

BAYESIAN CALIBRATION AND MODEL ASSESSMENT OF CELL MIGRATION DYNAMICS WITH SURROGATE MODEL INTEGRATION*

CHRISTINA SCHENK[†], JACOBO AYENSA JIMÉNEZ[‡], AND IGNACIO ROMERO[§]

Abstract. Computational models provide crucial insights into complex biological processes such as cancer evolution, but their mechanistic nature often makes them nonlinear and parameter-rich, complicating calibration. We systematically evaluate parameter probability distributions in cell migration models using Bayesian calibration across four complementary strategies: parametric and surrogate models, each with and without explicit model discrepancy. This approach enables joint analysis of parameter uncertainty, predictive performance, and interpretability. Applied to a real data experiment of glioblastoma progression in microfluidic devices, surrogate models achieve higher computational efficiency and predictive accuracy, whereas parametric models yield more reliable parameter estimates due to their mechanistic grounding. Incorporating model discrepancy exposes structural limitations, clarifying where model refinement is necessary. Together, these comparisons offer practical guidance for calibrating and improving computational models of complex biological systems.

Relevance to Life Sciences Glioblastoma is an aggressive brain tumor where invasive migration drives poor treatment outcomes. Comparing calibration approaches for models of glioblastoma progression in bio-mimetic platforms highlights key trade-offs: surrogate models improve prediction and speed, while parametric models preserve mechanistic interpretability relevant for biological insight. Explicit discrepancy modeling identifies when predictions fail to capture observed tumor behavior, guiding biologically motivated model improvements. These findings demonstrate how computational strategies can advance cancer modeling and support therapeutic planning.

Mathematical Content We employ Bayesian calibration to estimate posterior parameter distributions across four frameworks: parametric versus surrogate models, each assessed with and without discrepancy terms. Surrogates accelerate inference through approximated dynamics, while parametric formulations directly encode migration processes. The model discrepancy is formalized as an additive term that captures structural gaps between predictions and data. This systematic comparison reveals trade-offs between interpretability, uncertainty quantification, and efficiency, establishing a principled methodology for nonlinear model calibration in life sciences.

Key words. cell migration, Bayesian calibration, Bayesian inference, partial differential equations, glioblastoma progression, microfluidic devices

AMS subject classifications. 92C37, 62F15, 65C40, 68T05, 92-08, 35Q92

1. Introduction. Cell migration models are critical for predicting cancer progression. The complexity, heterogeneity, and dynamic nature of the Tumor Microenvironment hinder *in vivo* investigation of all tumor–stroma interactions. To overcome this, laboratory experiments at multiple scales aim to replicate the spatial and structural features of physiological and pathological environments. However, the limited predictive power of current *in vitro* models remains a major factor in the ongoing decline of new drug approvals relative to research investment [77] because of the limitations for reproducing important structural three-dimensionality, impacting on cell

*The first two authors are co-first authors. Submitted to the editors 09/23/2025.

Funding: The authors acknowledge financial support from the Spanish Ministry of Science and Innovation through a Juan de la Cierva Postdoctoral Fellowship for J.A.-J. (Grant No. JDC2023-052319-I).

[†]IMDEA Materials Institute, C/ Eric Kandel 2, Getafe, 28906, Spain (christina.schenk@imdea.org).

[‡]Universidad Politécnica de Madrid, José Gutiérrez Abascal, 2, Madrid, 29006, Spain, TME Lab, Aragón Institute of Engineering Research (I3A), Mariano Esquillor, S/N, Zaragoza, 50018, Spain (jacobo.ayensa@upm.es).

[§]Universidad Politécnica de Madrid, José Gutiérrez Abascal, 2, Madrid, 29006, Spain IMDEA Materials Institute, C/ Eric Kandel 2, Getafe, 28906, Spain (ignacio.romero@upm.es)

behavior through adhesive, mechanical, and chemical cues [30].

Mathematical models and computational tools have also emerged as powerful frameworks for the study of tumor progression, both as a complement of pre-clinical research and as predictive tools in computational oncology [29, 76, 91]. Parametric models, capable to assimilate biological data at different scales, are invaluable. The payoff of the mathematical abstractions (after sufficient validation) is to get information on the mechanisms driving tumor progression [3], build predictive tools [31], include treatment outcomes [59], explore “what if” situations, adapt the framework to complex geometries and patient-specific contexts [15], discard, validate or refine models [45, 47], design new experiments [65], support clinical decision making [67], or integrate multiple scales of analysis [4, 49, 89]. In fact, mathematical oncology is a valuable tool both when theory guides experiments and experiments guide new theory formulations [21] and nowadays is an opportunity for personalized medicine [90].

The use of Partial Differential Equations (PDEs) incorporating the specific behavior of the cells of the tumor under study has proven to be useful to reproduce tumor progression in experimental heterogeneous cell cultures [5] and in patient-specific geometries [56], both of which can be monitored over time while maintaining a moderate computational cost. Once calibrated, the models can be used to build digital twins [88] both to predict patient outcomes or to test drugs and treatment schedules *in silico*. This strategy was coined some years ago as *predictive medicine* [63].

Despite the benefits of parametric models, the calibration process is challenging. First, data quantity and quality are fundamental for parameter identifiability [70] and calibration [71]. Second, complex nonlinear models, mechanistically representing biological processes are usually multiparametric, thus being prone to overfitting. In any case, their validation is key, since these models will eventually be used in predictive scenarios [61]. Fortunately, the improvements in cell culture techniques, the development of new biomarkers, and the advent of new and more effective microscopy and imaging techniques triggers the assimilation of much richer although unstructured data [13].

Recently, enormous efforts have been made to establish a robust methodology for parametric adjustment in biology, particularly computational oncology. The seminal work [63] was the starting point for rigorous methodologies culminating in the OPAL method [53, 62], which has been successfully applied to gliomas [47] or breast cancer [68]. In addition, some context-dependent specific algorithms have been developed for learning purposes in real data scenarios [1] and nonconventional approaches have been proposed recently, such as the use of Physics-Informed Neural Networks [26] or hybrid approaches that combine PDEs with some elements of Machine Learning [22].

However, even if model selection is easily incorporated in most of the aforementioned workflows, they require both calibration and validation scenarios to decide on the adequacy of a particular model to be adopted. The concept of model discrepancy or model inadequacy is underutilized. The existence of a discrepancy term is discussed in many works (see, for instance, [25] for a Bayesian approach), but it is not explicitly included and therefore it is not exploited for decision-making. In addition, the computational cost of PDE models may require ancillary methods to accelerate their simulation for applications that require a high number of model runs, such as uncertainty quantification or reliability analysis [55], where the full *a posteriori* distribution is needed. Therefore, the use of inadequacy-aware surrogates is key. Indeed, in many contexts, we do not need the model to be right, but not to be wrong [32], so the model adequacy has to be evaluated, without having validation scenarios. Our main goal is to establish a framework for biological problems, where parametric calibration

can be applied systematically, the model inadequacy can be evaluated, and whose evaluation might be replaced with that of a surrogate, if needed.

With the idea of exploiting the advantages of Bayesian approaches for inverse problems [51], we present a methodology for model calibration in biology and biomedical engineering based on the Kennedy and O'Hagan error split [50]. The presented approach leverages this error decomposition to evaluate model adequacy, to assess the use of model surrogates, and to integrate or reconstruct the probability distribution of the experimental error. In addition, the posterior probability distribution of the parameters does not assimilate neither data noise nor model discrepancy, thus offering an improved biological characterization and therefore impacting downstream predictions, in terms of bias and uncertainty.

We illustrate the approach with a parametric model of spatial *in vitro* cell progression using real experimental data. This example is relevant enough from the biomedical point of view: we calibrate a parametric model based on PDEs that attempts to recreate the progression of glioblastoma (GBM), the most common and aggressive brain cancer [18, 82] in microfluidic devices, the most biomimetic cell culture technique [75]. Microfluidic platforms are very well-suited for recreating GBM physiological and pathological situations for research purposes [11, 12] and for drug testing [17]. In addition, we focus on the parameters related to the go-or-grow metabolic switch [42], a hallmark of GBM progression, using a previous parametric model [7] whose use has been demonstrated to be relevant for incorporating spatial cell cultures techniques in personalized medicine [66].

This paper is organized as follows. First, in Section 2, we introduce the mathematical model describing glioblastoma cell evolution, which accounts for growth, migration saturation, and oxygen moderation, leading to our case study based on a parametric model of tumor progression. We then present the model calibration framework and discuss the different approaches considered in Section 3. The results in Section 4 compare outcomes across four approaches, distinguishing between the use of the parametric versus surrogate model for calibration, and the presence or absence of a model discrepancy term. We also discuss the implications and significance of these findings as a foundation for future work. Finally, we conclude with a summary of our main contributions and propose avenues for further research in Section 5.

2. Mathematical model of glioblastoma cell culture evolution. In this section we describe the mathematical model of GBM progression in microfluidic devices. The setup of the experiment is shown in Fig. 1.

2.1. Governing equations. Our starting point is a nonlinear reaction-diffusion system of PDEs. In the devices under consideration, live and dead cell populations coexist. The PDEs that model their evolution are formulated in terms of the density of live and dead cells (v and u respectively) and of oxygen concentration (w). Following the work [7], the governing equations are:

$$(2.1a) \quad \frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left(D_n \frac{\partial u}{\partial x} - \chi F_{go}(u) \Pi_{go}(w) u \frac{\partial w}{\partial x} \right) + \frac{1}{\tau_n} F_{gr}(u) \Pi_{gr}(w) u - \frac{1}{\tau_d} \Pi_d(w) u,$$

$$(2.1b) \quad \frac{\partial v}{\partial t} = \frac{1}{\tau_d} \Pi_d(w) u,$$

$$(2.1c) \quad \frac{\partial w}{\partial t} = \frac{\partial}{\partial x} \left(D_{O_2} \frac{\partial w}{\partial x} \right) - \alpha \Pi_c(w) u,$$

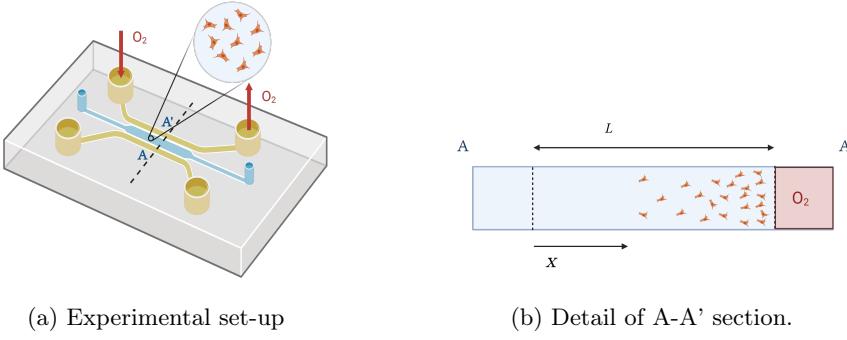


Fig. 1: **Experimental set-up.** A scheme of the microfluidic device is illustrated, together with the geometry and main variables of the model.

where the three unknown fields u, v, w are functions of time $t \geq 0$ and the space coordinate $x \in [0, L]$.

The first term of the right-hand side of Eq. (2.1a) represents the flow associated with cell culture migration and has two contributions: the non-oriented motility term $D_n \frac{\partial u}{\partial x}$ (modeled here as a random diffusion process) and the chemotaxis term $-\chi F_{go} \Pi_{go} u \frac{\partial w}{\partial x}$, where cell motility is induced by the oxygen gradient. F_{go} and Π_{go} are correction factors that account respectively for the limitations in the motion due to cell crowding and the changes occurring in cell metabolism associated with migration, depending on the oxygen levels, whose mathematical structure is discussed later. The second and third term corresponds to the reaction contribution and it is associated with cell growth and death. The correction terms F_{gr} , Π_{gr} and Π_d have a similar meaning and are also explained later.

With respect to the dead cell equation, Eq. (2.1b), it is assumed that dead cells do not migrate, neither by pedesis nor by chemotaxis. Therefore, we only consider the source term coming from the death of the cells.

The third equation, Eq. (2.1c), refers to the oxygen evolution equation; the first term of the right-hand side models the flow contribution, consisting solely of oxygen diffusion, and the second corresponds to the oxygen consumption by cells. As for the cell equation, the function Π_c is a nonlinear correction function.

2.2. Boundary and initial conditions. Eqs. (2.1a), (2.1b) and (2.1c) must be complemented with appropriate boundary and initial conditions, which are the data at a specific time. Note that we will refer to this time as $t = 0$ even if it is not necessarily the experiment starting time, but the instant when the cell culture profile is measured with the microfluidic device fully oxygenated.

$$(2.2a) \quad u(x, t = 0) = u_0(x),$$

$$(2.2b) \quad v(x, t = 0) = v_0(x),$$

$$(2.2c) \quad w(x, t = 0) = w_0,$$

with u_0 and v_0 given known function and w_0 the (homogeneous) ambient oxygen level.

With respect to the boundary conditions, we assume that some cells can escape from the central chamber to the lateral channels, but this cell flow is proportional to the cell concentration at the interface. No migration is considered for dead cells, so

no boundary conditions associated with Eq. (2.1b) are necessary. Regarding oxygen evolution, the oxygen concentration at the boundaries is assumed to be known and equal to w_0 , that is, the ambient concentration. Therefore, the boundary conditions for live cell population are:

$$(2.3a) \quad u(x = 0, t) - j f_u(x = 0, t) = 0,$$

$$(2.3b) \quad u(x = L, t) + j f_u(x = L, t) = 0,$$

where j is a parameter that controls the proportionality between cell concentration and cell flow given by

$$f_u(x, t) = D_n \frac{\partial u}{\partial x}(x, t) - \chi F_{\text{go}}(u(x, t)) \Pi_{\text{go}}(w(x, t)) u(x, t) \frac{\partial w}{\partial x}(x, t),$$

whereas for oxygen concentration

$$(2.3c) \quad w(x = 0, t) = w_0,$$

$$(2.3d) \quad w(x = L, t) = w_0.$$

2.3. Model corrections. To complete the model, we need to define the corrections that are specific to GBM, that is, the functions F_{go} , F_{gr} , Π_{go} , Π_{gr} , Π_{d} and Π_{c} that encode the cell behavior.

2.3.1. Growth and migration saturation. The functions F_{gr} and F_{go} are related to saturation considerations as cell growth and migration are limited by the amount of space. Therefore, F_{go} and F_{gr} depend on the cell density u . Following the work [7], we write

$$(2.4a) \quad F_{\text{gr}}(u, v) = 1 - \frac{u + v}{c_{\text{sat}}},$$

$$(2.4b) \quad F_{\text{go}}(u) = 1 - \frac{u}{c_{\text{sat}}},$$

where c_{sat} is the cell saturation. With respect to the nonlinear growth correction, it is required because dead cells occupy physical space, restricting the ability of viable cells to divide and also that cellular debris and necrotic material released by dead cells may alter the local microenvironment by generating toxic by-products, disrupting pH homeostasis, and affecting signaling pathways essential for cell growth [40]. Chemical cues released by dead cells inhibiting migration are neglected, hence only space needs are considered which are mainly modulated by live cell density.

2.3.2. Oxygen moderation. For some cellular mechanisms, oxygen acts as a moderator variable. In fact, cell hypoxia has been shown to play an important role in cell culture dynamics [11, 12, 19, 74], particularly in cell proliferation, death, and migration. These hypoxia-mediated dynamics have been proposed as one of the main driving forces of GBM invasion and aggressiveness [20].

The go-or-grow paradigm. The metabolic behavior of the GBM cells, in particular, its response to changes of oxygen pressure, is encoded in the functions Π_{go} and Π_{gr} . These two functions regulate the activation/deactivation of migration and proliferation. There is sound evidence in the scientific literature that the switch between proliferative and migratory activity in a cell population is hypoxia-mediated [23, 57],

that is, $\Pi_{\text{go}} = \Pi_{\text{go}}(w)$ and $\Pi_{\text{gr}} = \Pi_{\text{gr}}(w)$. We know that $\Pi_{\text{gr}} : \mathbb{R}^+ \rightarrow [0; 1]$ is a continuous, non-decreasing function and $\Pi_{\text{go}} : \mathbb{R}^+ \rightarrow [0; 1]$ is another, continuous non-increasing function. A slightly more restrictive condition is to consider that, in addition to monotonicity, we must have

$$\lim_{w \rightarrow 0^+} \Pi_{\text{go}}(w) = \lim_{w \rightarrow +\infty} \Pi_{\text{gr}}(w) = 1, \quad \lim_{w \rightarrow +\infty} \Pi_{\text{go}}(w) = \lim_{w \rightarrow 0^+} \Pi_{\text{gr}}(w) = 0.$$

Finally, based on energetic considerations, we may assume that cells invest their resources either in proliferation or migration, so it is plausible to postulate that there exists a function $Q : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ such that $\forall w \in \mathbb{R}^+$,

$$Q(\Pi_{\text{go}}(w), \Pi_{\text{gr}}(w)) = 1.$$

In the work by [7], a piecewise linear function is proposed to represent the go-or-grow behavior, namely,

$$(2.5a) \quad \Pi_{\text{gr}}(w) = \begin{cases} \frac{w}{h_1} & \text{if } 0 \leq w \leq h_1 \\ 1 & \text{if } w > h_1 \end{cases},$$

$$(2.5b) \quad \Pi_{\text{go}}(w) = \begin{cases} 1 - \frac{w}{h_1} & \text{if } 0 \leq w \leq h_1 \\ 0 & \text{if } w > h_1 \end{cases}.$$

Here, h_1 is the hypoxia threshold, the oxygen characteristic scale below which the cell metabolism changes from a proliferative to migratory nature. It is clear that for the considered model, the function that takes into account energy competing resources is $Q(x, y) = x + y$.

Cell death: apoptosis and necrosis. Cell death is a natural process that depends on many factors and agents with intrinsic stochastic nature [38]. Anoxia is one fundamental cause of cell death [80]. We can therefore assume that $\Pi_d : \mathbb{R}^+ \rightarrow [0, 1]$ is a non-increasing function. However, since apoptosis is always a cell death mechanism, it is reasonable to assume that $\forall w \in \mathbb{R}^+, \Pi_d(w) > 0$. In addition, we can impose that

$$\lim_{w \rightarrow 0^+} \Pi_d(w) = 1, \quad \lim_{w \rightarrow +\infty} \Pi_d(w) = 0.$$

The work [7] proposes the following nonlinear correction:

$$(2.6) \quad \Pi_d(w) = \frac{1}{2} \left[1 - \tanh \left(\frac{w - h_2}{\Delta h_2} \right) \right].$$

In Eq. (2.6), h_2 is the anoxia threshold and Δh_2 the anoxia sensitivity. These two parameters regulate the impact of oxygen on the metabolism of cell death, that is, whether the cell dead is due solely to apoptosis or to both apoptosis and necrosis.

Oxygen consumption. Finally, oxygen consumption is related with the kinetics of oxidative phosphorylation that occurs in the membrane of cellular mitochondria, usually described using a non-decreasing function $\Pi_c : \mathbb{R}^+ \rightarrow [0, 1]$ such that

$$\lim_{w \rightarrow 0^+} \Pi_c(w) = 0, \quad \lim_{w \rightarrow +\infty} \Pi_c(w) = 1.$$

It is usual in biology and pharmacology to use a Hill function for this purpose [39]. Following again the work [7], we assume Michaelis-Menten kinetics for oxidative phosphorylation, thus

$$\Pi_c(w) = \frac{w}{w + k_m}.$$

2.4. Case study: parametric model of glioblastoma progression. At this point, the presented model depends on eleven parameters, namely, D_n , D_{O_2} , χ , τ_n , τ_d , α , c_{sat} , h_1 , h_2 , Δh_2 , k_m . Some of them have a well-identified value in the scientific literature. For example, for those associated to the oxygen evolution equation, the oxygen diffusion $D_{O_2} = 1 \times 10^{-5} \text{ cm}^2 \cdot \text{s}^{-1}$ has been reported in many works [27, 83] as well as the Michaelis-Menten constant, $k_m = 2.5 \text{ mmHg}$ [2, 27]. For some others, the value is not so well characterized, although simple conventional experiments could be used for calibration purposes:

- The parameters related to cell growth and death under fully oxygenated conditions, τ_n , τ_d and c_{sat} , can be determined in cell growth experiments in ambient conditions and in the absence of oxygen gradients, both in microfluidic devices [52, 84], using cell spheroids [86], or even in conventional Petri dish experiments.
- Assuming that k_m is known, the oxygen consumption rate, α is easily obtained by measuring the ambient oxygen pressure in an isolated system with a controlled cell culture population. It is even possible to determine both k_m and α from the oxygen pressure using an Eadie–Hofstee diagram [46] or a Lineweaver–Burk plot [54].
- The identification and calibration of the parameters related to the metabolic changes of cell death mechanism, that is, h_2 and Δh_2 is more subtle, since it requires the experimental measurement of both the ambient oxygen pressure and the dead cell density evolution during the experiment.
- Finally, the non-oxygen-mediated pedesis constant, D_n , is more complicated to determine as spatial cell cultures are necessary. However, spheroid cultures [10] and microfluidic devices [60] offer a great opportunity for cell migration evaluation. If full oxygenation is guaranteed in the whole culture, no oxygen gradient is formed so non-oxygen mediated pedesis is easily computed, for instance, once given τ_n , D_n can be determined by evaluating the cell migration radial velocity V and using the Fisher's model [34], $V = 2\sqrt{D_n\tau_n}$.

The values of χ and h_1 are nevertheless significantly more difficult to determine. To identify their values we need to experimentally establish heterogeneous oxygen distributions. Indeed, as the cell migration depends on the oxygen level (and not only on the oxygen gradient), it is difficult to estimate this value from a single experiment or measurement. However, we can measure cell culture migration under an oxygen gradient in a very localized region where the oxygen level may be considered almost constant [37]. From now on, we will define $b = 1/h_1$, the inverse of the hypoxic threshold, as the model parameter for comparison purposes.

In addition to the model parameters, we have two constants j and w_0 associated with the boundary conditions, assuming the initial profiles $u_0(x)$ and $v_0(x)$ as given. The ambient oxygen condition w_0 can be easily measured, but the parameter j is related to cell fluxes and depends both on the properties of the cell culture gel, the physics of the interface lateral/central chambers, and the cell properties, so it is difficult to fix it *a priori*.

Therefore, we consider the following parametric model $\mathcal{M}_{\boldsymbol{\theta}}$:

$$(2.7) \quad \eta(x; \boldsymbol{\theta}) = \hat{u}(x, t = T; \boldsymbol{\theta}),$$

where \hat{u} is the live cell density field obtained by solving numerically the PDEs (2.1) with initial conditions (2.2) and boundary conditions (2.3), T is a given experimental and simulation horizon and the (unknown) parameter vector is $\boldsymbol{\theta} = (\tau_n, \chi, b, j)$. The reason for including the parameter τ_n is because the triad of parameters (τ_n, χ, h)

fully characterize the *go-or-grow* metabolic switch [42], that has been the subject of scientific research (see for instance [6] for an extensive review). The remaining model parameters are the ones that are considered in [7].

3. Model calibration. The model described in Section 2 can be employed to predict cell migration, provided that the boundary and initial conditions are set up, and all the parameters appearing in the equations are given appropriate values. This last aspect, however, entails many more difficulties that might naively appear. Moreover, an incorrect parameter selection can compromise the modeling and computational efforts spent towards migration predictions.

3.1. Calibration approaches. In general, determining the correct values for the parameters of a model is not trivial. These must be adjusted using information obtained from experimental data but, since the latter do not provide direct parameter values, some type of *inference* must be performed. We refer to as *calibration* to the process of selecting the best values for the parameters given the information available. Naturally, since there is not a single choice for what is *best* and different sources of information can be incorporated or disregarded, several alternative and valid calibration strategies exist.

The simplest and most common approach to model calibration is based on mismatch minimization: given some norm (for example, the Euclidean norm) measuring the discrepancy between the model predictions and the measurements on a set of experiments, the parameters are selected to minimize this error (see, e.g., [41] for a comprehensive overview). Even if often employed, this calibration strategy has well-known limitations: it is very sensitive to outliers, it might require some regularization that affects the solution, it might select parameter values that only yield local minima to the mismatch error, and, finally, can only provide point estimates possibly with often very wide confidence intervals for the optimal parameter values [14].

More powerful methods for calibration can be obtained by leveraging Bayesian statistics. Building on the classical work of [50], numerous applications have appeared in the literature that refine the original model [43, 44, 64, 85] or apply it to various fields [24, 28, 87]. The Bayesian point of view corrects many of the deficiencies of the least-squares approach, being especially suitable when the number of experiments that can be conducted is scarce.

There is a fundamental difference between the minimization approach to calibration and the Bayesian one. In the first class of methods, optimal *values* are sought for the unknown parameters. In the latter, optimal *probability distributions* are the object of interest. In this statistical picture, it is therefore accepted from the outset that parameters of (almost) all models are random variables. The goal of Bayesian calibration is not to find a single value for each parameter, but rather understand the probability distribution of each value, the correlations between pairs of parameters, etc. This probabilistic point of view is indispensable to tackle all the critical issues associated with the *uncertainty* of models and their predictions.

In this work, and in contrast with our previous article [7], we embrace the Bayesian approach to calibration of models for cell migration, and provide a detailed exploration of its benefits. We start next by summarizing its main features in an abstract fashion before applying it to the models of interest in this work.

3.2. Bayesian calibration. To start, let us consider the output of a physical phenomenon (or economical, chemical, social, etc.) which is denoted as φ and is a function of some controllable variables, an array $\mathbf{x} \in \mathbb{R}^k, k \geq 1$. When a measure-

ment z is made for $\varphi(\mathbf{x})$, the result obtained is never exact and thus we can write

$$(3.1) \quad z = \varphi(\mathbf{x}) + \epsilon(\mathbf{x}),$$

where ϵ is a random variable accounting for the experimental error, usually assumed to be normally distributed. From an experimental campaign, one often has access to a dataset $\mathcal{D} = \{(\mathbf{x}_i, z_i)\}_{i=1}^M$ of M (inexact) measurements.

Let us now consider an analytical or computational model $\eta : \mathbb{R}^k \times \mathbb{R}^p \rightarrow \mathbb{R}$ of the same phenomenon as above where $\eta(\mathbf{x}; \boldsymbol{\theta})$ is now a function of the controllable variables \mathbf{x} and an array of parameters $\boldsymbol{\theta} \in \mathbb{R}^p$. In contrast with the previous situation, now the model η is completely known; in turn, the parameters are either unknown or, more commonly, known only approximately. This situation is the one we face with the cell migration model of Section 2.

The Bayesian approach to model calibration starts from the assumption that searching directly for the optimal value of the model parameters $\boldsymbol{\theta}$ is inconsistent with our knowledge and the information provided by the experiments. Rather, based on the prior information that one might have about these parameters and the experimental results, there exists a probability distribution for $\boldsymbol{\theta}$ that best explains the measurements, incorporating the previously existing information. The ultimate goal of *Bayesian inference* is the determination of the latter probability distribution, the so-called *posterior* distribution.

3.3. The Kennedy and O'Hagan split. Next, we adapt the Kennedy and O'Hagan (KOH) ideas for calibrating a computer code that solves an initial boundary value problem of the type describing cell migration. The starting point of this Bayesian strategy is, as anticipated, an *a priori* multivariate probability distribution for the parameters whose probability density function will be denoted as $\pi = \pi(\boldsymbol{\theta})$. This distribution should reflect, as best as possible, the previous knowledge that we have about the model parameters, be it from previous models, other research groups, preliminary tests, etc.

To *infer* a better probability distribution for the parameters, any Bayesian method requires two more ingredients: data and a model for the likelihood of these data as a function of the parameters. The first one is provided by an experimental campaign \mathcal{D} as defined above. To construct the second and more complex element let us note, first, that the model η is typically insufficient to predict correctly all the experimental results and thus we assume that

$$(3.2) \quad z_i = \eta(\mathbf{x}_i; \boldsymbol{\theta}) + \delta(\mathbf{x}_i) + \epsilon(\mathbf{x}_i), \quad i = 1, 2, \dots, M,$$

where $\delta : \mathbb{R}^k \rightarrow \mathbb{R}$ is the *discrepancy* error, an unknown function that accounts for the inability of the model η to reproduce φ , even with the best choice of parameters. The objective of the KOH calibration is thus not only to obtain the best *a posteriori* distribution for $\boldsymbol{\theta}$, but also to reveal δ , at least in a way that is as consistent with the data and prior information as possible.

A convenient way to work with a general class of *non-parametric* functions and select from them the discrepancy error is to assume it is a Gaussian process. In addition, the model η might be expensive to evaluate so it also proves convenient to introduce a surrogate or meta-model $\tilde{\eta}$ for it, one that can be evaluated much faster. Since the error is normally distributed and the discrepancy is a Gaussian process, the calculations become simple when also $\tilde{\eta}$ is selected among Gaussian processes.

3.4. Gaussian processes. Gaussian processes (GP) play a key role in the calibration framework proposed by Kennedy and O'Hagan. GPs can be understood as generalizations of Gaussian multivariate distributions. Specifically, they are a collection of random variables indexed by a real number such that every finite subset of these variables has a multivariate normal distribution. We indicate that f is a GP with mean function m and covariance c using the notation $f \sim \mathcal{GP}(m(\mathbf{x}), c(\mathbf{x}, \mathbf{x}'))$. See [69] for a comprehensive description of these concepts.

In practice, here, we will always employ GPs with $m \equiv 0$ and covariance functions of the form

$$(3.3) \quad c(\mathbf{x}, \mathbf{x}') = \lambda \exp \left[-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\beta^2} \right],$$

with λ, β referred to as *hyperparameters*, λ denoting the signal variance and β the lengthscale. This corresponds to a stationary isotropic squared exponential kernel. In the calibration, the hyperparameters are unknown and their optimal probability distributions will have to be obtained, as for any other unknown parameter of the model.

In particular, the hyperparameter β has the same dimensions as the input variable \mathbf{x} , as it is thus referred to as a *lengthscale*. Note, also, that the norm of the input variable \mathbf{x} is only well-defined if the dimensions of all its components are the same. It therefore becomes practical to nondimensionalize the model equations from the outset, ensuring the consistency of the norm operations.

3.5. Calibration procedure. The Bayesian calibration consists, essentially, in calculating a posterior probability distribution given the prior probability distribution of the parameters and the hyperparameters, plus the likelihood function of the data.

In this work, we consider four different calibration procedures that illustrate the main features of Bayesian calibration. In the following, we succinctly introduce them and refer to the literature for more details on the actual calculations involved in their computation [43, 50, 64].

Bayesian inference (BI). The simplest calibration strategy is a prototypical example of Bayesian inference. In this case, the discrepancy of the model is ignored, but we have some prior knowledge for the parameters of the model and the measurement error in experimental data, which is assumed to be normally distributed. Then, a straightforward application of Bayes' theorem provides the posterior probability distribution for all the unknown random variables, that is, the parameters and the experimental error.

Although some analytical results enable the calculation of the posterior for simple priors and low dimensions, for the sake of generality, this probability distribution is often approximated using MCMC methods (see, for example, general monographs on the subject [16, 72], or a recent work with references to some of the latest algorithms [73]). Let us note that these numerical techniques are extremely efficient, but demand a large number of likelihood and prior evaluations.

Bayesian calibration of expensive models (BCE). It might turn out that the model to be calibrated is computationally expensive and thus impossible — or impractical — to run thousands of MCMC steps that explore the parameter and hyperparameter space. In these situations, the user might decide to replace the model with a *surrogate* or *meta-model* that, once trained, can be evaluated with a small cost and used for calibration. Since the true model is replaced with a surrogate, naturally, the calibration will be less accurate than the one described before.

While replacing an expensive model with a fast surrogate seems appealing, one should be aware of its limitations. To construct the surrogate, a small set of model evaluations needs to be used in the training phase. For this set to be representative of the model, the input and parameter spaces need to be sufficiently covered. In high-dimensional spaces, this methodology is very costly and efficient sampling methods such as the Latin hypercube are employed [48, 58]. Note that although Latin hypercube sampling is specifically designed to avoid clustering and ensure a uniform distribution of samples across each dimension, it may suffer from clustering effects when the dimensionality is high. This issue becomes even more apparent in the presence of constraints [33, 78]. Moreover, these methods cannot compensate for a vague definition of the parameter range, especially if the model is very sensitive to some of them. When the input and parameters are poorly sampled, we cannot expect the surrogate to be accurate, and the probability distributions of the model parameters will be uninformative. In these situations, it often happens that the posterior parameter distributions are almost identical to the prior distributions provided in the calibration, and the MCMC acceptance ratio is small. These two are signs that the likelihood is very small in the MCMC chains, and that the posterior is controlled by the prior distributions, meaning that the provided data is uninformative. If any of these flags is raised, the analyst is advised to narrow the parameter range in the surrogate training, if possible, or abandon this strategy altogether.

To calibrate a model using a surrogate, one needs to first generate a set of *synthetic* data points, obtained by evaluating the model with input and parameter values $\{\tilde{\mathbf{x}}_i, \tilde{\boldsymbol{\theta}}_i\}_{i=1}^M$ randomly selected in input and parameter space using, as indicated, e.g., the Latin hypercube method.

Bayesian calibration with discrepancy (BCD). When calibrating models, it is not uncommon that the optimal parameters result in a model whose predictions are not sufficiently good. This could be due to a bad model choice, often the price for selecting a model that is too simple. One of the outcomes of this situation is that the surrogate model variant (cf. previous subsection) can represent the experimental data even better than the physical model itself. Even then, the discrepancy of the model can still be estimated. Calibrations of this type do not typically improve the calibrated model; however, they can quantify the mismatch of the model and the predictions, hinting at the need to select a better model if more accuracy is desired.

Bayesian calibration of expensive models with discrepancy (BCED). When evaluating a computationally expensive model, and the goal is to assess the quality of the model, here represented by a surrogate, both the physical model and the discrepancy are approximated using Gaussian Process (GP). This allows for quantifying the mismatch between the surrogate and the resulting predictions, thereby informing the selection of more effective synthetic data to construct such a surrogate and/or a more appropriate kernel representation. Moreover, incorporating the discrepancy explicitly can increase the predictive uncertainty, which transparently reflects the combined uncertainties from both the surrogate approximation and possible model inadequacies, ultimately leading to more cautious and reliable predictive statements.

4. Results and discussion. Next, we present the main findings from the four Bayesian calibration strategies described in Section 3 evaluating their performance and discussing how the probabilistic results can be leveraged for downstream analyses and decision-making. Specific data and model configurations can be found in the Appendix.

4.1. Evaluating the calibration procedures. We first evaluate the calibration methodology using the parametric model proposed in [7] (see Section 2.4). The results are shown in Fig. 2 for the BI approach, Fig. 3 for the BCE approach, Fig. 4 for the BCD approach and Fig. 5 for the BCED approach and the corresponding posterior distributions in Fig. 7, 8, 9, 10, respectively. For comparison purposes, we define the prediction error of the live cell profile as

$$(4.1) \quad e = \frac{1}{c_{\text{sat}}} \left(\int_0^L (\hat{u}(x) - u_{\text{exp}}(x))^2 dx \right)^{1/2}.$$

In Fig. 2 we show the live cell profile obtained when the model parameters and the experimental error are adjusted using the BI approach and the experimental data. We show the predicted cell profiles and the experimental error by both MAP (Maximum a posteriori) and (sample) mean estimation of the parameters from the posterior distributions. The predictions obtained using the MAP give an error of $e = 0.0326$, improving the prediction in [7], with $e = 0.0768$. The mismatch between experimental data and model predictions is mostly absorbed by the experimental error, which is not negligible ($\sigma = 7.2 \times 10^{-2} c_{\text{sat}}$, so 7.2% of the saturation concentration). Interestingly, the predictions using the MAP and mean values of the parameters are very similar, illustrating the low manageable variability of the posterior parameter distributions.

In Fig. 3 we show the results obtained when using the BCE approach. In addition to the predictions obtained with the full model, we include the mean \pm standard deviation of the surrogate. Although the surrogate reduces the error to $e = 0.0075$, the error of the predictions obtained using the MAP is $e = 0.1161$, illustrating the failure to match the parametric model and the experimental data in the surrogate. Unlike in the previous case, the experimental error estimation is drastically reduced, at the cost of greatly increasing the variability of the model parameters as suggested by the differences when estimating the cell profile using the MAP and the mean of the parameters.

In Fig. 4, we show the predictions of the cell profile using the BCD approach. With the inclusion of the discrepancy, the predictions are of the same order as ones reported in previous works, with $e = 0.0948$ achieved using the MAP value of the parameters. The error is reduced to $e = 0.0068$ when considering the model discrepancy. The difference between these latter two errors is an indication of model inadequacy. However, when compared to the previous approaches, both the experimental error and parameter variability (as indicated by the overlap of MAP and mean predictions) are kept small. The mismatch between experimental data and model predictions is fundamentally absorbed by the discrepancy term. This is, indeed, the main component of the model that is able to detect model inadequacy, regardless of experimental error, while keeping the parameter approach valid enough, that is, assuming a posterior distribution of the parameters informative enough.

Finally, in Fig. 5 we show the predictions using the BCED approach. The error obtained using the MAP values is $e = 0.1132$, similar to the one obtained without the discrepancy. When including the discrepancy, the error decreases to $e = 0.0185$, which is not as small as that obtained without the use of a surrogate (it is the price to pay to reduce computational costs), but is still smaller than any error obtained without the inclusion of the discrepancy. The use of a surrogate also impacts both the uncertainty of the parameters (the difference between the predictions using mean and MAP) and the predictions using the model and the discrepancy (broader band). The prediction using the surrogate yields an error of $e = 0.0116$, which is larger than

the error obtained without the inclusion of the discrepancy, although we should recall that the use of a surrogate model here is merely instrumental.

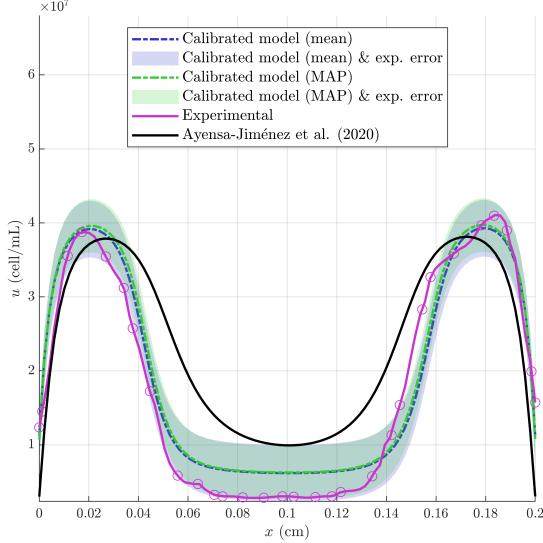


Fig. 2: Live cell profile generated. BI approach. Earlier results [7] showed an error of $\epsilon = 0.0768$, when compared to the prediction using the MAP parameters, with $\epsilon = 0.0326$.

As discussed previously, the inclusion of a discrepancy error δ accounts for the inability of the parametric model η to reproduce the actual phenomenon, even with the best possible choice of parameters. This allows for an initial assessment of the suitability of a model without the need for validation scenarios. In Fig. 6, we illustrate this fact by showing the GP representing the discrepancy $\delta(x)$, the standard deviation of the experimental error $\epsilon(x)$, and the deviation $d(x) = z(x) - \eta(x; \hat{\theta})$, where $\hat{\theta}$ are the MAP estimation of the parameters. In both approaches, BCD and BCED, the discrepancy is close to the deviation, showing a good predictive capacity of the overall approach, whereas it is far from the experimental error, illustrating the limited ability of the parametric model to explain the data. In other words, we can recreate the process φ with good accuracy, but there is room for improvement for an interpretable or mechanistic description. In fact, the use of δ instead of d is more enlightening, since d also incorporates the measurement error ϵ . Decision-making is less obvious when using a surrogate, which is the price to pay for reducing computational cost.

In particular, for the case study analyzed, we observe negative bias close to the boundaries of the microfluidic device and positive bias at the center of the chamber, hinting at mechanisms that could be modified or enhanced in the parametric model.

4.2. Using the probabilistic model for downstream tasks. Once the parametric model is calibrated, the posterior distribution of the model parameters θ , denoted by \mathcal{F}_θ , may be used for many downstream tasks, such as uncertainty quantification of a certain Quantity of Interest (QoI) (velocity of a propagating wave, GBM tumor burden growth, amount of cells extravasating the lateral channels, that is an

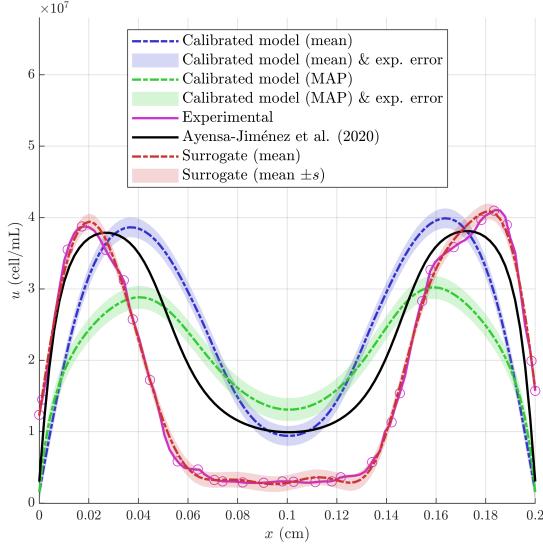


Fig. 3: Live cell profile generated using BCE approach. Earlier results [7] showed an error of $e = 0.0768$, when compared to the prediction using the MAP parameters, with $e = 0.1161$ and the prediction of the surrogate $e = 0.0075$.

indicator of the metastasis process, among others), goodness of fit evaluation of statistical tests (for instance to evaluate if a certain drug affects cell behavior through some model parameter). Therefore, an appropriate characterization of the statistics of the posterior distribution of the parameters is paramount. Here we illustrate this issue by analyzing the distributions \mathcal{F}_θ for each of the approaches, as well as the distribution of the experimental error σ and the eventual hyperparameters. We also discuss the statistics of the cell profile $\hat{u}(x)$, $x \in [0; L]$, $L = 2$ cm, as an example of QoI.

The results of the BI approach are summarized in the corner plot shown in Figure 7. The first four parameters (τ_n, χ, b, j) correspond to the model parameters, while σ quantifies the standard deviation of the experimental error. On the diagonal, the marginal posterior distributions for each parameter are displayed, with the median value and corresponding 95% credible intervals indicated. The off-diagonal subplots illustrate the joint posterior distributions, capturing potential correlations between parameter pairs. Each of the model parameters is reasonably well constrained by the data, with unimodal marginal posteriors and moderate credible intervals. The standard deviation of the experimental error, σ , has a relatively narrow distribution, suggesting a robust estimation of the measurement noise and indicating that the model is capable of explaining the majority of experimental variability. Inspecting the joint distributions, weak to moderate correlations are apparent between some parameter pairs, e.g., between b and j or χ and b , while other pairs, such as j and σ , show minimal dependency, which supports the identifiability of the parameter set under the current data and model configuration. MAP and mean estimates for all parameters are marked in blue and red, and found to be consistent with each other, further reflecting the unimodal nature of the posterior distributions, indicating convergence and well-defined parameter estimates. Taken together, these results show that this BI

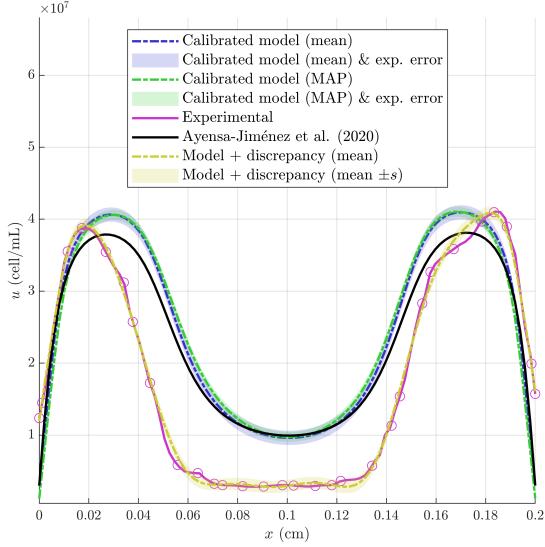


Fig. 4: Live cell profile generated using the BCD approach. Earlier results [7] showed an error of $e = 0.0768$, when compared to the prediction using the MAP parameters, with $e = 0.0948$ and the prediction including the discrepancy $e = 0.0068$.

approach delivers precise parameter estimates along with a transparent uncertainty quantification, capturing both model and experimental errors, thus improving the reliability of model predictions.

The corner plot in Figure 8 presents the results of the BCE approach that include the first four parameters of the physical model (τ_n, χ, b, j), followed by three surrogate hyperparameters from the GP ($\beta_x, \beta_\theta, \lambda_x$), and finally the standard deviation of the experimental error, σ . The physical model parameters are well constrained with unimodal marginal posteriors, indicating the data provide strong information about the system behavior. The GP hyperparameters, which govern the surrogate model's kernel structure and smoothness, exhibit distinct posterior distributions reflecting uncertainty in surrogate model tuning. Some weak to moderate correlations appear mainly among the hyperparameters and between hyperparameters and some model parameters, visible as elliptical shapes in the joint distributions. This reveals interdependencies in how surrogate model tuning interacts with physical parameters. The experimental error standard deviation, σ , is tightly estimated, indicating a clear distinction between measurement noise and model-based uncertainty. Consistent MAP (blue) and mean (red) values across all parameters underscore unimodality and stable convergence of the BCE approach.

Figure 9 displays the results of the BCD approach, now extended to estimate the model discrepancy. The first four parameters (τ_n, χ, b, j) correspond to the physical model, while the next two (β_d, λ_d) are the GP hyperparameters of the model discrepancy, and as before, σ denotes the standard deviation of the experimental error. The posterior distributions for τ_n, χ, b , and j remain unimodal; however, the credible intervals are generally wider compared to schemes without explicit discrepancy modeling. This reflects the additional uncertainty introduced by acknowledging a possible mis-

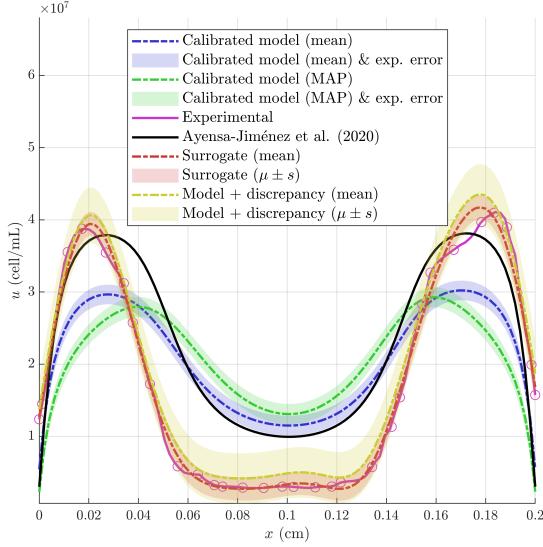


Fig. 5: Live cell profile generated using the BCED approach. Earlier results [7] showed an error of $e = 0.0768$, when compared to the prediction using the MAP parameters, with $e = 0.1132$, the prediction including the discrepancy with $e = 0.0138$ and the one using the surrogate with $e = 0.0116$.

match between model and reality. The GP hyperparameters for model discrepancy, β_d and λ_d , display broad and asymmetric posterior distributions, indicating significant uncertainty about the nature and scale of the discrepancy. Their inference enables the calibration framework to more flexibly account for systematic model errors. The experimental noise standard deviation, σ , stabilizes at a relatively low value, suggesting that the remaining variability attributed to measurement errors is well separated from both, parameter and model discrepancy uncertainties. Consistency between MAP (blue) and mean (red) estimates across most parameters continues to support stable convergence and essentially unimodal posteriors, despite the added flexibility of the discrepancy term. Several off-diagonal panels reveal correlations, especially between physical model parameters and the discrepancy hyperparameters, highlighting the statistical interplay between parameter estimation and model inadequacy correction. Including an explicit model discrepancy component in the Bayesian calibration thus increases the overall predictive uncertainty, more transparently combining the model uncertainty with the potential inadequacy of the model structure. This leads to more realistic uncertainty quantities and improves the trustworthiness of the predictive inferences.

Figure 10 illustrates the results of the BCED approach, incorporating both a surrogate model and a model discrepancy term. The first four parameters (τ_n, χ, b, j) come from the physical model; the next three $(\beta_x, \beta_\theta, \lambda_x)$ are the GP hyperparameters for the surrogate model, followed by two GP hyperparameters (β_d, λ_d) for the model discrepancy; finally, the experimental error standard deviation, σ . The marginal posterior distributions for all parameters are generally unimodal, indicating the parameters remain identifiable even with increased model flexibility. The credible intervals

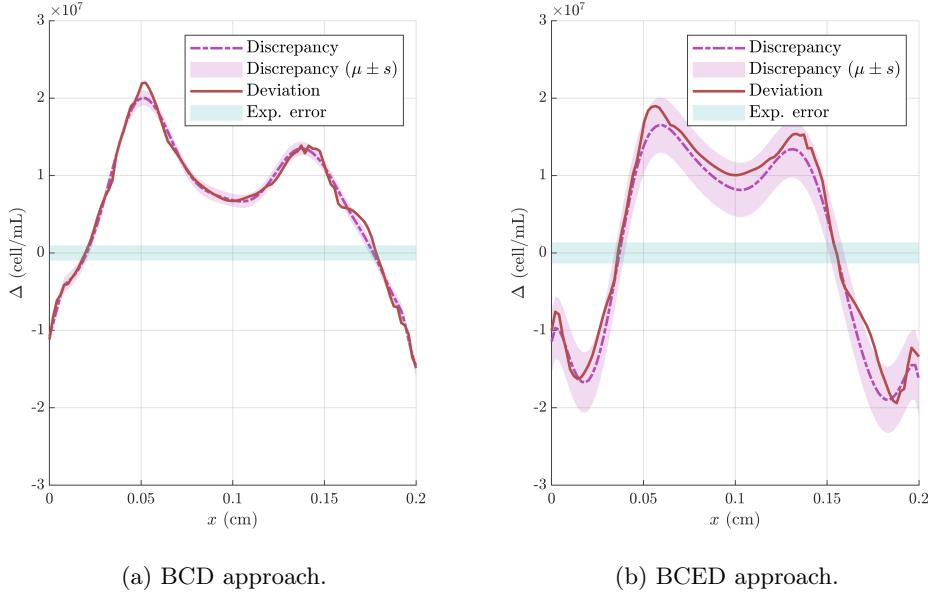


Fig. 6: Evaluation of the model adequacy. The discrepancy indicates the model inability to reproduce the experimental data; the deviation indicates the difference between the calibrated model and the data.

for the physical and surrogate model parameters are somewhat broader compared to models without a discrepancy component, reflecting the honest propagation of combined uncertainties. The GP hyperparameters for the surrogate ($\beta_x, \beta_\theta, \lambda_x$) and the discrepancy model (β_d, λ_d) possess clearly informative posteriors, demonstrating that the data provide nontrivial constraints on not just the physical and surrogate models but also the nature and scale of the discrepancy. Notable correlations are present between some surrogate and discrepancy hyperparameters, as evidenced by the elliptical shapes in their joint posterior panels. These dependencies illustrate the statistical interaction among model fidelity, surrogate expressiveness, and the allocated flexibility to correct for model inadequacy. The standard deviation of the experimental error, σ , remains tightly constrained and distinct from other sources of uncertainty, facilitating a clear partitioning between measurement noise and modeling uncertainty. The MAP (blue) and mean (red) estimates are close for all parameters, supporting the unimodal nature of the posteriors. In summary, the inclusion of both a surrogate and a model discrepancy term leads to a richer and more transparent quantification of predictive uncertainty, pointing in particular to potential shortcomings in the physical model and the surrogate representation.

Including the discrepancy term has many practical consequences beyond a more appropriate characterization of the distribution of the parameters. In Fig. 11, we show the cell profile and the experimental error obtained by Monte Carlo sub-sampling from the posterior distribution of the model parameters. More specifically, we generate N cell profiles u_i , $i = 1, \dots, N$, where each profile is obtained by means of

$$u_i(x) = \hat{u}_i(x) + \epsilon_i,$$

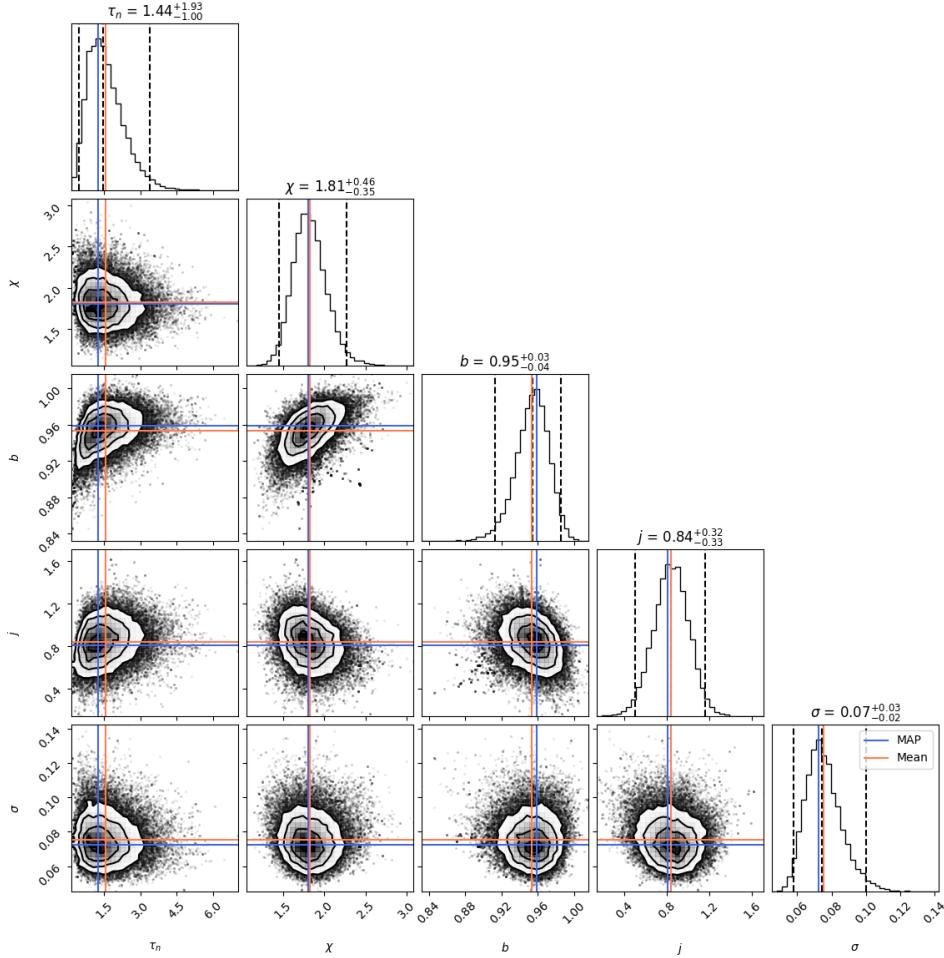


Fig. 7: **Posterior distribution for the BI approach.** The posterior distributions are obtained for 8000 samples, 16 chains, and 0.2 burn-in.

where $\hat{u}_i(x) = u(x, \boldsymbol{\theta}_i)$ is the cell profile predicted using the parameter $\boldsymbol{\theta}_i$ sampled from the posterior parameter distribution $\mathcal{F}_{\boldsymbol{\theta}}$ and ϵ_i is sampled from the error distribution $\mathcal{N}(0, \sigma^2)$. The figure shows the mean $\bar{u}(x)$ of a sample of size $N = 1000$ as well as the region $\bar{u}(x) \pm 2s_u(x)$, where $s_u(x)$ is the standard deviation:

$$\bar{u}(x) = \frac{1}{N} \sum_{i=1}^N \hat{u}_i(x), \quad s_u(x) = \left(\frac{1}{N-1} \sum_{i=1}^N (u_i(x) - \bar{u}(x))^2 \right)^{1/2}.$$

The benefit of including the discrepancy, in terms of uncertainty quantification, is illustrated in Fig. 12. In terms of predictive capacity, this approach can approximate the experimental data while keeping the uncertainty small. In fact, compared to the approach using the surrogate and the one using the inclusion of the discrepancy, we observe that the mean behavior is better described, and the confidence band defined

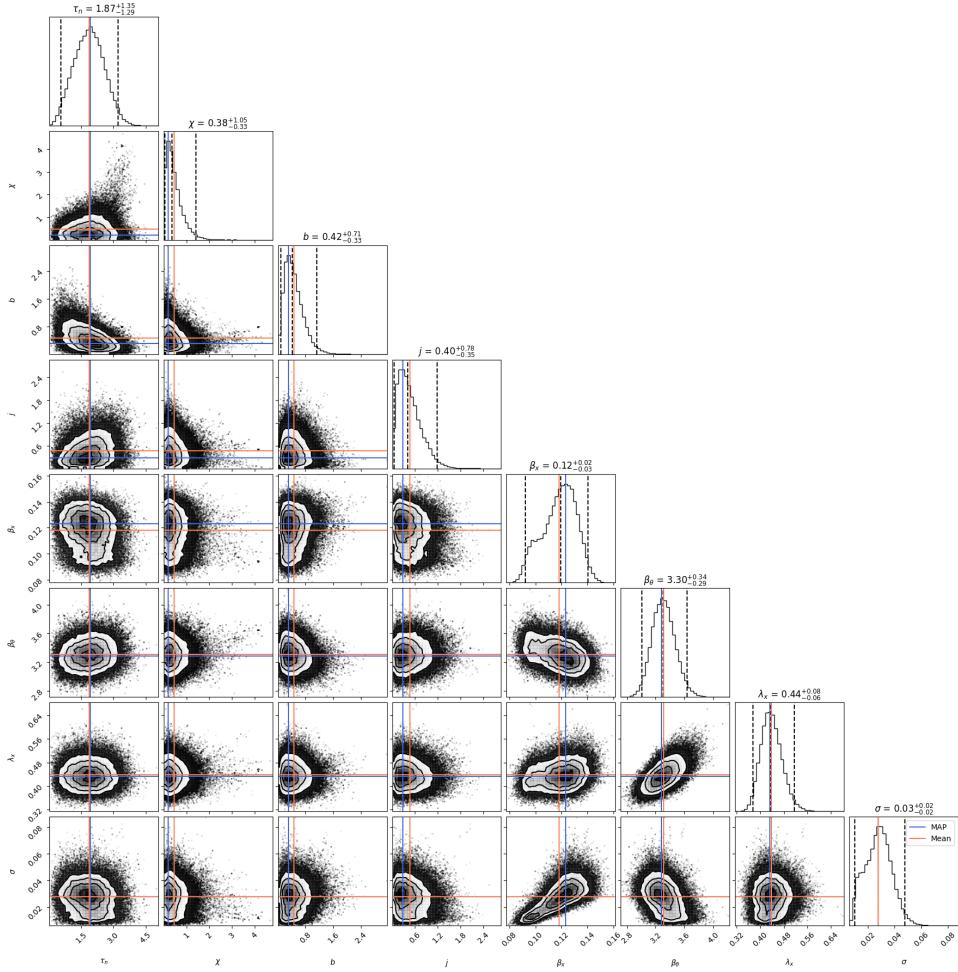


Fig. 8: **Posterior distribution for the BCD approach.** The posterior distributions are obtained for 20000 samples, 16 chains, and 0.2 burn-in.

as $\bar{u}(x) \pm 2s_u(x)$ is narrower and more regular.

As discussed before, in the presented approaches, the use of a surrogate model is merely instrumental to reduce computational cost. However, for completeness, in Fig. 13 we show the ability of the GP to recreate the biological process φ . The results are very similar in both cases, but adding the discrepancy term slightly increases the bias and uncertainty because it lets the model adjust for differences between the surrogate and the true biological process. This correction adds flexibility but introduces uncertainty and shifts the predictions since the calibration must fit data and account for possible model errors, instead of fitting data alone. In other words, modeling discrepancy acknowledges imperfection in the surrogate, giving more reliable but less certain and slightly adjusted predictions.

The computational times of the studied algorithms were evaluated to provide a comparative perspective, acknowledging variability due to system specifications,

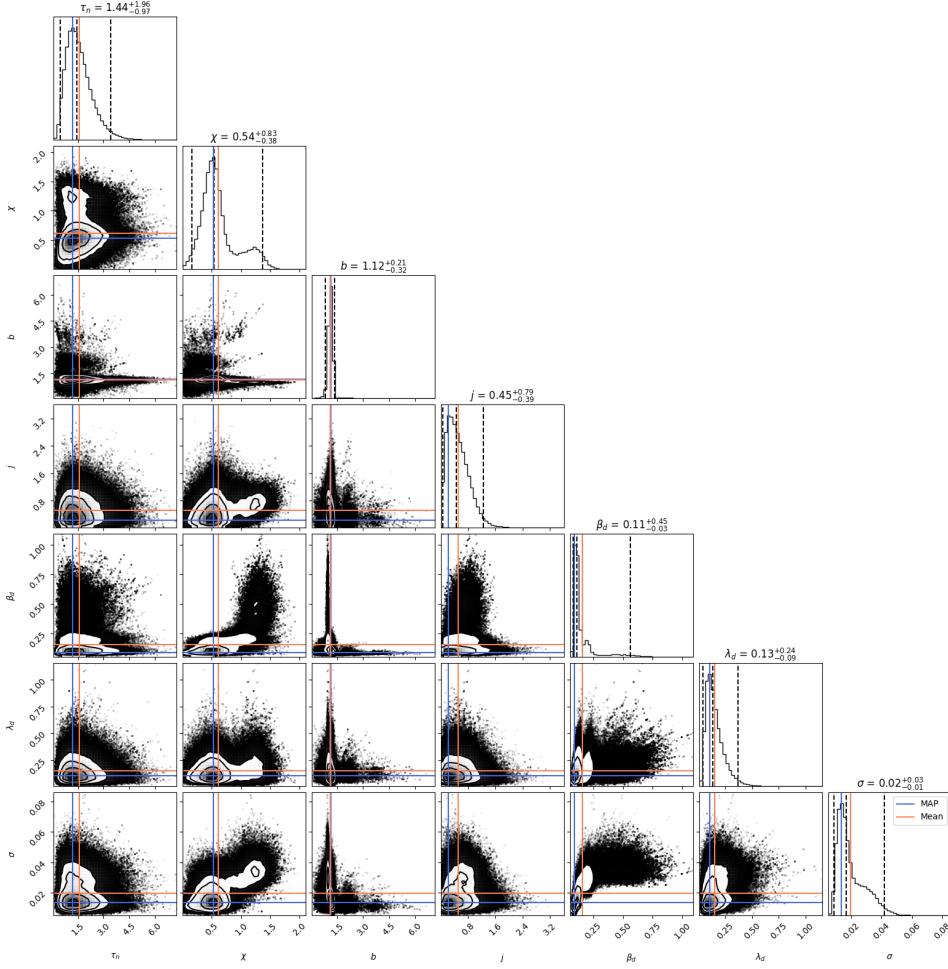


Fig. 9: **Posterior distribution for the BCE approach.** The posterior distributions are obtained for 100000 samples, 16 chains, and 0.2 burn-in.

network traffic, and other external factors. Specifically, BI required approximately 5 seconds per iteration; BCE achieved around 42 iterations per second, leading to a notable speedup using a surrogate model. BCD averaged around 5 seconds per iteration, while BCED reached approximately 22 iterations per second, also resulting in significant speed improvements through surrogate modeling. Compared to BI, the BCE approach achieves approximately a 208-fold speedup, while BCED attains about an 111-fold speedup relative to BCD. These improvements highlight the effectiveness of surrogate modeling in accelerating iteration performance.

The discrepancy in our modeling framework provides valuable information about the performance and limitations of the physical model. Specifically, a large discrepancy highlights that the current physical model is not sufficiently accurate, indicating room for improvement in its formulation. Although incorporating discrepancy enriches our understanding of model inadequacies, it does not lead to improved esti-

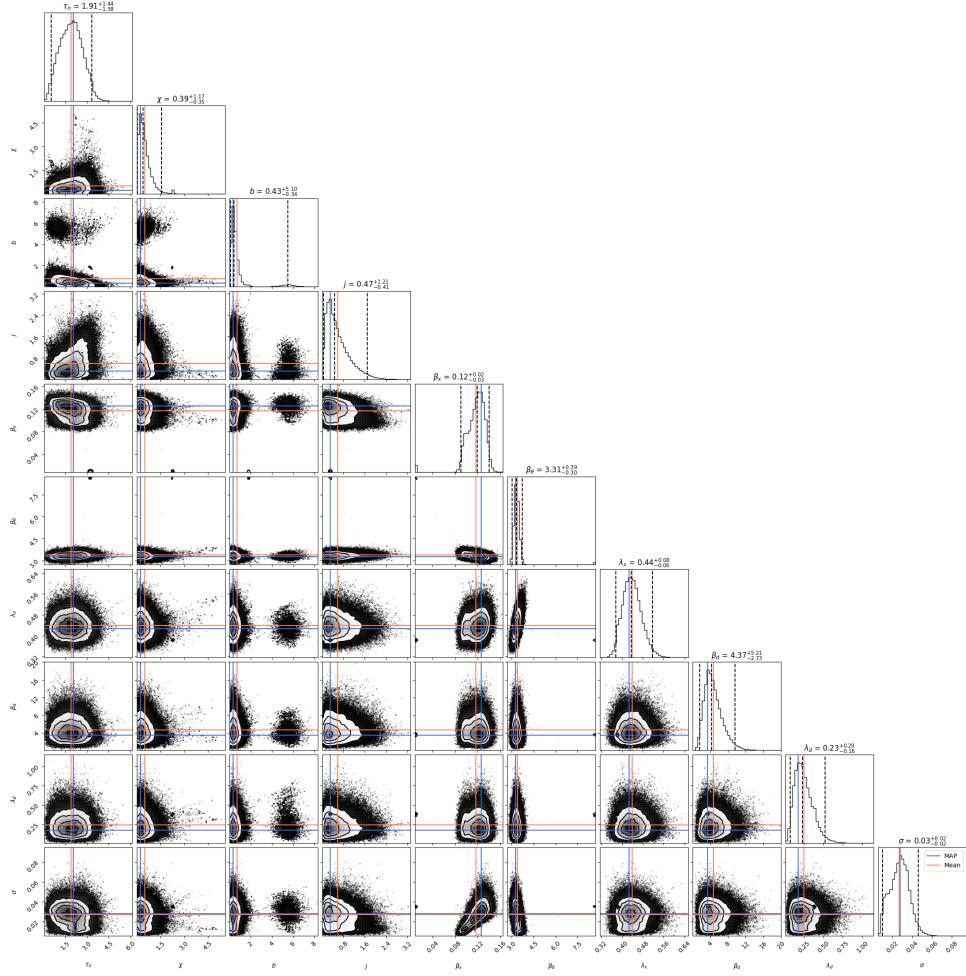


Fig. 10: **Posterior distribution for the BCED approach.** The posterior distributions are obtained for 30000 samples, 32 chains, and 0.2 burn-in.

mates of the model parameters themselves, suggesting challenges such as parameter identifiability or model structural errors that cannot be resolved by calibration alone. Indeed, in Table 1 we show the values of the parameters obtained using the BI and BCD approaches and we compare them to the ones reported in [7]. We focus on the parameters related with tumor progression (τ_n , χ and b) and not with the experimental setup (j). Even if the BI approach shows the smallest error between the prediction given by the model and the experimental profile, the value of the chemotaxis parameter χ is overestimated. This is because the work in [7] employed a different experimental scenario (pseudopalisade configuration), for which GBM progression has more migratory characteristics, and for which the identifiability of χ was clearer. The inclusion of the discrepancy does not guarantee increased accuracy in the estimated parameter, but it does indicate that the model calibration may not be appropriate, as here, without the need for a different validation scenario. The value of the remaining

parameters remain close to their reference. It is interesting to analyze the results in light of previous studies that have focused on characterizing the parameters related to go-or-grow [6]. The correlation that was reported between the parameters τ_n , χ and b because of the lack of identifiability for such a complex model (see for instance [8,65]) is also here revealed. However, both the spread and correlations of these parameters are reduced when including the discrepancy term, indicating that part of this uncertainty was due to model inadequacy. The variability of each parameter is also reduced when compared to the ranges found in the specialized literature [7].

	Values by [7]	BI	BCD
τ_n (s)	7.5×10^5	6.5×10^5	6.1×10^5
χ ($\text{cm}^2 \cdot \text{mmHg}^{-1} \cdot \text{s}^{-1}$)	7.5×10^{-9}	20×10^{-9}	5.8×10^{-9}
b (mmHg^{-1})	0.14	0.14	0.16
J ($\text{s} \cdot \text{cm}^{-1}$)	1×10^6	2.1×10^6	0.6×10^6

Table 1: **Values of the parameters estimated with and without the discrepancy inclusion.** The values reported are the MAP using physical units.

Using a surrogate model offers an alternative, computationally cheaper, that can be more predictive than the physical model. However, this benefit comes at a cost: the surrogate may compromise the quality of parameter estimates since it does not explicitly embed the underlying physics but relies instead on information derived from synthetic data. Consequently, while the surrogate is useful for saving computational cost, it lacks the interpretability and physical grounds necessary for robust parameter inference.

By employing a Bayesian calibration approach that integrates different methodologies, we obtain a more complete and nuanced picture of the modeling problem. This analysis enables quantification of various sources of uncertainty, including those from the physical model, the surrogate, and observational noise. As a result, we gain insights into the strengths and weaknesses not only of the original physical model, but also of the surrogate, informing subsequent model development and uncertainty management strategies. In practice, these findings emphasize the need to carefully select modeling approaches based on the specific goals of a study. Surrogate models are well-suited for prediction tasks that demand computational efficiency but lack interpretability, whereas physical models remain essential when physical interpretability and accurate parameter estimation are prioritized.

5. Conclusion. This paper has investigated the interplay of various Bayesian calibration strategies, including the use of parametric and surrogate models, with and without discrepancy, in the context of models for glioblastoma progression.

We first describe the modeling framework and detailed the calibration methodologies employed. Through comparative analysis of prediction accuracy, parameter identifiability, and uncertainty quantification, our results reveal the distinct strengths and complementarities of each approach, providing a detailed understanding of how these modeling choices impact both practical outcomes and theoretical insights.

Our findings demonstrate that leveraging Bayesian inference for this glioblastoma progression case study can provide complementary perspectives: parametric models retain interpretability and physical grounding, while surrogates deliver computational efficiency and strong predictive performance. The explicit consideration of model discrepancy enriches the process by highlighting when model structure is insufficient and

where improvements are needed. A key outcome of our investigation is the realization that these calibration strategies serve distinct but interconnected goals. While standard GP surrogate models excel in accurate and efficient prediction, particularly when computational resources are limited, they are less reliable for parameter inference due to the lack of embedded physical constraints. Conversely, calibration of the physical model, especially when accounting for discrepancy, provides greater insight into parameter identifiability and model limitations, revealing structural errors that cannot be addressed by calibration alone.

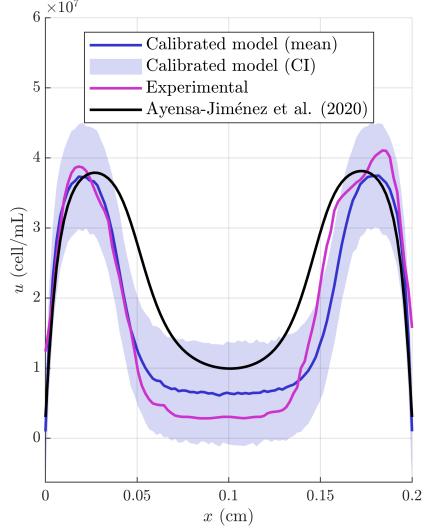
By systematically quantifying uncertainties from all sources – parametric model, surrogate, and observational noise – our Bayesian framework allows for a nuanced selection of modeling strategies based on whether prediction or physical understanding is prioritized. The results here illustrate that integrating multiple calibration methodologies not only improves predictive skill but also builds a deeper, more transparent understanding of model trustworthiness and limitations.

Overall, this work highlights the importance of an integrated calibration approach tailored to the specific objectives of the considered study. For future research on these types of glioblastoma progression models, we recommend continued development of hybrid and physics-informed surrogate models, as well as refinement of physical models to enhance their descriptive capacity and reduce structural errors. These efforts are also expected to foster calibration and prediction strategies that are less tightly bound to specific goals, promoting broader applicability across modeling scenarios.

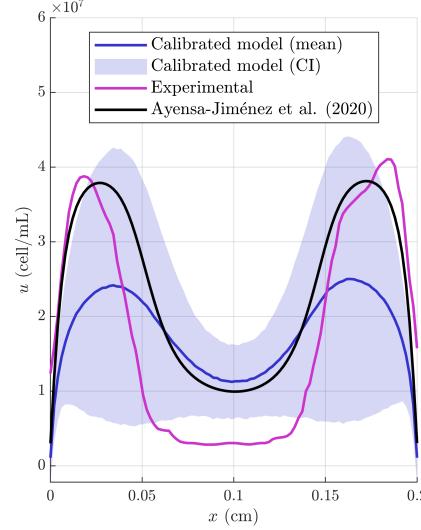
Author contributions. All authors have made substantial intellectual contributions to the study conception, execution, and design of the work. All authors have read and approved the final manuscript. In addition, the following contributions occurred: Conceptualization: Christina Schenk (C.S.), Ignacio Romero (I.R.); Methodology: C.S., Jacobo Ayensa-Jiménez (J.A-J.), I.R.; Formal analysis and investigation: C.S., J.A-J.; Writing: C.S., J.A-J., I.R.; original draft preparation: C.S., J.A-J., I.R.; Writing - review and editing: C.S., J.A-J., I.R.; Data Curation: C.S., J.A-J.; Visualization: C.S., J.A-J.; Software: C.S., J.A-J., I.R.; Funding acquisition: J.A-J..

Data and code availability statement. The data and the Matlab code that support the findings of this study are openly available in the GBM-SIMUL GitHub repository [9] at <https://github.com/Ayensa-Jimenez/GBM-SIMUL>. The Bayesian calibration code is provided as an attached .zip file for reviewer access and will be made publicly available through an open repository upon publication of this work.

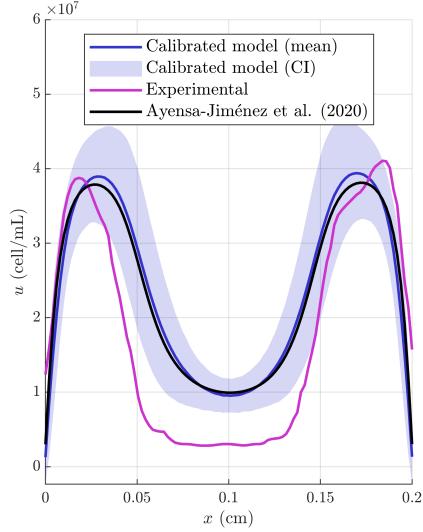
Acknowledgments. The authors acknowledge financial support from the Spanish Ministry of Science and Innovation through a Juan de la Cierva Postdoctoral Fellowship for J.A-J. (Grant No. JDC2023-052319-I). The authors acknowledge the use of the large language model Perplexity AI to assist in reformulating the wording of several sentences in the manuscript for clarity.



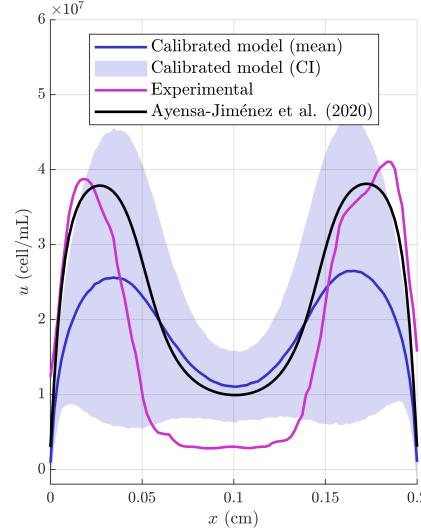
(a) Parametric model.



(b) Parametric model and surrogate.



(c) Parametric model and discrepancy.



(d) Par. model, surrogate and discrepancy.

Fig. 11: Robustness of the approach using the parametric model. The robustness is evaluated by means of the prediction bias and the prediction uncertainties when considering the *a posteriori* parameter distribution.

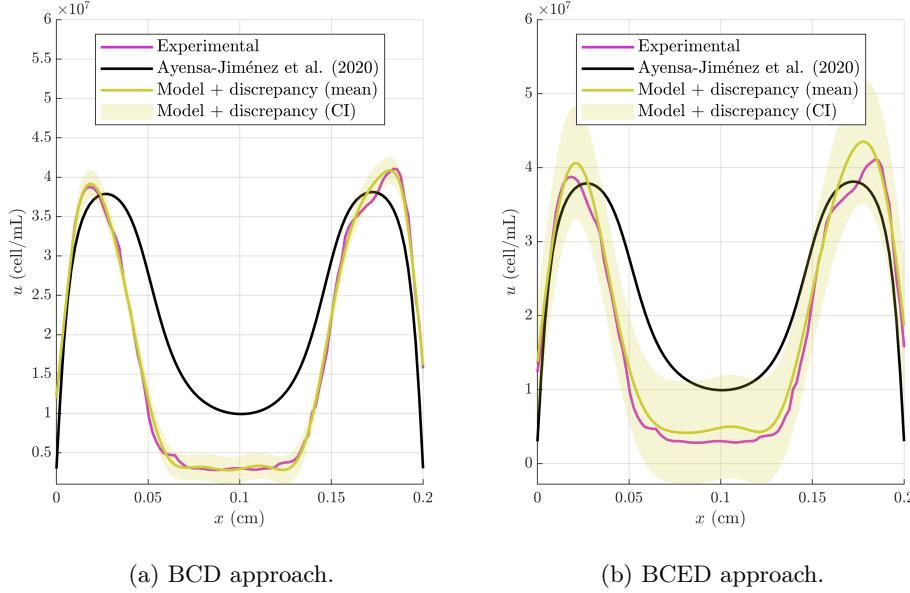


Fig. 12: **Prediction capacity when considering model discrepancy.** The robustness is evaluated by means of the prediction bias and the prediction uncertainties when considering the *a posteriori* parameter distribution.

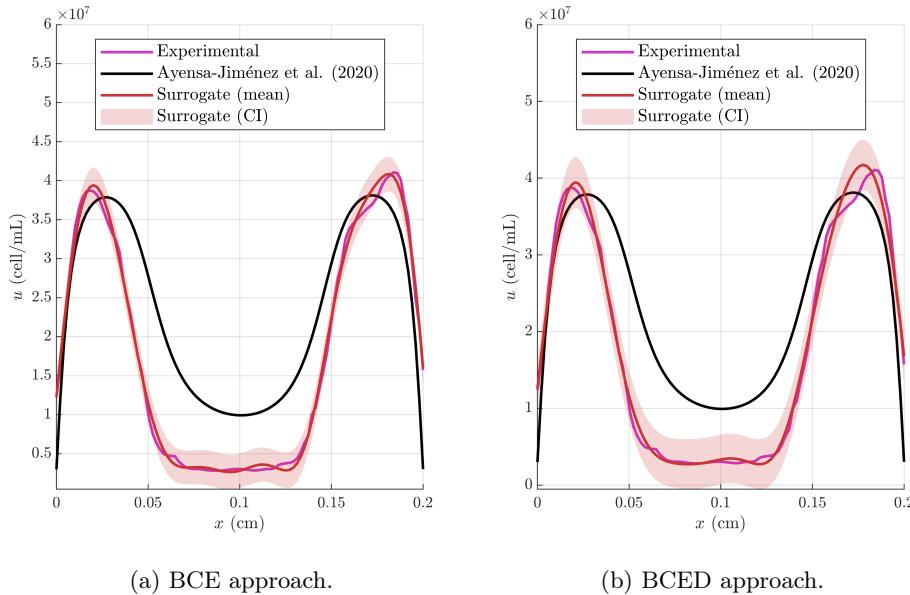


Fig. 13: **Performance of the surrogate evaluation.** We show the ability of the GP for accurately predict the experimental data for the BCD and BCED approaches.

REFERENCES

- [1] A. AGOSTI, P. CIARLETTA, H. GÄRCKE, AND M. HINZE, *Learning patient-specific parameters for a diffuse interface glioblastoma model from neuroimaging data*, Mathematical Methods in the Applied Sciences, 43 (2020), pp. 8945–8979.
- [2] T. ALPER AND P. HOWARD-FLANDERS, *Role of oxygen in modifying the radiosensitivity of *E. coli**, Nature, 178 (1956), pp. 978–979.
- [3] P. M. ALTROCK, L. L. LIU, AND F. MICHLER, *The mathematics of cancer: integrating quantitative models*, Nature Reviews Cancer, 15 (2015), pp. 730–745.
- [4] A. R. ANDERSON AND V. QUARANTA, *Integrative mathematical oncology*, Nature Reviews Cancer, 8 (2008), pp. 227–234.
- [5] J. AYENSA-JIMÉNEZ, M. H. DOWEIDAR, M. DOBLARÉ, AND E. A. GAFFNEY, *A mathematical modelling study of chemotactic dynamics in cell cultures: The impact of spatio-temporal heterogeneity*, Bulletin of Mathematical Biology, 85 (2023), p. 98.
- [6] J. AYENSA-JIMÉNEZ, M. PÉREZ-ALIACAR, M. H. DOWEIDAR, E. A. GAFFNEY, AND M. DOBLARÉ, *An overview from physically-based to data-driven approaches of the modelling and simulation of glioblastoma progression in microfluidic devices*, Archives of Computational Methods in Engineering, (2025), pp. 1–37.
- [7] J. AYENSA-JIMÉNEZ, M. PÉREZ-ALIACAR, T. RANDÈLOVIC, S. OLIVÁN, L. FERNÁNDEZ, J. A. SANZ-HERRERA, I. OCHOA, M. H. DOWEIDAR, AND M. DOBLARÉ, *Mathematical formulation and parametric analysis of in vitro cell models in microfluidic devices: application to different stages of glioblastoma evolution*, Scientific Reports, 10 (2020), pp. 1–21.
- [8] J. AYENSA-JIMÉNEZ, M. PÉREZ-ALIACAR, T. RANDÈLOVIC, J. A. SANZ-HERRERA, M. H. DOWEIDAR, AND M. DOBLARÉ, *Analysis of the parametric correlation in mathematical modeling of in vitro glioblastoma evolution using copulas*, Mathematics, 9 (2020), p. 27.
- [9] J. AYENSA-JIMÉNEZ, *Gbm-simul*. <https://github.com/Ayensa-Jimenez/GBM-SIMUL>, 2025. Commit d09ce5a, accessed: 2025-08-25.
- [10] J. M. AYUSO, H. A. BASHEER, R. MONGE, P. SÁNCHEZ-ÁLVAREZ, M. DOBLARÉ, S. D. SHNYDER, V. VINADER, K. AFARINKIA, L. J. FERNÁNDEZ, AND I. OCHOA, *Study of the chemotactic response of multicellular spheroids in a microfluidic device*, PloS one, 10 (2015), p. e0139515.
- [11] J. M. AYUSO, R. MONGE, A. MARTÍNEZ-GONZÁLEZ, M. VIRUMBRALES-MUÑOZ, G. A. LLAMAZARES, J. BERGANZO, A. HERNÁNDEZ-LAÍN, J. SANTOLARIA, M. DOBLARÉ, C. HUBERT, ET AL., *Glioblastoma on a microfluidic chip: Generating pseudopalisades and enhancing aggressiveness through blood vessel obstruction events*, Neuro-oncology, 19 (2017), pp. 503–513.
- [12] J. M. AYUSO, M. VIRUMBRALES-MUÑOZ, A. LACUEVA, P. M. LANUZA, E. CHECA-CHAVARRIA, P. BOTELLA, E. FERNÁNDEZ, M. DOBLARÉ, S. J. ALLISON, R. M. PHILLIPS, ET AL., *Development and characterization of a microfluidic model of the tumour microenvironment*, Scientific reports, 6 (2016), pp. 1–16.
- [13] J. M. AYUSO, M. VIRUMBRALES-MUÑOZ, J. M. LANG, AND D. J. BEEBE, *A role for microfluidic systems in precision medicine*, Nature communications, 13 (2022), p. 3086.
- [14] T. W. H. BACKMAN, C. SCHENK, T. RADIVOJEVIC, D. ANDO, J. SINGH, J. J. CZAJKA, Z. COSTELLO, J. D. KEASLING, Y. TANG, E. AKHMATSKAYA, AND H. GARCIA MARTIN, *Bayflux: A bayesian method to quantify metabolic fluxes and their uncertainty at the genome scale*, PLOS Computational Biology, 19 (2023), pp. 1–26, <https://doi.org/10.1371/journal.pcbi.1011111>, <https://doi.org/10.1371/journal.pcbi.1011111>.
- [15] A. BALDOCK, R. ROCKNE, A. BOONE, M. NEAL, A. HAWKINS-DAARUD, D. CORWIN, C. BRIDGE, L. GUYMAN, A. TRISTER, M. MRUGALA, ET AL., *From patient-specific mathematical neuro-oncology to precision medicine*, Frontiers in oncology, 3 (2013), p. 62.
- [16] A. BARBU, S.-C. ZHU, ET AL., *Monte Carlo Methods*, vol. 35, Springer, 2020.
- [17] C. BAYONA, L. ALZA, T. RANDÈLOVIĆ, M. C. SALLÁN, A. VISA, C. CANTÍ, I. OCHOA, S. OLIVÁN, AND J. HERREROS, *Tetralol derivative nnc-55-0396 targets hypoxic cells in the glioblastoma microenvironment: an organ-on-chip approach*, Cell Death & Disease, 15 (2024), p. 127.
- [18] D. J. BRAT, *Glioblastoma: biology, genetics, and behavior*, American Society of Clinical Oncology Educational Book, 32 (2012), pp. 102–107.
- [19] D. J. BRAT, A. A. CASTELLANO-SANCHEZ, S. B. HUNTER, M. PECOT, C. COHEN, E. H. HAMMOND, S. N. DEVI, B. KAUR, AND E. G. VAN MEIR, *Pseudopalisades in glioblastoma are hypoxic, express extracellular matrix proteases, and are formed by an actively migrating cell population*, Cancer research, 64 (2004), pp. 920–927.
- [20] D. J. BRAT AND E. G. VAN MEIR, *Vaso-occlusive and prothrombotic mechanisms associated*

- with tumor hypoxia, necrosis, and accelerated growth in glioblastoma*, Laboratory investigation, 84 (2004), pp. 397–405.
- [21] H. M. BYRNE, T. ALARCON, M. R. OWEN, S. D. WEBB, AND P. K. MAINI, *Modelling aspects of cancer dynamics: a review*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 364 (2006), pp. 1563–1578.
- [22] D. CAMACHO-GOMEZ, C. BORAU, J. M. GARCIA-AZNAR, M. J. GOMEZ-BENITO, M. GIROLAMI, AND M. A. PEREZ, *Physics-informed machine learning digital twin for reconstructing prostate cancer tumor growth via psa tests*, npj Digital Medicine, 8 (2025), p. 485.
- [23] A. CARREAU, B. E. HAFNY-RAHBI, A. MATEJKU, C. GRILLON, AND C. KIEDA, *Why is the partial oxygen pressure of human tissues a crucial parameter? small molecules and hypoxia*, Journal of cellular and molecular medicine, 15 (2011), pp. 1239–1253.
- [24] A. CHONG AND K. MENBERG, *Guidelines for the bayesian calibration of building energy models*, Energy and Buildings, 174 (2018), p. 527–547, <https://doi.org/10.1016/j.enbuild.2018.06.028>.
- [25] J. COLLIS, A. J. CONNOR, M. PACZKOWSKI, P. KANNAN, J. PITT-FRANCIS, H. M. BYRNE, AND M. E. HUBBARD, *Bayesian calibration, validation and uncertainty quantification for predictive modelling of tumour growth: a tutorial*, Bulletin of mathematical biology, 79 (2017), pp. 939–974.
- [26] M. DANEKER, Z. ZHANG, G. E. KARNIADAKIS, AND L. LU, *Systems biology: Identifiability analysis and parameter identification via systems-biology-informed neural networks*, in Computational Modeling of Signaling Networks, Springer, 2023, pp. 87–105.
- [27] A. DAŞU, I. TOMA-DAŞU, AND M. KARLSSON, *Theoretical simulation of tumour oxygenation and results from acute and chronic hypoxia*, Physics in Medicine & Biology, 48 (2003), p. 2829.
- [28] J. DE PABLOS, I. SABIROV, AND I. ROMERO, *A methodology for the calibration of complex material models: application to thermo-elasto-plastic materials for high-velocity impact simulations*, Archives of Computational Methods in Engineering, (2023), <https://doi.org/10.1007/s11831-023-09888-y>.
- [29] T. S. DEISBOECK, Z. WANG, P. MACKLIN, AND V. CRISTINI, *Multiscale cancer modeling*, Annual review of biomedical engineering, 13 (2011), pp. 127–155.
- [30] R. EDMONDSON, J. J. BROGLIE, A. F. ADCOCK, AND L. YANG, *Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors*, Assay and drug development technologies, 12 (2014), pp. 207–218.
- [31] H. ENDERLING, J. C. L. ALFONSO, E. MOROS, J. J. CAUDELL, AND L. B. HARRISON, *Integrating mathematical modeling into the roadmap for personalized adaptive radiation therapy*, Trends in cancer, 5 (2019), pp. 467–474.
- [32] H. ENDERLING AND O. WOLKENHAUER, *Are all models wrong?*, Computational and systems oncology, 1 (2021), p. e1008.
- [33] J. M. ESPOSITO, *Concentration of measure phenomenon and its implications for sample-based planning algorithms in very-high dimensional configuration spaces*, in 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023, pp. 7865–7871, <https://doi.org/10.1109/ICRA48891.2023.10160286>.
- [34] R. A. FISHER, *The wave of advance of advantageous genes*, Annals of eugenics, 7 (1937), pp. 355–369.
- [35] D. FOREMAN-MACKEY, W. M. FARR, M. SINHA, A. M. ARCHIBALD, D. W. HOGG, J. S. SANDERS, J. ZUNTZ, P. K. G. WILLIAMS, A. R. J. NELSON, M. DE VAL-BORRO, T. ERHARDT, I. PASHCHENKO, AND O. A. PLA, *emcee v3: A python ensemble sampling toolkit for affine-invariant mcmc*, Journal of Open Source Software, 4 (2019), p. 1864, <https://doi.org/10.21105/joss.01864>.
- [36] D. FOREMAN-MACKEY, D. W. HOGG, D. LANG, AND J. GOODMAN, *<tt>emcee</tt>: The mcmc hammer*, Publications of the Astronomical Society of the Pacific, 125 (2013), p. 306–312, <https://doi.org/10.1086/670067>, <http://dx.doi.org/10.1086/670067>.
- [37] K. FUNAMOTO, I. K. ZERVANTONAKIS, Y. LIU, C. J. OCHS, C. KIM, AND R. D. KAMM, *A novel microfluidic platform for high-resolution imaging of a three-dimensional cell culture under a controlled hypoxic environment*, Lab on a chip, 12 (2012), pp. 4855–4863.
- [38] L. GALLUZZI, I. VITALE, S. A. AARONSON, J. M. ABRAMS, D. ADAM, P. AGOSTINIS, E. S. ALNEMRI, L. ALTUCCI, I. AMELIO, D. W. ANDREWS, ET AL., *Molecular mechanisms of cell death: recommendations of the nomenclature committee on cell death 2018*, Cell Death & Differentiation, 25 (2018), pp. 486–541.
- [39] S. GOUTELLE, M. MAURIN, F. ROUGIER, X. BARBAUT, L. BOURGUIGNON, M. DUCHER, AND P. MAIRE, *The hill equation: a review of its capabilities in pharmacological modelling*, Fundamental & clinical pharmacology, 22 (2008), pp. 633–648.

- [40] D. HANAHAN, *Hallmarks of cancer: new dimensions*, Cancer discovery, 12 (2022), pp. 31–46.
- [41] P. HANSEN, V. PEREYRA, AND G. SCHERER, *Least Squares Data Fitting with Applications*, Least Squares Data Fitting with Applications, Johns Hopkins University Press, 2013.
- [42] H. HATZIKIROU, D. BASANTA, M. SIMON, K. SCHALLER, AND A. DEUTSCH, ‘go or grow’: the key to the emergence of invasion in tumour progression?, Mathematical medicine and biology: a journal of the IMA, 29 (2012), pp. 49–65.
- [43] D. HIGDON, J. GATTIKER, B. WILLIAMS, AND M. RIGHTLEY, Computer model calibration using high-dimensional output, Journal of the American Statistical Association, 103 (2008), p. 570–583, <https://doi.org/10.1198/016214507000000888>, <http://dx.doi.org/10.1198/016214507000000888>.
- [44] D. HIGDON, M. KENNEDY, J. C. CAVENDISH, J. A. CAFEZO, AND R. D. RYNE, Combining Field Data and Computer Simulations for Calibration and Prediction, SIAM Journal on Scientific Computing, 26 (2004), pp. 448–466, <https://doi.org/10.1137/S1064827503426693>.
- [45] K. C. HIREMATH, K. ATAKISHI, E. A. LIMA, M. FARHAT, B. PANTHI, H. LANGSHAW, M. D. SHANKER, W. TALPUR, S. THROWER, J. GOLDMAN, ET AL., Identifiability and model selection frameworks for models of high-grade glioma response to chemoradiation, Philosophical Transactions A, 383 (2025), p. 20240212.
- [46] B. HOFSTEE, Non-inverted versus inverted plots in enzyme kinetics, Nature, 184 (1959), pp. 1296–1298.
- [47] D. A. HORMUTH, J. A. WEIS, S. L. BARNES, M. I. MIGA, E. C. RERICHA, V. QUARANTA, AND T. E. YANKEELOV, A mechanically coupled reaction–diffusion model that incorporates intra-tumoural heterogeneity to predict in vivo glioma growth, Journal of The Royal Society Interface, 14 (2017), p. 20161010.
- [48] R. IMAN, J. HELTON, AND J. CAMPBELL, An approach to sensitivity analysis of computer models: Part i—introduction, input variable selection and preliminary variable assessment, Journal of Quality Technology, 13 (1981), p. 174–183, <https://doi.org/10.1080/00224065.1981.11978748>, <http://dx.doi.org/10.1080/00224065.1981.11978748>.
- [49] A. S. KAZEROUNI, M. GADDE, A. GARDNER, D. A. HORMUTH, A. M. JARRETT, K. E. JOHNSON, E. A. LIMA, G. LORENZO, C. PHILLIPS, A. BROCK, ET AL., Integrating quantitative assays with biologically based mathematical modeling for predictive oncology, Iscience, 23 (2020).
- [50] M. C. KENNEDY AND A. O’HAGAN, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B. (Statistical Methodology), 63 (2001), p. 425–464, <https://doi.org/10.1111/1467-9868.00294>, <http://dx.doi.org/10.1111/1467-9868.00294>.
- [51] J. LATZ, Bayesian inverse problems are usually well-posed, SIAM Review, 65 (2023), pp. 831–865.
- [52] K. F. LEI, M.-H. WU, C.-W. HSU, AND Y.-D. CHEN, Real-time and non-invasive impedimetric monitoring of cell proliferation and chemosensitivity in a perfusion 3d cell culture microfluidic chip, Biosensors and Bioelectronics, 51 (2014), pp. 16–21.
- [53] E. LIMA, J. ODEN, D. HORMUTH, T. YANKEELOV, AND R. ALMEIDA, Selection, calibration, and validation of models of tumor growth, Mathematical Models and Methods in Applied Sciences, 26 (2016), pp. 2341–2368.
- [54] H. LINEWEAVER AND D. BURK, The determination of enzyme dissociation constants, Journal of the American chemical society, 56 (1934), pp. 658–666.
- [55] G. LORENZO, S. R. AHMED, D. A. HORMUTH, B. VAUGHN, J. KALPATHY-CRAMER, L. SOLORIO, T. E. YANKEELOV, AND H. GOMEZ, Patient-specific, mechanistic models of tumor growth incorporating artificial intelligence and big data, Annual Review of Biomedical Engineering, 26 (2024), pp. 529–560.
- [56] G. LORENZO, M. A. SCOTT, K. TEW, T. J. HUGHES, Y. J. ZHANG, L. LIU, G. VILANOVA, AND H. GOMEZ, Tissue-scale, personalized modeling and simulation of prostate cancer growth, Proceedings of the National Academy of Sciences, 113 (2016), pp. E7663–E7671.
- [57] X. LU AND Y. KANG, Hypoxia and hypoxia-inducible factors: master regulators of metastasis, Clinical cancer research, 16 (2010), pp. 5928–5935.
- [58] M. MCKAY, R. BECKMAN, AND W. CONOVER, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21 (1979), pp. 55–61, <https://doi.org/10.2307/1268522>, <http://dx.doi.org/10.2307/1268522>.
- [59] M. T. MCKENNA, J. A. WEIS, A. BROCK, V. QUARANTA, AND T. E. YANKEELOV, Precision medicine with imprecise therapy: computational modeling for chemotherapy in breast cancer, Translational oncology, 11 (2018), pp. 732–742.
- [60] M. MEHLING AND S. TAY, Microfluidic cell culture, Current opinion in Biotechnology, 25 (2014), pp. 95–102.

- [61] J. METZCAR, C. R. JUTZELER, P. MACKLIN, A. KÖHN-LUQUE, AND S. C. BRÜNINGK, *A review of mechanistic learning in mathematical oncology*, Frontiers in Immunology, 15 (2024), p. 1363144.
- [62] J. T. ODEN, I. BABUŠKA, AND D. FAGHIHI, *Predictive computational science: Computer predictions in the presence of uncertainty*, Encyclopedia of Computational Mechanics Second Edition, (2017), pp. 1–26.
- [63] J. T. ODEN, E. A. LIMA, R. C. ALMEIDA, Y. FENG, M. N. RYLANDER, D. FUENTES, D. FAGHIHI, M. M. RAHMAN, M. DEWITT, M. GADDE, ET AL., *Toward predictive multiscale modeling of vascular tumor growth: computational and experimental oncology for tumor prediction*, Archives of Computational Methods in Engineering, 23 (2016), pp. 735–779.
- [64] A. O'HAGAN, *Bayesian analysis of computer code outputs: A tutorial*, Reliability Engineering & System Safety, 91 (2006), pp. 1290–1300, <https://doi.org/10.1016/j.ress.2005.11.025>, <https://linkinghub.elsevier.com/retrieve/pii/S0951832005002383>.
- [65] M. PÉREZ-ALIACAR, J. AYENSA-JIMÉNEZ, M. H. DOWEIDAR, AND M. DOBLARÉ, *Massive data probabilistic framework for parameter estimation in biological problems*, in Digital Human Modeling and Medicine, Elsevier, 2023, pp. 19–62.
- [66] M. PÉREZ-ALIACAR, M. H. DOWEIDAR, M. DOBLARÉ, AND J. AYENSA-JIMÉNEZ, *Predicting cell behaviour parameters from glioblastoma on a chip images. a deep learning approach*, Computers in Biology and Medicine, 135 (2021), p. 104547.
- [67] Á. PÉREZ-BENITO, J. M. GARCÍA-AZNAR, M. J. GÓMEZ-BENITO, AND M. Á. PÉREZ, *Patient-specific prostate tumour growth simulation: A first step towards the digital twin*, Frontiers in Physiology, 15 (2024), p. 1421591.
- [68] C. M. PHILLIPS, E. A. LIMA, C. WU, A. M. JARRETT, Z. ZHOU, N. ELSHAFFEY, J. MA, G. RAUCH, AND T. E. YANKEELOV, *Assessing the identifiability of model selection frameworks for the prediction of patient outcomes in the clinical breast cancer setting*, Journal of Computational Science, 69 (2023), p. 102006.
- [69] C. E. RASMUSSEN AND C. K. I. WILLIAMS, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Massachusetts, 2006.
- [70] A. RAUE, J. KARLSSON, M. P. SACCOMANI, M. JIRSTRAND, AND J. TIMMER, *Comparison of approaches for parameter identifiability analysis of biological systems*, Bioinformatics, 30 (2014), pp. 1440–1448.
- [71] M. N. READ, K. ALDEN, J. TIMMIS, AND P. S. ANDREWS, *Strategies for calibrating models of biology*, Briefings in Bioinformatics, 21 (2020), pp. 24–35.
- [72] C. P. ROBERT AND G. CASELLA, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer New York, 2004, <https://doi.org/10.1007/978-1-4757-4145-2>, <http://dx.doi.org/10.1007/978-1-4757-4145-2>.
- [73] I. ROMERO AND M. ORTIZ, *An energy-stepping markov Monte Carlo method*, Meccanica, 60 (2025), p. 1411–1436, <https://doi.org/10.1007/s11012-025-01997-1>, <http://link.springer.com/article/10.1007/s11012-025-01997-1>.
- [74] Y. RONG, D. L. DURDEN, E. G. VAN MEIR, AND D. J. BRAT, ‘*pseudopalisading’ necrosis in glioblastoma: a familiar morphologic feature that links vascular pathology, hypoxia, and angiogenesis*’, Journal of Neuropathology & Experimental Neurology, 65 (2006), pp. 529–539.
- [75] E. K. SACKMANN, A. L. FULTON, AND D. J. BEEBE, *The present and future role of microfluidics in biomedical research*, Nature, 507 (2014), pp. 181–189.
- [76] S. SANGA, H. B. FRIEBOES, X. ZHENG, R. GATENBY, E. L. BEARER, AND V. CRISTINI, *Predictive oncology: a review of multidisciplinary, multiscale in silico modeling linking phenotype, morphology and growth*, Neuroimage, 37 (2007), pp. S120–S134.
- [77] J. W. SCANNELL, A. BLANCKLEY, H. BOLDON, AND B. WARRINGTON, *Diagnosing the decline in pharmaceutical r&d efficiency*, Nature reviews Drug discovery, 11 (2012), pp. 191–200.
- [78] C. SCHENK AND M. HARANCZYK, *A novel constrained sampling method for efficient exploration in materials and chemical mixture design*, Computational Materials Science, 252 (2025), p. 113780, <https://www.sciencedirect.com/science/article/pii/S0927025625001235?dgcid=author>.
- [79] C. SCHENK AND I. ROMERO, *ACBICI - A Configurable Bayesian Calibration and Inference package*, <https://materials.imdea.org/acbici-a-configurable-bayesian-calibration-and-inference-package/>, 2025.
- [80] A. SENDOEL AND M. O. HENGARTNER, *Apoptotic cell death under hypoxia*, Physiology, 29 (2014), pp. 168–176.
- [81] R. D. SKEEL AND M. BERZINS, *A method for the spatial discretization of parabolic equations in one space variable*, SIAM journal on scientific and statistical computing, 11 (1990),

- pp. 1–32.
- [82] A. C. TAN, D. M. ASHLEY, G. Y. LÓPEZ, M. MALINZAK, H. S. FRIEDMAN, AND M. KHASRAW, *Management of glioblastoma: State of the art and future directions*, CA: a cancer journal for clinicians, 70 (2020), pp. 299–312.
 - [83] I. F. TANNOCK, *Oxygen diffusion and the distribution of cellular radiosensitivity in tumours*, The British journal of radiology, 45 (1972), pp. 515–524.
 - [84] F. F. TAO, X. XIAO, K. F. LEI, AND I.-C. LEE, *Based cell culture microfluidic system*, BioChip Journal, 9 (2015), pp. 97–104.
 - [85] M. VAN OIJEN, J. ROUGIER, AND R. SMITH, *Bayesian calibration of process-based forest models: bridging the gap between models and data*, Tree Physiology, 25 (2005), p. 915–927, <https://doi.org/10.1093/treephys/25.7.915>, <http://dx.doi.org/10.1093/treephys/25.7.915>.
 - [86] M. VINCI, S. GOWAN, F. BOXALL, L. PATTERSON, M. ZIMMERMANN, C. LOMAS, M. MEDDIOLA, D. HARDISSON, S. A. ECCLES, ET AL., *Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation*, BMC biology, 10 (2012), pp. 1–21.
 - [87] H. WILLMANN, J. NITZLER, S. BRANDSTÄTER, AND W. A. WALL, *Bayesian calibration of coupled computational mechanics models under uncertainty based on interface deformation*, Advanced Modeling and Simulation in Engineering Sciences, 9 (2022), <https://doi.org/10.1186/s40323-022-00237-5>, <http://dx.doi.org/10.1186/s40323-022-00237-5>.
 - [88] C. WU, G. LORENZO, D. A. HORMUTH, E. A. LIMA, K. P. ŠLAVKOVA, J. C. DiCARLO, J. VIROSTKO, C. M. PHILLIPS, D. PATT, C. CHUNG, ET AL., *Integrating mechanism-based modeling with biomedical imaging to build practical digital twins for clinical oncology*, Biophysics reviews, 3 (2022).
 - [89] T. E. YANKEELOV, G. AN, O. SAUT, E. G. LUEBECK, A. S. POPEL, B. RIBBA, P. VICINI, X. ZHOU, J. A. WEIS, K. YE, ET AL., *Multi-scale modeling in clinical oncology: opportunities and barriers to success*, Annals of biomedical engineering, 44 (2016), pp. 2626–2641.
 - [90] T. E. YANKEELOV, D. A. HORMUTH, E. A. LIMA, G. LORENZO, C. WU, L. C. OKEREKE, G. M. RAUCH, A. M. VENKATESAN, AND C. CHUNG, *Designing clinical trials for patients who are not average*, Iscience, 27 (2024).
 - [91] T. E. YANKEELOV, V. QUARANTA, K. J. EVANS, AND E. C. RERICHA, *Toward a science of tumor forecasting for clinical oncology*, Cancer research, 75 (2015), pp. 918–923.

Appendix A. Computational implementation. GBM model was implemented in Matlab R2023b using the `pdepe` solver for systems of 1-D parabolic and elliptic PDEs. The `pdepe` routine uses a space-time integrator based on a piecewise nonlinear Galerkin approach that is second order accurate in space [81], computing the solution over a specified spatial mesh of size $N_x = 100$ and time span with $N_t = 100$ (although the numerical solver is adaptive in time). Linear interpolation of the experimental data corresponding to the initial condition was performed using `interp1` to map the initial cell profile to the values at given spatial coordinates. The GBM progression simulation code is free and publicly available [9] (<https://github.com/Ayensa-Jimenez/GBM-SIMUL>).

The Bayesian calibration procedure was performed in Python using ACBICI 2.0.0, an in-house software package [79], leveraging core scientific libraries including `scipy`, `numpy`, and `math`. Experimental design for synthetic datasets was conducted using the `pyDOE` package, employing Latin Hypercube sampling to ensure efficient coverage of the parameter space. For Bayesian inference, the `emcee` package [35,36] was utilized to perform Markov Chain Monte Carlo (MCMC) sampling, enabling efficient exploration of the posterior distribution. Graphical illustrations and visualizations were generated with the `matplotlib`, `corner` and `seaborn` libraries. ACBICI 2.0.0 and the code used for the case studies are provided in an attached .zip file for reviewer access and will be made publicly available in an open repository upon publication of this work.

Appendix B. Data and model configurations.

B.1. Data preparation. A selection of experimental data points was performed to ensure comprehensive coverage of the input domain while maintaining computational efficiency. Initially, the data points corresponding to the minimum and max-

imum values of the key input variables were identified and included to anchor the dataset boundaries. Subsequently, 28 additional points were selected using Latin Hypercube Sampling (LHS) to uniformly cover the input space in a stratified manner.

For synthetic data generation, a set of 200 data points was selected at random from a larger pool of 500 simulations conducted in MATLAB. These synthetic simulations were performed by sampling the model parameters uniformly in the nominal interval [0.1, 6]. This selection reflects plausible variations of the parameters within the identified operational range and provides a sufficiently large and diverse surrogate dataset compromising sufficient knowledge and computational efficiency for subsequent model calibration and validation.

B.2. Priors. The default priors from ACBICI 2.0.0 were chosen. Specifically, the hyperparameters were assigned $\text{Gamma}(\alpha, \beta)$ priors scaled to the characteristic magnitudes of the data. When no prior information was available, the shape parameter was fixed at $\alpha = 5$, while the rate parameter β was determined from empirical statistics of the inputs or outputs. For the length-scale parameters β_x and β_t , the rate was set to $\beta = 5/d$, with d denoting the average pairwise distance in the experimental inputs or synthetic time coordinates, respectively. For the variance-related parameters λ_x and λ_d , the rate was chosen as $\beta = 5/\sigma$, where σ is the empirical standard deviation of the experimental responses.

For the observational error standard deviation σ_ϵ , the prior was constructed from the mean absolute response amplitude, $\bar{y} = \text{mean}(|y^{\text{exp}}|)$. The prior mean was set to $0.1\bar{y}$, with a standard deviation equal to 0.1 of this prior mean. These values defined the parameters of a Gamma distribution, ensuring that the prior expectation and variance reflected a small but non-negligible fraction of the observed signal magnitude. This specification yields weakly informative priors that are scaled to the natural variability of the observed data, preventing unrealistic hyperparameter values while retaining calibration flexibility.