrix P, and that $d|\theta$ follows a Gaussian distribution $N(H\theta, R)$, i.e. d is linearly related to θ by the expression $H\theta$ for a matrix H of appropriate dimension, and the uncertainty in the data is described by the covariance matrix R. Then the posterior distribution $p(\theta|d)$ follows the Gaussian distribution $N(\mu^*, P^*)$ where

$$\mu^* = \mu + K(d - H\mu) \tag{21}$$

and

$$P^* = (H^T R^{-1} H + P^{-1})^{-1} = P - PH^T(HPH^T + R)^{-1}HP \tag{22}$$

where I is the identity matrix of appropriate dimension, and K is the Kalman gain matrix

$$K = PH^T(HPH^T + R)^{-1}. \tag{23}$$

More details can be found in e.g., [5,32,61]. It is worthwhile to note that the same solution is obtained if one formulates the parameter estimation problem as solving the linear-least square problem

$$\min_\theta (d - H\theta)^T R^{-1}(d - H\theta) + (\theta - \mu)^T P^{-1}(\theta - \mu). \tag{24}$$

Then the minimum is achieved at μ^* given as in (21), cf. [5,61].

In general, the relationship between the observation vector d and the model parameter θ is nonlinear. Let us assume it can be described by $d = h(\theta)$. A set of samples that can serve as an approximative set of samples from the posterior distribution can then be found by the following algorithm:

Algorithm 1 *(Ensemble smoother).* An updated set of ensemble members, taking the measurements into account is calculated as follows (i is the index of the ensemble member):

1. Generate an ensemble of samples $\{\theta_i\}$ from the prior distribution $p(\theta_i)$.
2. Calculate a corresponding set of predicted observed values, $d_i = h(\theta_i)$.
3. Find an updated ensemble $\{\theta_i^*\}$ tacitly assuming that the ensemble of predicted observed values $\{d_i\}$ follows a Gaussian distribution, utilizing approximations of (21)–(23).

In our case the updated ensemble is calculated using the ensemble transform Kalman filter [27]. (For a range of other approaches for handling Step 3 of the above algorithm see for instance [64].) Let N_{ens} denote the number of samples in the ensemble. The update equations can then be expressed using the ensemble mean

$$\bar{\theta} = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \theta_i, \tag{25}$$

the mean for the predicted observed values,

$$\overline{h(\theta)} = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} h(\theta_i), \tag{26}$$

and the two matrices containing the ensemble of anomalies for θ and $h(\theta)$

$$A = \left[\theta_1 - \bar{\theta}, \ldots, \theta_{N_{\text{ens}}} - \bar{\theta}\right], \tag{27}$$

and

$$A_H = \left[h(\theta_1) - \overline{h(\theta)}, \ldots, h(\theta_{N_{\text{ens}}}) - \overline{h(\theta)}\right], \tag{28}$$

respectively. Then, the updated mean $\bar{\theta}^*$ is calculated using an empirical Kalman gain matrix $\hat{K}$

(i) obtained by replacing P with

$$\hat{P} = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} (\theta_i - \bar{\theta})(\theta_i - \bar{\theta})^T = \frac{1}{N_{\text{ens}} - 1} AA^T;$$

(ii) the expression PH^T with the empirical estimate $\widehat{PH^T} = \frac{1}{N_{\text{ens}}-1} AA_H^T$;

(iii) HPH^T with $\widehat{HPH^T} = \frac{1}{N_{\text{ens}}-1} A_H A_H^T$.

This gives an empirical formula for the Kalman gain (cf. (23)) as

$$\hat{K} = \widehat{PH^T}(\widehat{HPH^T} + R)^{-1}.$$

From this we can estimate the mean of the posterior distribution as

$$\bar{\theta}^* = \bar{\theta} + \hat{K}(d - \overline{h(\theta)}), \tag{29}$$

which completes our approximations for (21) and (23). A set of posterior anomalies can be obtained by finding a transformation matrix T such that $A^* = AT$, such that in light of (22)

$$\hat{P}^* = \hat{P} - \hat{P}H^T(H\hat{P}H^T + R)^{-1}H\hat{P}.$$

The above expression does not have a unique solution. Our solution is based on the solution from the ensemble transform Kalman filter [27] which is given as

$$T = \left[I + \frac{1}{N_{\text{ens}} - 1} A_H^T R^{-1} A_H\right]^{-1/2}$$

where the power $-1/2$ in the above expression denotes the inverse of the unique positive definite square root of a positive definite matrix. The updated ensemble members are then given as

$$\left[\theta_1^* \ldots \theta_n^*\right] = [\theta_1 \ldots \theta_n]\, T + \left[\bar{\theta}^* \ldots \bar{\theta}^*\right], \tag{30}$$

which gives us an updated ensemble that is giving an approximation of the posterior covariance matrix in (22). Note that by this choice of T the mean of $[\theta_1^*, \ldots, \theta_n^*]$ equals $\bar{\theta}^*$ making (29) and (30) consistent.

For better handling of any non-linearities in $h(\theta)$ we utilize an iterative approach called the ensemble smoother – multiple data assimilation (ES-MDA) [13]. This approach is based on the fact that if we make repeated observations of one (or more) data point(s), but with a certain higher uncertainty in the observation, we get the same solution as for a single observation. To be more precise, select a set of $\beta_j(\geq 1)$'s and a N such that $\sum_{j=1}^{N} 1/\beta_j = 1$. In the linear Gaussian case, assuming that d is observed N times with uncertainty provided by the covariance matrices $\beta_1 R$, $\beta_2 R$, ..., $\beta_N R$, we obtain the same solution of (20) as the one originally obtained by (21)-(23) recursively. We start with $p(\theta)$ given by the Gaussian distribution $N(\mu, P)$, utilize (21)-(23), but with the covariance matrix $\beta_1 R$ for the first uncertainty in the data. The posterior mean μ_1^* and P_1^* are used to describe the prior while we utilize the second data vector which has uncertainty in the data described with the covariance matrix $\beta_2 R$. After N set of observations are utilized we get the same solution as having one data vector with covariance matrix R, i.e. the posterior is given as $N(\mu^*, P^*)$. This argument is now utilized to change Algorithm 1 into an iterative approach where the measurement uncertainty is increased in each step, but we return back to Step 2 until N updates of the parameters have been done. This has the advantage that we limit the change in the parameter vector in each of the steps, and a better linear approximation of $h(\theta)$ in the update Step 3 can be achieved.

Appendix C. Numerical computations

The forward problem associated with (1) is solved in a semi-implicit manner. Pressure is computed implicitly (wrt time stepping) whereas advection of fluid and cells takes place through an explicit time stepping calculation [66]. For the calculations in this work we consider 61×61 grid blocks on a domain of size 1 cm$\times$1 cm, i.e., grid blocks size $\Delta x = \Delta y \approx 0.16$ mm. Simulation over a period $[0, T_2]$ (which amounts to 6 days) by using around 850 time steps per day takes around 140 sec. With this number of grid blocks it takes around 6-7 hours to do the simulation with 4 iterations (Ensemble smoother) and with an ensemble of 100 members where the simulated time period from T_0 to $T_1 = T_2/2$ represents approximately 3 days. For the computations we used a HP desktop, Intel(R) Core (TM) i9-9900K CPU @ 3.60 GHz 3.60 GHz.

To consider simulations in a preclinical setting as explored, for example, in [24], we may consider a physical domain of size 2.5 cm$\times$2.5 cm. Again, we may use 61×61 grid blocks which amount to $\Delta x = \Delta y \approx 0.4$ mm. Considering a simulated time period $[0, T_2]$ of 20 days we find that it takes around 140 sec computing time. This amounts to using around 250 time steps per day. Extension to 3D will of course increase the computing time, however, it seems still to be within the reach on this scale of the physical domain. For comparison we note that in [24] a numerical grid corresponding to $\Delta x = \Delta y = 0.25$ mm and $\Delta z = 1$ mm was used in combination with MR images acquired over a $32 \times 32 \times 16$ mm^3 field of view sampled with a $128 \times 128 \times 16$ matrix.

References

[1] S.I. Aanonsen, G. Nævdal, D.S. Oliver, A.C. Reynolds, B. Vallès, The ensemble Kalman filter in reservoir engineering – a review, SPE J. 14 (3) (2009) 393–412.
[2] R. Alert, X. Trepat, Physical models of collective cell migration, Annu. Rev. Condens. Matter Phys. 11 (2020) 77–101.
[3] L.M.K. Andersen, C.S. Wegner, T.G. Simonsen, R. Huang, J.V. Gaustad, A. Hauge, K. Galappathi, E.K. Rofstad, Lymph node metastasis and the physico-chemical micro-environment of pancreatic ductal adenocarcinoma xenografts, Oncotarget 8 (29) (2017) 48060.
[4] S. Angeli, T. Stylianopoulos, Biphasic modeling of brain tumor biomechanics and response to radiation treatment, J. Biomech. 49 (2016) 1524–1531.
[5] R.C. Aster, B. Borchers, C.H. Thurber, Parameter Estimation and Inverse Problems, International Geophysics Series, vol. 90, Elsevier, 2005.
[6] J. Barbazán, D.M. Vignjevic, Cancer associated fibroblasts: is the force the path to the dark side?, Curr. Opin. Cell Biol. 56 (2019) 71–79.
[7] H. Byrne, J. King, D. McElwain, L. Preziosi, A two-phase model of solid tumour growth, Appl. Math. Lett. 16 (2003) 567–573.
[8] J. Chen, D. Weihs, F.J. Vermolen, Computational modeling of therapy on pancreatic cancer in its early stages, Biomech. Model. Mechanobiol. (2019).
[9] X. Chen, R.M. Summers, J. Yao, Kidney tumor growth prediction by coupling reaction-diffusion and biomechanical model, IEEE Trans. Biomed. Eng. 60 (1) (2013) 169–173.
[10] J.P. Chilès, P. Delfiner, Geostatistics. Modeling Spatial Uncertainty, Wiley, 1999.
[11] A. Desmouliere, C. Guyot, G. Gabbiani, The stroma reaction myofibroblast: a key player in the control of tumor cell behavior, Int. J. Dev. Biol. 48 (5–6) (2004) 509–517.
[12] A. d'Esposito, P.W. Sweeney, M. Ali, M. Saleh, R. Ramasawmy, T.A. Roberts, G. Agliardi, A. Desjardins, M.F. Lythgoe, R.B. Pedley, et al., Computational fluid dynamics with imaging of cleared tissue and of in vivo perfusion predicts drug uptake and treatment responses in tumours, Nat. Biomed. Eng. 2 (10) (2018) 773–787.
[13] A.A. Emerick, A.C. Reynolds, Ensemble smoother with multiple data assimilation, Comput. Geosci. 55 (2013) 3–15.
[14] G. Evensen, Data Assimilation: The Ensemble Kalman Filter, Springer, 2007.
[15] G. Evensen, The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Syst. Mag. 29 (3) (2009) 83–104.
[16] K.R. Fischer, A. Durrans, S. Lee, J. Sheng, F. Li, S.T. Wong, H. Choi, T. El Rayes, S. Ryu, J. Troeger, et al., Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance, Nature 527 (7579) (2015) 472–476.
[17] G. Follain, D. Herrmann, S. Harlepp, V. Hyenne, N. Osmani, S.C. Warren, P. Timpson, J.G. Goetz, Fluids and their mechanics in tumour transit: shaping metastasis, Nat. Rev. Cancer 20 (2020) 107–124.
[18] H. Frieboes, F. Jina, Y.L. Chuang, S. Wise, J. Lowengrub, V. Cristini, Three-dimensional multispecies nonlinear tumor growth–II: tumor invasion and angiogenesis, J. Theor. Biol. 264 (4) (2010) 1254–1278.
[19] A. Fyles, M. Milosevic, M. Pintilie, A. Syed, W. Levin, L. Manchul, R.P. Hill, Long-term performance of interstitial fluid pressure and hypoxia as prognostic factors in cervix cancer, Radiother. Oncol. 80 (2) (2006) 132–137.
[20] C. Gaggioli, S. Hooper, C. Hidalgo-Carcedo, R. Grosse, J.F. Marshall, K. Harrington, E. Sahai, Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells, Nat. Cell Biol. 9 (12) (2007) 1392–1400.
[21] J. Guo, W. Lou, Y. Ji, S. Zhang, Effect of CCR7, CXCR4 and VEGF-C on the lymph node metastasis of human pancreatic ductal adenocarcinoma, Oncol. Lett. 5 (5) (2013) 1572–1578.
[22] L.M.K. Hansem, R. Huang, C.S. Wegner, T.G. Simonsen, J.V. Gaustad, A. Hauge, E.K. Rofstad, Intratumor heterogeneity in interstitial fluid pressure in cervical and pancreatic carcinoma xenografts, Transl. Oncol. 12 (8) (2019) 1079–1085.

[23] T. Hompland, C. Ellingsen, K.M. Øvrebø, E.K. Rofstad, Interstitial fluid pressure and associated lymph node metastasis revealed in tumors by dynamic contrast-enhanced MRI, Cancer Res. 72 (19) (2012) 4899–4908.
[24] D.A. Hormuth, A.M. Jarrett, X. Feng, T.E. Yankeelov, Calibrating a predictive model of tumor growth and angiogenesis with quantitative MRI, Ann. Biomed. Eng. 47 (7) (2019) 1539–1551.
[25] D.A. Hormuth, J.A. Weis, S.L. Barnes, M.I. Miga, E.C. Rericha, V. Quaranta, T.E. Yankeelov, A mechanically coupled reaction–diffusion model that incorporates intra-tumoural heterogeneity to predict in vivo glioma growth, J. R. Soc. Interface 14 (128) (2017) 20161010.
[26] M. Hubbard, H.M. Byrne, Multiphase modelling of vascular tumour growth in two spatial dimensions, J. Theor. Biol. 316 (2013) 70–89.
[27] B.R. Hunt, E.J. Kostelich, I. Szunyogh, Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, Physica D 230 (2007) 112–126.
[28] R.K. Jain, J.D. Martin, T. Stylianopoulos, The role of mechanical forces in tumor growth and therapy, Annu. Rev. Biomed. Eng. 16 (2014) 321–346.
[29] A.M. Jarrett, D.A. Hormuth, S.L. Barnes, X. Feng, W. Huang, T.E. Yankeelov, Incorporating drug delivery into an imaging-driven, mechanics-coupled reaction diffusion model for predicting the response of breast cancer to neoadjuvant chemotherapy: theory and preliminary clinical results, Phys. Med. Biol. 63 (2018) 820–830.
[30] A.M. Jarrett, D.A. Hormuth II, C. Wu, A.S. Kazerouni, D.A. Ekrut, J. Virostko, A.G. Sorace, J.C. DiCarlo, J. Kowalski, D. Patt, B. Goodgame, S. Avery, T.E. Yankeelov, Evaluating patient-specific neoadjuvant regimens for breast cancer via a mathematical model constrained by quantitative magnetic resonance imaging data, Neoplasia 22 (2020) 820–830.
[31] R. Kalluri, M. Zeisberg, Fibroblasts in cancer, Nat. Rev. Cancer 6 (5) (2006) 392–401.
[32] H. Kaper, H. Engler, Mathematics & Climate, SIAM, 2013.
[33] K.M. Kingsmore, D.K. Logsdon, D.H. Floyd, S.M. Peirce, B.W. Purow, J.M. Munson, Interstitial flow differentially increases patient-derived glioblastoma stem cell invasion via CXCR4, CXCL12, and CD44-mediated mechanisms, Integr. Biol. 8 (2016) 1246.
[34] K.M. Kingsmore, A. Vaccari, D. Abler, S.X. Cui, F.H. Epstein, R.C. Rockne, S.T. Acton, J.M. Munson, MRI analysis to map interstitial flow in the brain tumor microenvironment, APL Bioeng. 2 (2018) 031905.
[35] E.J. Kostelich, Y. Kuang, J.M. McDaniel, N.Z. Moore, N.L. Martirosyan, M.C. Preul, Accurate state estimation from uncertain data and models: an application of data assimilation to mathematical models of human brain tumors, Biol. Direct 6 (64) (2011), http://www.biology-direct.com/content/6/1/64.
[36] A. Labernadie, T. Kato, A. Brugués, X. Serra-Picamal, S. Derzsi, E. Arwert, A. Weston, V. González-Tarragó, A. Elosegui-Artola, L. Albertazzi, et al., A mechanically active heterotypic E-cadherin/N-cadherin adhesion enables fibroblasts to drive cancer cell invasion, Nat. Cell Biol. 19 (3) (2017) 224–237.
[37] E. Lakiotaki, S. Sakellariou, K. Evangelou, G. Liapis, E. Patsouris, I. Delladetsima, Vascular and ductal elastotic changes in pancreatic cancer, APMIS, Acta Pathol. Microbiol. Immunol. Scand. 124 (3) (2016) 181–187.
[38] H.M. Langevin, D.M. Rizzo, J.R. Fox, G.J. Badger, J. Wu, E.E. Konofagou, D. Stevens-Tuttle, N.A. Bouffard, M.H. Krag, Dynamic morphometric characterization of local connective tissue network structure in humans using ultrasound, BMC Syst. Biol. 1 (25) (2007).
[39] T.D. Lewin, P.K. Maini, E.G. Moros, H. Enderling, H.M. Byrne, A three phase model to investigate the effects of dead material on the growth of avascular tumours, Math. Model. Nat. Phenom. 15 (22) (2020).
[40] Y. Liu, J. Li, S. Sun, B. Yu, Advances in Gaussian random field generation: a review, Comput. Geosci. 23 (2019) 1011–1047.
[41] R.J. Lorentzen, T. Bhakta, D. Grana, X. Luo, R. Valestrand, G. Nævdal, Simultaneous assimilation of production and seismic data: application to the Norne field, Comput. Geosci. 24 (2020) 907–920.
[42] S.J. Lunt, A. Fyles, R.P. Hill, M. Milosevic, Interstitial fluid pressure in tumors: therapeutic barrier and biomarker of angiogenesis, Future Oncol. 4 (6) (2008).
[43] P. Mascheroni, J.C.L. Alfonso, M. Kalli, T. Stylianopoulos, M. Meyer-Hermann, H. Hatzikirou, On the impact of chemo-mechanically induced phenotypic transitions in gliomas, Cancers 11 (2019) 716.
[44] P. Mascheroni, C. Stigliano, M. Carfagna, D.P. Boso, L. Preziosi, P. Decuzzi, B.A. Schrefler, Predicting the growth of glioblastoma multiforme spheroids using a multiphase porous media model, Biomech. Model. Mechanobiol. 15 (2016) 1215–1228.
[45] F. Mpekris, S. Angeli, A.P. Pirentis, T. Stylianopoulos, Stress-mediated progression of solid tumors: effect of mechanical stress on tissue oxygenation, cancer cell proliferation and drug delivery, Biomech. Model. Mechanobiol. 14 (6) (2015) 1391–1402.
[46] J.M. Munson, R.V. Bellamkonda, M.A. Swartz, Interstitial flow in a 3D microenvironment increases glioma invasion by a CXCR4-dependent mechanism, Cancer Res. 73 (5) (2013) 1536–1546.
[47] G. Nævdal, E.K. Rofstad, K. Søreide, S. Evje, Fluid-sensitive migration mechanisms predict association between metastasis and high interstitial fluid pressure in pancreatic cancer, J. Biomech. 145 (154) (2022) 111362.
[48] H.T. Nia, H. Liu, G. Seano, M. Datta, D. Jones, N. Rahbari, et al., Solid stress and elastic energy as measures of tumour mechanopathology, Nat. Biomed. Eng. 1 (1) (2017) 716.
[49] W.J. Polacheck, J.L. Charest, R.D. Kamm, Interstitial flow influences direction of tumor cell migration through competing mechanisms, Proc. Natl. Acad. Sci. 108 (27) (2011) 11115–11120.
[50] Y. Qiao, P. Andersen, S. Evje, D. Standnes, A mixture theory approach to model co- and counter-current two-phase flow in porous media accounting for viscous coupling, Adv. Water Resour. 112 (2018) 170–188.
[51] Y. Qiao, Q. Li, S. Evje, On the numerical discretization of a tumor progression model driven by competing migration mechanisms, Math. Eng. 4 (6) (2022).
[52] M. Raissi, P. Perdikaris, G. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2018) 686–707.
[53] M. Raissi, A. Yazdani, G.E. Karniadakis, Hidden fluid mechanics: a Navier-Stokes informed deep learning framework for assimilating flow visualization data, arXiv:1808.04327, 2018.
[54] M. Raissi, A. Yazdani, G.E. Karniadakis, Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations, Science 367 (6481) (2020) 1026–1030.
[55] A.C. Shieh, H.A. Rozansky, B. Hinz, M.A. Swartz, Tumor cell invasion is promoted by interstitial flow-induced matrix priming by stromal fibroblasts, Cancer Res. 71 (3) (2011) 790–800.
[56] J.D. Shields, M.E. Fleury, C. Yong, A.A. Tomei, G.J. Randolph, M.A. Swartz, Autologous chemotaxis as a mechanism of tumor cell homing to lymphatics via interstitial flow and autocrine CCR7 signaling, Cancer Cell 11 (6) (2007) 526–538.
[57] J. Sperveslage, et al., Lack of CCR7 expression is rate limiting for lymphatic spread of pancreatic ductal adenocarcinoma, Int. J. Cancer 131 (4) (2012) E371–E381.
[58] T. Stylianopoulos, L. Munn, R.K. Jain, Reengineering the physical microenvironment of tumors to improve drug delivery and efficacy: from mathematical modeling to bench to bedside, Trends Cancer 4 (2018) 292–319.
[59] P.W. Sweeney, A. d'Esposito, S. Walker-Samuel, R.J. Shipley, Modelling the transport of fluid through heterogeneous, whole tumours in silico, PLoS Comput. Biol. 15 (6) (2019).
[60] I.F. Tannock, C.M. Lee, J.K. Tunggal, D.S. Cowan, M.J. Egorin, Limited penetration of anticancer drugs through tumor tissue: a potential cause of resistance of solid tumors to chemotherapy, Clin. Cancer Res. 8 (3) (2002) 878–884.
[61] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, 2005.

[62] M. Tektonidis, H. Hatzikirou, A. Chauvière, M. Simon, K. Schaller, A. Deutsch, Identification of intrinsic in vitro cellular mechanisms for glioma invasion, J. Theor. Biol. 287 (2011) 131–147.
[63] J. Urdal, J.O. Waldeland, S. Evje, Enhanced cancer cell invasion caused by fibroblasts when fluid flow is present, Biomech. Model. Mechanobiol. 18 (4) (2019) 1047–1078.
[64] S. Vetra-Cervalho, P.J. van Leeuwen, L. Nerger, A. Barth, M.U. Altaf, P. Brasseur, P. Kirchgessner, J.M. Beckers, State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems, Tellus, Ser. A Dyn. Meteorol. Oceanogr. 70 (1) (2018) 1445364.
[65] D. Von Ahrens, T.D. Bhagat, D. Nagrath, A. Maitra, A. Verma, The role of stromal cancer-associated fibroblasts in pancreatic cancer, J. Hematol. Oncol. 10 (1) (2017) 76.
[66] J.O. Waldeland, J.V. Gaustad, E.K. Rofstad, S. Evje, In silico investigations of intratumoral heterogeneous interstitial fluid pressure, J. Theor. Biol. 526 (2021) 110787.
[67] J.O. Waldeland, W.J. Polacheck, S. Evje, Collective tumor cell migration in the presence of fibroblasts, J. Biomech. 100 (2020) 109568.
[68] S. Walker-Samuel, T.A. Roberts, R. Ramasawmy, et al., Investigating low-velocity fluid flow in tumors with convection-MRI, Cancer Res. 78 (7) (2018) 1859–1872.
[69] M. Wang, J. Zhao, L. Zhang, F. Wei, Y. Lian, Y. Wu, Z. Gong, S. Zhang, J. Zhou, K. Cao, et al., Role of tumor microenvironment in tumorigenesis, J. Cancer 8 (5) (2017) 761.
[70] J.A. Weis, M.I. Miga, T.E. Yankeelov, Predicting the response of breast cancer to neoadjuvant therapy using a mechanically coupled reaction-diffusion model, Cancer Res. 75 (22) (2015) 4697–4707.
[71] J.A. Weis, M.I. Miga, T.E. Yankeelov, Three-dimensional image-based mechanical modeling for predicting the response of breast cancer to neoadjuvant therapy, Comput. Methods Appl. Mech. Eng. 314 (2017) 494–512.
[72] M. Wu, H.B. Frieboes, S.R. McDougall, M.A.J. Chaplain, V. Cristini, J. Lowengrub, The effect of interstitial pressure on tumor growth: coupling with the blood and lymphatic vascular systems, J. Theor. Biol. 320 (2013) 131–151.
[73] T.E. Yankeelov, V. Quaranta, K.J. Evans, E.C. Rericha, Toward a science of tumor forecasting for clinical oncology, Cancer Res. 75 (6) (2015) 918–923.
[74] S.G. Yeo, J.S. Kim, M.J. Cho, K.H. Kim, J.S. Kim, Interstitial fluid pressure as a prognostic factor in cervical cancer following radiation therapy, Clin. Cancer Res. 15 (19) (2009) 6201–6207.
[75] X. Zheng, J.L. Carstens, J. Kim, M. Scheible, J. Kaye, H. Sugimoto, C.C. Wu, V.S. LeBleu, R. Kalluri, Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer, Nature 527 (7579) (2015) 525–530.
[76] H. Zhou, P. Lei, T.P. Padera, Progression of metastasis through lymphatic system, Cells 10 (2021) 627.